

Bayesian Parametric Bootstrap for Models with Intractable Likelihoods

Brenda N. Vo^{*}, Christopher C. Drovandi[†], and Anthony N. Pettitt[‡]

Abstract. In this paper it is demonstrated how the Bayesian parametric bootstrap can be adapted to models with intractable likelihoods. The approach is most appealing when the computationally efficient semi-automatic approximate Bayesian computation (ABC) summary statistics are selected. The parametric bootstrap approximation is used to form a proposal distribution in ABC algorithms to improve the computational efficiency. The new approach is demonstrated through the sequential Monte Carlo and the ABC importance and rejection sampling algorithms. We found efficiency gains in two simulation studies, the univariate g-and-k quantile distribution, a toggle switch model in dynamic bionetworks, and in a stochastic model describing expanding melanoma cell colonies.

Keywords: Bayesian parametric bootstrap, approximate Bayesian computation, sequential Monte Carlo, melanoma cell spreading, agent-based model, quantile distribution.

1 Introduction

For many complex models in biological, medical and ecological sciences, the likelihood functions are not available in an analytical form and are computationally intractable, so computing the posterior distribution for these models is challenging. To overcome this limitation, approximate Bayesian computation (ABC), a class of Bayesian “likelihood-free” techniques, has emerged, which avoids direct evaluation of the likelihood through repeated simulation of data from the model. As such, ABC methods permit Bayesian inference for models with intractable likelihoods, when simulation from the model for a range of parameter values is feasible.

ABC methods have been successfully applied in a wide range of problems such as population genetics (Beaumont et al., 2002), infectious diseases (Drovandi and Pettitt, 2011c; Tanaka et al., 2006), astronomical model analysis (Cameron and Pettitt, 2012) and cell biology (Vo et al., 2015a,b). ABC is often implemented using one of the three main algorithms: importance and rejection sampling (ABC IS), Markov chain Monte Carlo (MCMC ABC) or sequential Monte Carlo (SMC ABC). In these ABC algorithms, a good proposal is crucial to improve the computational efficiency.

It has been shown that, for some cases, bootstrap methods are useful for numerical calculation of Bayes posterior distributions (Newton and Raftery, 1994; Efron, 2012;

^{*}School of Science and Technology, University of New England, Armidale, Australia, bvo3@une.edu.au

[†]ARC Centre of Excellence for Mathematical & Statistical Frontiers (ACEMS), School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, c.drovandi@qut.edu.au

[‡]ARC Centre of Excellence for Mathematical & Statistical Frontiers (ACEMS), School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, a.pettitt@qut.edu.au

Rubin, 1981). In particular, Efron (2012) proposed the use of a parametric bootstrap and a re-weighting scheme to approximate posterior distributions and their expectations. This approach is efficient and computationally straightforward. However, it depends upon an analytical expression for the sampling density of a statistic and a point estimate of the parameter, which generally requires the availability of the likelihood function. We show in this article that nevertheless parametric bootstrap samples can be used to construct a good proposal distribution in the context of ABC.

The main purpose of this paper is to present new and efficient ABC algorithms by introducing a good proposal distribution. The proposal distribution is constructed using the semi-automatic approach to ABC of Fearnhead and Prangle (2012) and the Bayesian parametric bootstrap (PB) of Efron (2012). The semi-automatic ABC is chosen because it provides summary statistics as point estimators, which are then used in the Bayesian PB process. In this paper, the resulting PB distributions, which require very few model simulations, are used as an initial importance distribution for SMC ABC algorithms and ABC IS. Hereafter, we refer to these new ABC algorithms as PB SMC ABC and PB ABC IS algorithms, respectively. We also note that, in some cases, the PB distribution can be a useful posterior approximation in its own right.

We first apply the methodology to simulated datasets from two examples: (i) the g-and-k quantile distribution of Rayner and MacGillivray (2002) and (ii) the toggle switch model of Bonassi et al. (2011) to validate the approach and compare it with the standard SMC ABC algorithm in Vo et al. (2015a).

We then apply the new collection of methods to a stochastic model that describes the expansion of human malignant melanoma cells (MM127) in a circular barrier assay (Vo et al., 2015a; Treloar et al., 2013). The cell expanding populations are governed by three parameters: cell motility, cell proliferation and cell-to-cell adhesion. Simulation of cell experiments from the stochastic model is highly computationally intensive, especially for a relatively large cell proliferation rate. Thus, inference requires an ABC algorithm that is efficient in terms of the number of model simulations.

This article is organized as follows. Background of ABC methodology is briefly reviewed in Section 2, followed by the Bayesian PB of Efron (2012) in Section 3. We demonstrate how the Bayesian PB can be efficiently adapted to likelihood-free problems and how to use this result to form a proposal for ABC algorithms in Section 4. Sections 5 and 6 contain the results from the two simulation studies, while Section 7 illustrates results from experimental data of human melanoma cells. The article is concluded with a discussion in Section 8.

