**Calibrating Models to Data: A Comparison of Methods**

**Introduction**

Most sciences today use mathematical and computer simulation models to approximate the real-world processes under study (Kennedy), (Fojo), (Vanni). For example, models play a significant role in health policymaking by estimating the impact of interventions in situations where empirical studies may be time-consuming, costly and impractical (Stout). Developing a model calls for a trade-off between computational cost and accuracy; simple models require little computation time but can be a poor description of the real-world process, whereas complex models allow for a more accurate description of the process at the cost of increased computational cost. After model development, it is imperative to know how well the model represents reality. Model calibration, or fitting the model to data, increases the confidence that the model provides a realistic approximation to the real-world process (Vanni), (Stout).

Calibration is the process of comparing model outputs with empirical data to identify the model parameter values that achieve a good fit to data (Menzies), (Vanni). Calibration improves the credibility and validity of the subsequent predictions made and inferences drawn from the model (Stout). It is also commonly used in the case where model parameters are not observable or available, to estimate such input parameters (Elske). The main components of calibration are summary statistics, parameter-search strategy, goodness-of-fit (GOF) measure and acceptance criteria.

Several methods have been used for model calibration and the number of studies that apply these calibration methods is proliferating in many research fields (Vanni). (Stout) broke the model calibration process into seven stages which were later discussed in detail by (Vanni). (Karnon) went through the seven stages of the calibration process using an early breast cancer model and produced a practical guidance on a more applicable calibration process. (Vanni), in their review article further examined different methods of calibration and reviewed some examples from health economic decision models. The model calibration methods applied in most studies are in two categories, optimisation methods and sampling methods (Menzies). For the purpose of this study, we focus on the sampling methods.

Because there are many model calibration methods with little or no consensus on their performance, we perform a simulation study to compare the performance of model calibration methods using a simple stochastic Susceptible-Infected-Recovered (SIR) model. The methods to be compared are Rejection Approximate Bayesian Computation (Rejection ABC), Sequential Approximate Bayesian Computation (Sequential ABC) and Bayesian Maximum Likelihood estimation (BMLE).

Outline to be completed when thesis is fully written…….

**METHODS**

**2.1. METHODS TO BE COMPARED**

1. **Bayesian calibration methods**

Approximate Bayesian Computation (ABC) consists of computational methods and techniques that make use of Bayesian statistics. These ABC techniques are relevant for calibrating stochastic models to empirical data because they are easy to implement and can be applied to any model (Sunnaker)

1. **Rejection ABC**

Rejection ABC is the first and most basic form of ABC. This method operates by sampling parameter values () from the prior distributionand given these sampled parameter values, data (*y*) is simulated under a model. A summary statistic () of the simulated data () must satisfy a proximity criterion with the target statistic () of the observed data (*x*) such that, where *d* expresses the metric distance between the data sets *x* and *y*, and *ϵ* represents a tolerance level. Corresponding parameter values are retained for simulated summaries (*s*) that are closer to the target statistic (t) than the tolerance (ϵ) (Sunnaker). *Figure 1* illustrates how the rejection ABC algorithm functions. The simulator (M) is run each time with a newly sampled parameter value (θ) from the prior distribution obtaining a simulated summary statistic which are the red and green dots. Based on the tolerance level (ϵ), a decision is made whether to retain the particular parameter value if the simulated summary statistic (red or green dot) is further from the target statistic () than (ϵ).