2 Approximate Bayesian computation

Let $\theta \in \Theta$ be the parameter of a statistical model with a prior distribution $p(\theta)$ and an intractable likelihood, $p(\mathbf{y}|\theta)$, where \mathbf{y} is the observed data. ABC avoids likelihood evaluations by simulating data \mathbf{x} from the model and comparing it with the observed data. In ABC, direct comparison between the observed and the simulated datasets is often inefficient, especially when the data is high dimensional (Blum, 2010). Thus, we

consider a vector of summary statistics $\mathbf{s}(\cdot) = \{s_1(\cdot), \dots, s_d(\cdot)\}$ and denote $\mathbf{s}_{obs} = \mathbf{s}(\mathbf{y})$ and $\mathbf{s} = \mathbf{s}(\mathbf{x})$. To measure the closeness between \mathbf{x} and \mathbf{y} , via the closeness between \mathbf{s} and \mathbf{s}_{obs} , we use a distance metric $\rho(\mathbf{s}, \mathbf{s}_{obs})$ and a kernel weighting function $K_\epsilon(\rho(\mathbf{s}, \mathbf{s}_{obs}))$, where $\epsilon > 0$ is a bandwidth that is referred to as the ABC tolerance. The ABC posterior is constructed as

$$p_\epsilon(\theta|\mathbf{s}_{obs}) \propto p(\theta)p_\epsilon(\mathbf{s}_{obs}|\theta), \quad (1)$$

with

$$p_\epsilon(\mathbf{s}_{obs}|\theta) = \int K_\epsilon(\rho(\mathbf{s}_{obs}, \mathbf{s}))p(\mathbf{s}|\theta)d\mathbf{s}. \quad (2)$$

The general guidance for choosing the summary statistic is that it should be low dimensional but still incorporates most of the information contained in the data (Blum et al., 2013). In addition to the loss of information introduced by the use of an insufficient summary statistic, the ABC posterior is also affected by the tolerance ϵ , which is necessary to ensure computational feasibility. The smaller ϵ leads to a more accurate approximation in (2) but more computation. In practice, the kernel weighting function $K_\epsilon(\rho(\mathbf{s}_{obs}, \mathbf{s}))$ is often chosen as an indicator function, $\mathbf{1}_{\{\rho(\mathbf{s}_{obs}, \mathbf{s}) \leq \epsilon\}}$, that is unity if the condition involving the discrepancy is satisfied and is zero otherwise.

ABC algorithms for sampling from (1) generally consist of three major steps: sampling a proposed parameter θ^* , simulating data \mathbf{x} as per the observed data structure from the model $p(\cdot|\theta^*)$ and comparing \mathbf{x} with the observed data \mathbf{y} . Different ABC algorithms are distinguished by the process of sampling proposed parameters. Some of these are discussed next. Despite on-going research on developing more efficient ABC algorithms, often many hundreds of thousands or millions of model simulations are required.

We suggest that the efficiency of ABC algorithms can be improved if there is a good analytical approximation to the posterior $p(\theta|\mathbf{s}_{obs})$ that can be obtained quickly. For example, such an approximation can be used to form importance distributions for importance sampling (IS) or sequential Monte Carlo (SMC) based algorithms. In Section 3, we describe an adaptation of the Bayesian parametric bootstrap that can be used to form this initial approximation. The remainder of this section briefly discusses the ABC IS algorithm (Fearnhead and Prangle, 2012) and the SMC ABC algorithm of Vo et al. (2015a).

2.1 ABC importance and rejection sampling

Fearnhead and Prangle (2012) provide an importance and rejection sampling implementation of ABC for which the output is a weighted sample of values from the ABC posterior distribution (Algorithm 1). For simplicity, we set up the acceptance-rejection step (line 6) using the indicator function $\mathbf{1}_{\{\rho(\mathbf{s}_{obs}, \mathbf{s}) \leq \epsilon\}}$.

In Algorithm 1, a proposed parameter θ^* is drawn from an importance distribution, $g(\theta)$. Each proposed value θ^* is assigned a weight proportional to $p(\theta^*)/g(\theta^*)$ if it produces simulated data that satisfies the discrepancy condition, otherwise its weight is zero. When $g(\theta) = p(\theta)$, this algorithm becomes ABC rejection sampling which is similar to the algorithm of Beaumont et al. (2002).

Algorithm 1: ABC importance and rejection sampling (ABC IS) (Fearnhead and Prangle, 2012).

```

1 Given observed data  $\mathbf{y}$ ,  $N > 0$ , summary statistics  $\mathbf{s}(\cdot)$ ; a proposal density  $g(\theta)$ ,
  with  $g(\theta) > 0$  when prior  $p(\theta) > 0$ ; a density kernel  $K(\cdot)$ , with  $\max\{K(\cdot)\} = 1$ 
  and a bandwidth  $\epsilon > 0$ 
2 compute  $\mathbf{s}_{obs} = \mathbf{s}(\mathbf{y})$ 
3 for  $j = 1$  to  $N$  do
4   | simulate  $\theta_i \sim g(\theta)$ 
5   | simulate  $\mathbf{x} \sim p(\cdot|\theta_i)$ , and calculate  $\mathbf{s} = \mathbf{s}(\mathbf{x})$ 
6   | with probability  $K\{(\mathbf{s} - \mathbf{s}_{obs})/\epsilon\}$  set  $W_i = p(\theta_i)/g(\theta_i)$ ; otherwise set  $W_i = 0$ 
7 end
```

The advantage of this algorithm is that it generates independent samples and the algorithm can easily be run in parallel. However, if a good importance distribution is not available and the prior distribution is substantially different from the posterior, this algorithm results in low acceptance rates and is computationally inefficient.

2.2 SMC ABC

In this paper, we use the SMC ABC algorithm of Vo et al. (2015a) (Algorithm 2), which, for the motivating application in cell biology, was shown to have improvements over the algorithms of Sisson et al. (2009); Beaumont et al. (2009); Drovandi and Pettitt (2011a). For a non-increasing sequence of tolerances $\{\epsilon_t\}_{t=1}^T$, the SMC ABC algorithm aims to obtain a set of N weighted particles from the following sequence of targets

$$p_{\epsilon_t}(\theta, \mathbf{s}|\mathbf{s}_{obs}) \propto p(\theta)p(\mathbf{s}|\theta)\mathbf{1}_{\{\rho(\mathbf{s}, \mathbf{s}_{obs}) \leq \epsilon_t\}}.$$

In brief, the SMC ABC algorithm of Vo et al. (2015a) integrates the advantages of automatically determining tolerance values from Drovandi and Pettitt (2011a), Del Moral et al. (2012), and the advantage of geometric sampling from a proposal distribution until an acceptable parameter value is obtained (Sisson et al., 2009; Beaumont et al., 2009). In this SMC ABC algorithm, $\mathcal{N}(\cdot, \cdot)$ denotes the multivariate normal distribution, and $N_\alpha = \lfloor \alpha N \rfloor$ is the number of particles to keep at each iteration among the N particles, with $\alpha \in [0, 1]$. The stopping criterion is either the minimal acceptance rate, $p_{acc_{min}}$, or a target tolerance, ϵ_T .

For many applications of ABC, the most computationally intensive procedure is the model simulation process. Therefore, we aim to develop efficient ABC algorithms that can achieve a low tolerance value within a manageable number of model simulations. To achieve this, we incorporate an importance distribution at the initial iteration, $t = 0$, of the SMC ABC algorithm. Section 3 will describe how to obtain such an importance distribution while the detail of the algorithms will be provided in Section 4.

Algorithm 2: SMC ABC algorithm of Vo et al. (2015a).

```

1 Given  $N, N_\alpha, \mathbf{s}_{obs} = \mathbf{s}(\mathbf{y}), p_{acc_{min}}, \epsilon_T$ .
2 set  $p_{acc} = 1, t = 0$ 
3 for  $i = 1$  to  $N$  do
4   simulate  $\theta_i^{(t)} \sim p(\theta)$  and  $\mathbf{x} \sim p(\cdot | \theta_i^{(t)})$ 
5   compute  $\mathbf{s} = \mathbf{s}(\mathbf{x}), \rho_i^{(t)} = \rho(\mathbf{s}_{obs}, \mathbf{s}), w_i^{(t)} = \frac{1}{N}$ 
6 end
7 compute  $\epsilon^{(t)} = \max_{i=1, \dots, N} \{\rho_i^{(t)}\}$ 
8 while  $(p_{acc} > p_{acc_{min}})$  and  $(\epsilon^{(t)} > \epsilon_T)$  do
9   sort the particle set  $(\theta_i^{(t)}, \rho_i^{(t)})_{i=1}^N$  by  $\rho_i^{(t)}$ 
10  normalise the weights  $W_i^{(t)} = w_i^{(t)} / \sum_{j=1}^{N_\alpha} w_j^{(t)}, i = 1, \dots, N_\alpha$ 
11  set  $\Sigma_t$  as twice the weighted empirical covariance using  $(\theta_i^{(t)}, W_i^{(t)})_{i=1}^{N_\alpha}$ 
12  set  $\epsilon^{(t)} = \rho_{N-N_\alpha}^{(t)}$  and the number of trials,  $N_{trials} = 0$ 
13  for  $i = N_\alpha + 1$  to  $N$  do
14    while  $\rho_i^{(t)} > \epsilon^{(t)}$  do
15      resample  $\theta_i^*$  from  $(\theta_j^{(t)}, W_j^{(t)})_{j=1}^{N_\alpha}$ 
16      generate  $\theta_i^{(t)} | \theta_i^* \sim \mathcal{N}(\theta_i^*, \Sigma_t)$  and simulate  $\mathbf{x} \sim p(\cdot | \theta_i^{(t)})$ 
17      compute  $\mathbf{s} = \mathbf{s}(\mathbf{x}), \rho_i^{(t)} = \rho(\mathbf{s}_{obs}, \mathbf{s})$ 
18       $N_{trials} = N_{trials} + 1$ 
19    end
20    set  $w_i^{(t)} = \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N_\alpha} W_j^{(t)} \mathcal{N}(\theta_i^{(t)}; \theta_j^{(t)}, \Sigma_t)}$ 
21  end
22  set  $p_{acc} = \frac{N-N_\alpha}{N_{trials}}$ 
23  normalise the weights  $W_i^{(t)} = w_i^{(t)} / \sum_{j=N_\alpha+1}^N w_j^{(t)}, i = N_\alpha + 1, \dots, N$ .
24  set  $w_i^{(t+1)} = \frac{N_\alpha}{N} W_i^{(t)}, i = 1, \dots, N_\alpha$  and  $w_i^{(t+1)} = \frac{N-N_\alpha}{N} W_i^{(t)},$ 
    $i = N_\alpha + 1, \dots, N$ 
25  set  $t = t + 1$ 
26 end

```

2.3 Summary statistics and discrepancy function

In the literature, various approaches have been proposed to choose useful summary statistics including a sequential scheme based on the principle of approximate sufficiency (Joyce and Marjoram, 2008), partial least-squares regression (Wegmann et al., 2009), indirect inference (Drovandi et al., 2011) and machine learning methods (Aeschbacher et al., 2012). In this paper, we implement the method proposed by Fearnhead and Prangle (2012) who use the estimates of the posterior means of θ as the summary statistics. These posterior means are obtained via regression.