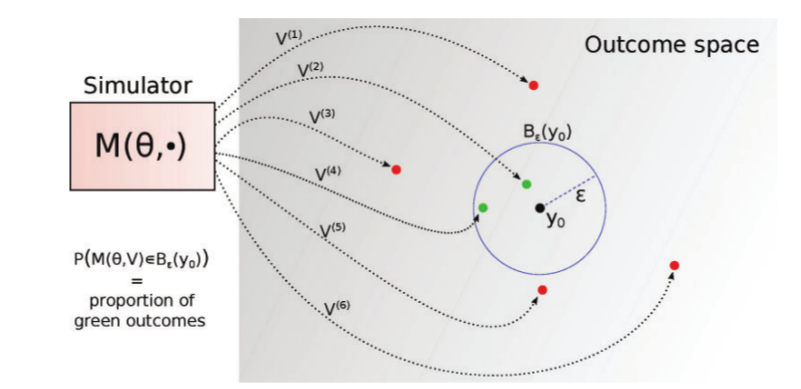


Figure 1 : Rejection ABC(Lintusaari)

The distribution of these retained parameter values is expected to assume posterior density without the explicit calculation of the likelihood. From Bayesian statistics, estimation of the posterior distribution depends on the prior distribution and the likelihood. The posterior is defined as

Where is a vector of parameters, *x* is the observed data, is the posterior distribution, is the likelihood and is the prior distribution. ABC techniques use this same knowledge in the approximation of the posterior. They use summary statistics of the observed data instead of the entire observed data. Thus, equation … is modified to approximate the posterior as follows

These summaries are compared to the summaries of the simulated data in order to approximate the posterior without performing explicit calculations of the likelihood since the likelihood is intractable.

1. **Sequential ABC**

Sequential ABC is a class of ABC methods which approximates the posterior progressively by drawing sequential samples from the prior (Lenormand). The prior for a particular sampling step depends on the previous retained sample except for the first sampling step which draws from the prior parameter space provided. Thus, the tolerance of the initial sampling step is less restrictive compared to the subsequent ones (Trevelyan). The sample at the current sampling step () is derived from the previous sample () using a decreasing sequence of tolerance levels. These methods determine by themselves the tolerance level to use at each sampling step and provides a stopping criterion. This choice of tolerance for the current sampling step is determined as a function of the metric distances simulated in the previous sampling step (McKinley). Figure 2 gives an illustration of how the sequential ABC algorithm works. The simulator (M) is run for the first time with newly sampled parameter values () selected from the prior distribution with tolerance obtaining a simulated sample. Having obtained the first retained sample of parameter values at a much bigger tolerance level, the simulator starts a second step of simulation with the first retained sample as the prior parameter space and a decreased tolerance. A second sample of parameter values is obtained at tolerance and this process if repeated until a stopping criterion is reached. At each sampling step, a decision is made whether to retain a particular parameter value or discard it. If a simulated summary statistic at that step is further from the target statistic () than the tolerance level () of that step, that particular parameter value is discarded, otherwise it is retained. The final sample approximates the posterior distribution.