Initially, M draws of $\{\theta_i\}_{i=1}^M$ are made from the prior distribution. If the prior $p(\theta)$ is diffuse then draws of θ_i can be restricted to regions of non-negligible posterior density found using a pilot run of ABC. Each parameter θ_i is then used to simulate a dataset \mathbf{x}_i from the model, $\mathbf{x}_i \sim p(\cdot|\theta_i)$, $i = 1, \dots, M$. We fit the regression model

$$\theta_i = \alpha + \beta^T f(\mathbf{x}_i) + \xi_i, \quad i = 1, \dots, M, \quad (3)$$

with zero mean error ξ_i and $f(\cdot)$ is a vector-valued function of the data (or a set of summary statistics if using the full data is not feasible). Different choices of $f(\cdot)$ could be considered to obtain a better fit in the regression. In this paper, to find the best regression model, we employ a stepwise (bidirectional) regression method and the Bayesian information criterion (BIC) for model selection.

The expected value of θ_i given the simulated data \mathbf{x}_i , $E[\theta_i|\mathbf{x}_i]$, is then estimated by $\hat{\alpha} + \hat{\beta}^T f(\mathbf{x}_i)$, $i = 1, \dots, M$, where the intercept parameter $\hat{\alpha}$ and the vector of regression coefficients $\hat{\beta}$ is estimated from the best regression model. The derived summary statistic, $\mathbf{s}^{FP}(\mathbf{y}) = \hat{\alpha} + \hat{\beta}^T f(\mathbf{y})$, is then interpreted as the estimated posterior mean of θ obtained from the regression procedure. Thus, using this dimension reduction approach, we have only one summary statistic per parameter. In practice, if the parameter θ is vector valued, then a multiple linear regression model (3) is fitted to each component of θ in turn, with possibly a different function $f(\cdot)$ and different estimates $\hat{\alpha}$ and $\hat{\beta}$.

The derived summary statistics \mathbf{s}^{FP} can have different scales and correlations between summaries. Thus, we consider the Mahalanobis distance to compare the summary statistics of the observed and the simulated data. This discrepancy function is given by

$$\rho(\mathbf{y}, \mathbf{x}) = (\mathbf{s}^{FP}(\mathbf{y}) - \mathbf{s}^{FP}(\mathbf{x}))^T \times W^{-1} \times (\mathbf{s}^{FP}(\mathbf{y}) - \mathbf{s}^{FP}(\mathbf{x})),$$

where W is an estimate of the covariance matrix of the summary statistics \mathbf{s}^{FP} . To estimate W , we generate 100 simulated datasets $\{\mathbf{x}_i|\hat{\theta}\}_{i=1}^{100}$, using our point estimate $\hat{\theta} = \mathbf{s}^{FP}(\mathbf{y})$, obtained from the regression step above. For each simulated dataset \mathbf{x}_i , we compute the vector of summary statistics $\mathbf{s}_i^{FP} = \mathbf{s}^{FP}(\mathbf{x}_i)$. W is subsequently estimated by $\text{cov}(\{\mathbf{s}_i^{FP}\}_{i=1}^{100})$.

3 Bayesian parametric bootstrap

In this section, the summary statistics $\mathbf{s}(\cdot)$ are assumed to be an estimator of θ . Given an observed data set, we can compute an estimate of θ , $\hat{\theta}$, as a function of \mathbf{s}_{obs} . For simplicity we denote $\hat{\theta} = \mathbf{s}_{obs}$.

The Bayesian PB of Efron (2012) independently generates B values of the statistic $\mathbf{s}_j = \mathbf{s}(\mathbf{x}_j)$, $j = 1, \dots, B$ where \mathbf{x}_j is a simulated data set from the model $p(\cdot|\hat{\theta})$. Each sample estimate of θ , $\theta_j = \mathbf{s}_j$, $j = 1, \dots, B$, is a PB replication of $\hat{\theta}$. By re-weighting these points with an importance weight

$$W_j \propto \frac{[p(\theta)]_{\theta=\mathbf{s}_j} [p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_{obs}, \theta=\mathbf{s}_j}}{[p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_j, \theta=\mathbf{s}_{obs}}}, \quad (4)$$

we obtain a weighted sample from the posterior distribution of θ given $\hat{\theta}$ (Efron, 2012). If the likelihood function of the summary statistics $p(\mathbf{s}|\theta)$ can be evaluated then the importance weights (4) can be found. However, for models with intractable likelihoods, $p(\mathbf{s}|\theta)$ cannot be evaluated.

We consider a special case where the weights in (4) can be simplified. If the likelihood for \mathbf{s} is symmetric in $\mathbf{s}-\theta$ (\mathbf{s} and θ must be the same dimension), there exists a symmetric density h such that $h(x) = h(-x)$ for all x . Denote $p(\mathbf{s}|\theta) = h(\mathbf{s}-\theta)$, then the bootstrap provides values of

$$\begin{aligned} [p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_j, \theta=\mathbf{s}_{obs}} &= [h(\mathbf{s}-\theta)]_{\mathbf{s}=\mathbf{s}_j, \theta=\mathbf{s}_{obs}} \\ &= [h(\theta-\mathbf{s})]_{\mathbf{s}=\mathbf{s}_j, \theta=\mathbf{s}_{obs}} \\ &= [p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_{obs}, \theta=\mathbf{s}_j}, \end{aligned} \quad (5)$$

for $j = 1, \dots, B$. Therefore, the importance weights for the posterior become just the prior evaluated at the bootstrap replication \mathbf{s}_j

$$W_j \propto [p(\theta)]_{\theta=\mathbf{s}_j}, \quad j = 1, \dots, B. \quad (6)$$

Without the assumption of exact symmetry, the PB sample may give a poor approximation to the likelihood $[p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_{obs}, \theta=\mathbf{s}_j}$, $j = 1, \dots, B$. The weighted samples $\{\theta_j, W_j\}_{j=1}^B$ derived from (6), where $\theta_j = \mathbf{s}_j$, $j = 1, \dots, B$, may give a poor approximation to the posterior, $p(\theta|\mathbf{s}_{obs})$.

4 Coupling Bayesian parametric bootstrap with ABC

This section demonstrates two main innovations: how to obtain the PB distribution for models with intractable likelihoods and how to incorporate this PB distribution in ABC algorithms to improve their efficiency.

To apply the Bayesian PB idea for models with intractable likelihoods, it is computationally too intensive to take $\hat{\theta}$ as a point estimate of the ABC posterior $p_\epsilon(\theta|\mathbf{s}_{obs})$ obtained from the standard ABC IS or SMC ABC (Algorithms 1 and 2 above). The main idea here is to perform the Bayesian PB with $\hat{\theta}$ derived from the semi-automatic approach of Fearnhead and Prangle (2012) since it is very fast to obtain.

Sampling simulated data \mathbf{x} for ABC requires different values of θ while obtaining the PB distribution only requires sampling \mathbf{x} for fixed $\hat{\theta} = \mathbf{s}^{FP}(\mathbf{y})$ obtained from the regression approach in Section 2.3. Assuming that the likelihood for the summary statistic $p(\mathbf{s}|\theta)$ has the symmetry property so that the following holds

$$[p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_j, \theta=\mathbf{s}_{obs}} = [p(\mathbf{s}|\theta)]_{\mathbf{s}=\mathbf{s}_{obs}, \theta=\mathbf{s}_j}, \quad (7)$$

then the weighted samples $\{\theta_j, W_j\}_{j=1}^B$ give an approximation to the posterior $p(\theta|\mathbf{s}_{obs})$. Here $\theta_j = \mathbf{s}_j$ and the importance weights W_j are proportional to the prior density (6). This approximation is extremely computationally efficient having used only $(N_{pilot} + M + B)$ simulations from the model $p(\cdot|\theta)$. Here, N_{pilot} is the number of model simulations from the ABC pilot run. Pseudo code to perform the Bayesian parametric

Algorithm 3: Likelihood-free Bayesian PB algorithm.

```

1 Given observed data  $\mathbf{y}$ , prior distribution  $p(\theta)$  and integers  $M, B > 0$ 
2 Optional: Perform an ABC pilot run to obtain a training region of  $\theta$ 
3 Generate  $M$  synthetic data sets for the regression: Simulate  $\theta_i$  from the
  truncated region as appropriate, and generate  $\mathbf{x}_i \sim p(\cdot|\theta_i)$ ,  $i = 1, \dots, M$ 
4 Perform a regression analysis:  $\theta_i = \alpha + \beta^T f(\mathbf{x}_i) + \epsilon_i$ ,  $i = 1, \dots, M$ , for each
  component in  $\theta$ 
5 Compute the point estimate  $\hat{\theta} = \hat{\alpha} + \hat{\beta}^T f(\mathbf{y})$ , for each component in  $\theta$ 
6 for  $j = 1$  to  $B$  do
7   | Simulate  $\mathbf{x}_j \sim p(\cdot|\hat{\theta})$ 
8   | Compute the bootstrap value,  $\theta_j = \hat{\alpha} + \hat{\beta}^T f(\mathbf{x}_j)$  for each component in  $\theta$ 
9   | Compute the weight  $W_j \propto p(\theta_j)$ 
10 end

```

bootstrap in this section is provided in Algorithm 3. For the ABC pilot run, one could adopt any ABC algorithm.

The appeal of likelihood-free Bayesian PB is that once the regression model has been fitted, all the samples generated through the PB approach are kept. Li and Fearnhead (2016b) demonstrated that in standard ABC the ABC tolerance must reduce with an increase to the size of the sample, so that low acceptance rates are still required even in large sample settings to obtain estimators with good asymptotic properties. One exception as determined by Li and Fearnhead (2016a) is that when a good proposal distribution is used we are able to accept more of the proposals as the sample size increases if a post-hoc regression adjustment is applied. However, in finite sample scenarios that we are likely to encounter in practice, ABC methods typically suffer from very low acceptance rates. We show in the Supplementary Material (Vo et al., 2019) that the standard ABC IS in Algorithm 1 would require 70 times more model simulations than the PB ABC IS algorithm when a good proposal obtained from the PB distribution is used. Thus, the Bayesian PB approach in Algorithm 3 will be significantly more computationally efficient than standard ABC algorithms, however the PB distribution is likely to differ in some way to the standard ABC posterior. Furthermore, Li and Fearnhead (2016b) show that the asymptotic efficiency of ABC does at least improve with a good proposal distribution. Therefore, we consider harnessing the PB ABC approximation, which can be obtained quickly, to form a proposal distribution for ABC algorithms.