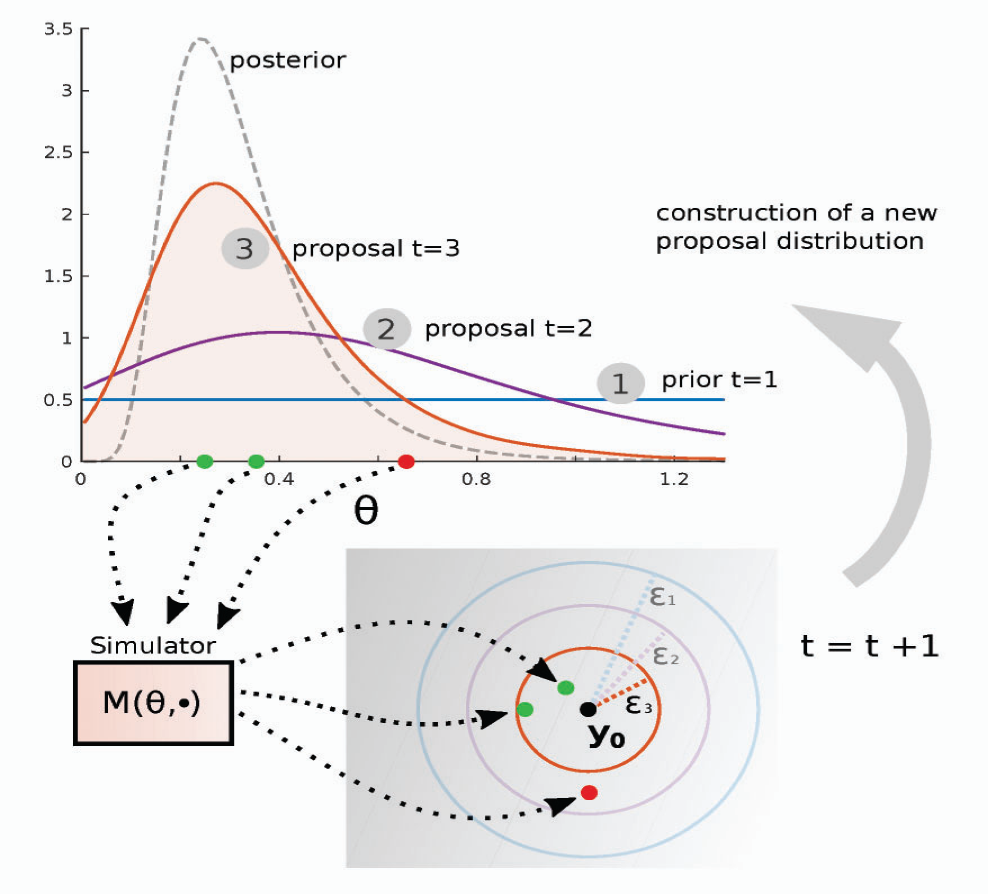


Figure 2 : Sequential ABC(Lintusaari)

1. **Bayesian Maximum Likelihood Estimation (BMLE)**

Data simulation with Bayesian Maximum Likelihood Estimation (BMLE)

Parameter combinations with high values of the likelihood are more consistent with the target supplied. This property allows the assessment of how the data supports one parameter combination compared with another.

**Steps**

1. Draw a large number of parameter sets from the prior distribution
2. For each parameter set, run the model and estimate model outcomes
3. Using these model outcomes, estimate the likelihood for the parameter set and retain this value (log likelihood)
4. Resample from the original parameter sample with replacement, using the likelihood values as sampling weights.

**THE SIMULATION MODEL**

**The SIR model**

The ABC algorithm requires an input model under which simulation is performed. In this study, the ABC methods performed simulations firstly under a simple SIR (Susceptible - Infected - Recovered) model which was used to generate datasets. The SIR model is an epidemiological model that computes the number of infected individuals with an infectious disease in a closed population over time. A closed population implies that the population size remains constant over time. This model involves a system of three non-linear ordinary differential equations (ODEs) that relates the number of susceptible, number of infected, and number of recovered individuals (Weiss). The SIR model functions under several assumptions such as: the population under study is homogeneous in nature (all individuals behave the same), the mode of transmission of the disease from infected to susceptible individuals is through direct contact between infected and susceptible individuals, recovered individuals gain permanent immunity to the disease. The closed population is usually grouped into compartments (health states) denoted by, and. The following system of ODEs governs the dynamics of the SIR model

Where is the disease transmission rate, is the recovery rate, is the duration of infection and the basic reproductive number. is the proportion of individuals in the population that are susceptible to the disease and represents the proportion of individuals that are infectious. Susceptible individuals become infectious at a rate. At a rate , infectious individuals recover from the disease (gain permanent immunity to the disease) (LStone).

The population stays constant throughout the transmission dynamics over the set time period such that

Figure 3 illustrates the dynamics of a stochastic SIR model run in the R software over a time period of 75 days for a population of 1000 individuals. The blue curve indicates the Susceptible compartment, the red curve indicates the Infected compartment and the green curve indicates the Recovered individuals. The susceptible compartment reduces to zero as the infected compartment gradually picks up.