We denote an analytic approximation of the weighted sample $\{\theta_j, W_j\}_{j=1}^B$ resulting from the Bayesian PB algorithm above as $g(\theta)$, which can be taken as a parametric distribution such as a multivariate normal or a kernel density estimate. It is important that the tails of the proposal distribution are thicker than that of the target distribution (for an argument of this in the ABC context, see Li and Fearnhead (2016b)). This can be achieved by extending the multivariate normal proposal to a multivariate t by choosing a small number of degrees of freedom or inflating the variance of the multivariate normal

Algorithm 4: PB SMC ABC algorithm. SMC ABC algorithm with initial proposal density $g(\theta)$. In this paper we form $g(\theta)$ using the PB method, creating the PB SMC ABC algorithm.

```

1  Given  $N, N_\alpha, p_{acc_{min}}, \epsilon_T$ , a summary statistic function  $\mathbf{s}(\cdot)$  and  $\mathbf{s}_{obs} = \mathbf{s}(\mathbf{y})$ .
2  Obtain the Bayesian PB distribution,  $g(\theta)$ , as described in Algorithm 3
3  Set  $p_{acc} = 1, t = 0$ 
4  for  $i = 1$  to  $N$  do
5      Simulate  $\theta_i^{(t)} \sim g(\theta)$  and  $\mathbf{x} \sim p(\cdot | \theta_i^{(t)})$ 
6      compute  $\mathbf{s} = \mathbf{s}(\mathbf{x}), \rho_i^{(t)} = \rho(\mathbf{s}_{obs}, \mathbf{s})$ 
7       $w_i^{(t)} = \frac{\pi(\theta_i)}{g(\theta_i)}$ 
8  end
9   $\epsilon^{(t)} = \max_{i=1, \dots, N} \{\rho_i^{(t)}\}$ 
10 while  $(p_{acc} > p_{acc_{min}})$  and  $(\epsilon^{(t)} > \epsilon_T)$  do
11     Sort the particle set  $(\theta_i^{(t)}, \rho_i^{(t)})_{i=1}^N$  by  $\rho_i^{(t)}$ , such that  $\rho_1^{(t)} \leq \rho_2^{(t)} \leq \dots \leq \rho_N^{(t)}$ 
12     Normalise the weights  $W_i^{(t)} = w_i^{(t)} / \sum_{j=1}^{N_\alpha} w_j^{(t)}$  for  $i = 1, \dots, N_\alpha$ 
13     Set  $\Sigma_t$  as twice the weighted empirical covariance using  $(\theta_i^{(t)}, W_i^{(t)})_{i=1}^{N_\alpha}$ 
14     Set  $\epsilon^{(t)} = \rho_{N-N_\alpha}^{(t)}$  and the number of trials,  $N_{trials} = 0$ 
15     for  $i = N_\alpha + 1$  to  $N$  do
16         while  $\rho_i^{(t)} > \epsilon^{(t)}$  do
17             Draw  $\theta_i^*$  from  $(\theta_j^{(t)}, W_j^{(t)})_{j=1}^{N_\alpha}$ 
18             Generate  $\theta_i^{(t)} | \theta_i^* \sim \mathcal{N}(\theta_i^*, \Sigma_t)$  and simulate  $\mathbf{x} \sim p(\cdot | \theta_i^{(t)})$ 
19             Compute  $\mathbf{s} = \mathbf{s}(\mathbf{x}), \rho_i^{(t)} = \rho(\mathbf{s}_{obs}, \mathbf{s})$ 
20              $N_{trials} = N_{trials} + 1$ 
21         end
22         Set  $w_i^{(t)} = \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N_\alpha} W_j^{(t)} \mathcal{N}(\theta_i^{(t)}; \theta_j^{(t)}, \Sigma_t)}$ 
23     end
24     Set  $p_{acc} = \frac{N - N_\alpha}{N_{trials}}$ 
25     Normalise the weights  $W_i^{(t)} = w_i^{(t)} / \sum_{j=N_\alpha+1}^N w_j^{(t)}$  for  $i = N_\alpha + 1, \dots, N$ .
26     Set  $w_i^{(t+1)} = \frac{N_\alpha}{N} W_i^{(t)}$  for  $i = 1, \dots, N_\alpha$  and  $w_i^{(t+1)} = \frac{N - N_\alpha}{N} W_i^{(t)}$  for
        $i = N_\alpha + 1, \dots, N$ 
27     Set  $t = t + 1$ 
28 end

```

proposal. In this paper we choose the latter and scale the sample covariance matrix by a factor of 2. The approximation $g(\theta)$ is used as a proposal density in the ABC IS (see Algorithm 1, PB ABC IS) or an initial importance distribution for the SMC ABC algorithm (see Algorithm 4, PB SMC ABC), and we discuss other options in Section 8.

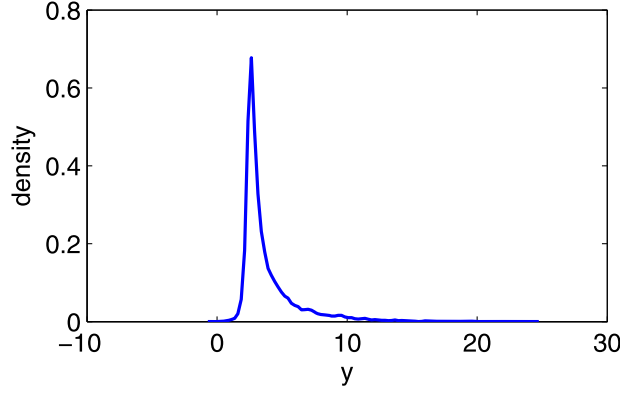


Figure 1: Estimated probability density based on the simulated dataset from the g-and-k distribution.

5 Quantile distribution

5.1 Model and data

We now validate our new collection of methods using synthetically generated data from the g-and-k quantile distribution (Rayner and MacGillivray, 2002). The g-and-k-distribution is a class of quantile distributions and it is defined by its quantile function, the inverse cumulative distribution function

$$Q(z(u); \theta) = F^{-1}(z(u); \theta) = a + b \left(1 + c \frac{1 - \exp(-gz(u))}{1 + \exp(-gz(u))} \right) (1 + z(u)^2)^k z(u), \quad (8)$$

where $z(u)$ is the u -quantile of the standard normal distribution and $\theta = (a, b, c, g, k)$ is the unknown parameter. Given a fixed value of c , $c = 0.8$ (Rayner and MacGillivray, 2002), the g-and-k distribution consists of four unknown parameters, a, b, g and k , which are related to location, scale, skewness and kurtosis, respectively. The simulated dataset consists of $n = 10^4$ independent draws from the g-and-k distribution with parameters $\theta = (a, b, g, k) = (3, 1, 2, 0.5)$. A uniform prior $(0, 10)^4$ is used for the parameters. This is similar to the example used in Fearnhead and Prangle (2012); Drovandi and Pettitt (2011b); Allingham et al. (2009). A plot of the estimated probability density function based on this dataset is shown in Figure 1. The data shows significant skewness and kurtosis.

5.2 Results

For the ABC pilot run, we use the SMC ABC approach in Algorithm 2, the set of octiles as summary statistics (Drovandi and Pettitt, 2011b) with the Euclidean distance between summary statistics and set $N = 1,000$. After 18 iterations, we find that the training regions for a, b, g and k are given by $(2.8, 3.2)$, $(0.7, 1.3)$, $(1, 4)$ and $(0, 1)$,

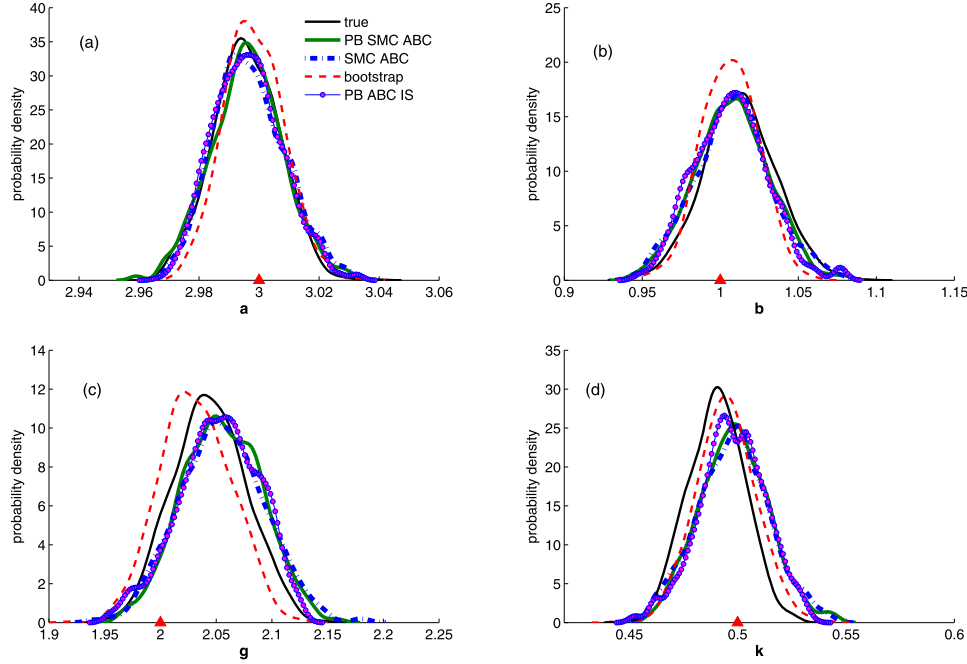


Figure 2: Posterior distributions for the parameters of the g-and-k simulated dataset. In all subfigures, results from using the Bayesian PB (dashed red), the PB SMC ABC (solid green) and the plain SMC ABC (dotted blue) are shown. The true values of a , b , g and k are plotted as red triangles.

respectively. The number of model simulations for the pilot run is 25,012 and the probability of acceptance in the last iteration is approximately 27%.

For the regression procedure, we simulate $M = 5,000$ datasets from the parameters that are drawn from the training regions above. For the covariates in the regression model (3), we consider a set of summary statistics, $\{L_i\}_{i=1}^{19}$, where L_i , $i = 1, \dots, 19$ is the $(0.05 \times i)$ th quantile, rather than the entire dataset. A bidirectional stepwise regression is then fitted to determine a 4-dimensional summary statistic \mathbf{s}^{FP} . The point estimate $\hat{\theta}$ obtained from the regression is $\hat{\theta} = (\hat{a}, \hat{b}, \hat{g}, \hat{k}) = (2.9970, 1.0064, 2.0426, 0.4965)$. It should be noted that the summary statistics $\{L_i\}_{i=1}^{19}$ are highly informative for all parameters a, b, g, k , since the adjusted R-square values of the regressions for the four parameters are 85.1%, 91.5%, 92.5% and 95.6%, respectively.