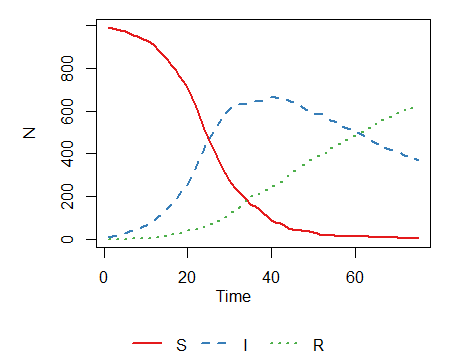


Figure 3: Plot of the stochastic SIR model

**Creating a raster**

A raster consists of a matrix of cells or pixels arranged into rows and columns to form a grid. Each cell contains a value which represents stored information. In order to compare the posterior densities of the two methods to the reference posterior density, we created a raster using the raster function from the raster library in the R software (R version 3.5.0 (2018-04-23)). The raster was created by considering the minimum and maximum values of beta and gamma accepted by both of the methods to be compared as well as the reference. This was done so that the same raster could be applied to both methods and the reference. The resulting parameter space was divided into equally sized bins with beta values on the x-axis and gamma values on the y-axis (see Figure *4*). This then formed a grid in which the posterior densities laid. We applied the grid to each posterior density in order to quantify the density of each cell or pixel.



Figure 4: 4x4 raster applied to a posterior density

**SIMULATION PROCEDURE**

1. **Obtaining targets**

A target is a data point from the observed data to be considered during the simulation procedure before a decision is finally taken, as to whether a certain parameter combination is to be accepted or discarded. As targets in this study, there were two scenarios. Scenario 1 considered two target features (prevalence at two time points, 50 and 75) and scenario 2 looked at three target features (considered the peak prevalence for each model run in addition to the two time points in scenario 1).

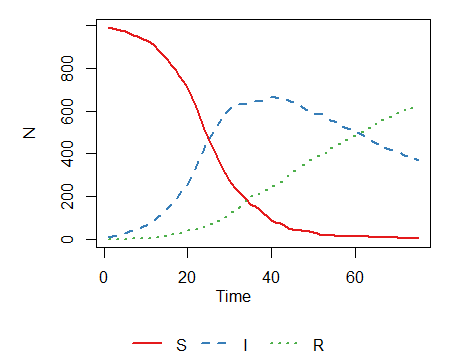


Figure 5: targets for scenarios 1 and 2. Scenario 1 considers only green points on the Infected (I) curve while scenario 2 considers both red and green points on the (I) curve.

R version 3.5.0 (2018-04-23) was used to perform the statistical analyses and datasets were obtained from a stochastic SIR model using the SIR function in the SimInf library (siminfref). To obtain targets for scenario 1, the SIR model was run one hundred times and for each run, the prevalence at times 50 and 75 were saved in a matrix. Targets for scenario 1 were then computed as the means of these saved prevalence at the two time points. Similarly for scenario 2, the SIR model was run one hundred times and for each run, the prevalence at times 50 and 75 as well as the peak prevalence were saved in a matrix. Targets for scenario 2 were then computed as the means of these three target features.

1. **Running Simulations for methods at equal time**

In order to set equal times for the simulations to run, the following steps were followed in each case

1. Run 10000 simulations with both methods and recorded the time each method took to run. We found out that Sequential ABC took longer to run the 10000 simulations compared to Rejection ABC.
2. Estimate the number of simulations Rejection ABC could run in Sequential time as
3. At tolerance of 1, run these estimated number of simulations using Rejection ABC and record the computation time.
4. Estimate the number of simulations run by Rejection ABC at the amount of time used by Sequential ABC as

This step fixes the time for the two methods

1. Calculate the tolerance for Rejection ABC such that the number of accepted parameter combinations is equal to that of sequential ABC as follows;
2. **Obtaining a Reference (“true posterior”)**

At a sufficiently small tolerance, rejection ABC approximates the true posterior (ref: Sunnaker et al). For each scenario, we run 500000 simulations and saved all parameter combinations (tolerance = 1). We then chose the tolerance as low as possible in order to retain the same number of parameter combinations accepted by the methods to be compared.

1. **Applying a raster**