Using this value of $\hat{\theta}$, we perform the Bayesian PB with $B = 1,000$ (Algorithm 3). The densities of the Bayesian PB samples are plotted in Figure 2 (dashed red). We fit a multivariate normal distribution to the PB samples and use it as an initial importance distribution, $g(\theta)$, in the PB SMC ABC algorithm. In order to help ensure coverage of the tails, the covariance matrix of $g(\theta)$ is set as twice the empirical covariance matrix based on the PB samples. The ABC posterior distributions of a, b, g and k from the

PB SMC ABC algorithm are plotted in Figure 2 (solid green) together with the results from using the standard SMC ABC algorithm of Vo et al. (2015a) (dotted blue).

It should be noted that, in the standard SMC ABC approach, we use an initial importance distribution formed from the pilot, a multivariate normal distribution with covariance matrix estimated from the pilot run samples inflated by a factor of two, rather than the prior distribution to avoid the extrapolation issue in the regression procedure (Fearnhead and Prangle, 2012).

For all four parameters, the posteriors resulting from standard SMC ABC and PB SMC ABC are well-defined and quite similar, as expected. However, PB SMC ABC required 35% fewer model simulations, about 303,000 simulations less than the standard SMC ABC algorithm. The PB distributions are also very close to the ABC posteriors given a very small number of model simulations (31,012).

In both algorithms, we set $N = 2,000$ and a tolerance $\epsilon = 0.78$. The effective sample size, ESS, for the PB SMC ABC and the standard SMC ABC algorithms are 1,413 and 1,390, respectively. We also compare ABC posterior distributions with the true posteriors obtained from the numerical evaluation of the likelihood function (Rayner and MacGillivray, 2002) (results not shown). The results for a and b are very accurate, suggesting that the summary statistics \mathbf{s}^{FP} are close to sufficient for these parameters. The results for g and k obtained from the two ABC algorithms show slight deviation from the true posteriors and also a small loss of precision.

6 Toggle switch model

6.1 Model

The performance of the new methods was also investigated on a toggle switch model describing gene expressions. Details of the model and its background were previously described in Gardner et al. (2000); Bonassi et al. (2011). Motivated from a study of dynamic cellular networks using E.coli bacterial cells (Gardner et al., 2000), Bonassi et al. (2011) proposed a discrete time stochastic model to capture the behaviour of a network with two genes in a synthetic toggle switch design. In brief, for each cell c , $c = 1, \dots, 2000$, we let $u_{c,t}$ and $v_{c,t}$ be the expressions of genes u and v for cell c at time t , respectively. Given an initial state $(u_{c,0}, v_{c,0})$ and a small discrete time step h , the states of $u_{c,t}$ and $v_{c,t}$ are updated using the equations below (Bonassi et al., 2011):

$$\begin{aligned} u_{c,t+h} &= u_{c,t} + h\alpha_u/(1 + v_{c,t}^{\beta_u}) - h(1 + 0.03u_{c,t}) + 0.5h\xi_{c,u,t}, \\ v_{c,t+h} &= v_{c,t} + h\alpha_v/(1 + u_{c,t}^{\beta_v}) - h(1 + 0.03v_{c,t}) + 0.5h\xi_{c,v,t}, \end{aligned} \quad (9)$$

where $t = 0, \dots, T$, with T is the assumed time to reach steady-state. The terms $\xi_{c,u,t}$ and $\xi_{c,v,t}$ are independent standard normal random variables and they represent the intrinsic noise within cell c .

For a particular cell c , the stochastic process is only observed for gene u , at the steady-state time T . Let y_c be the noisy measurements of $u_{c,T}$, which is modelled by Bonassi et al. (2011) as follows:

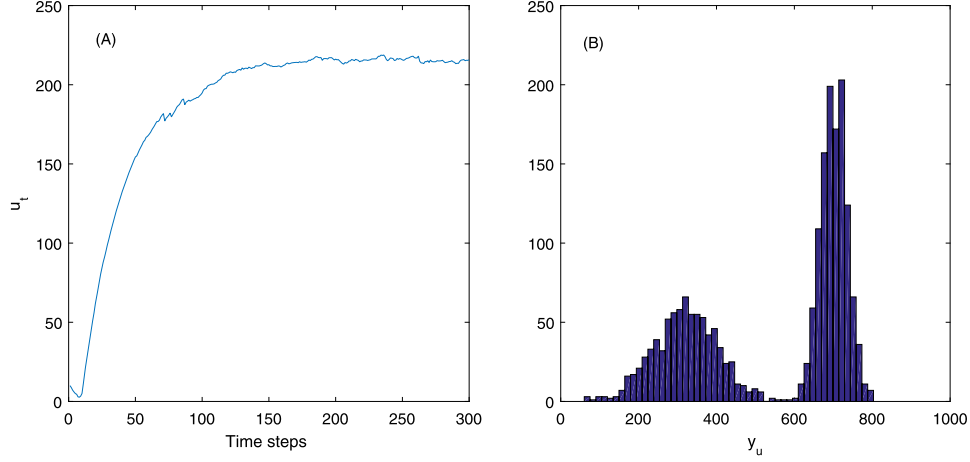


Figure 3: Simulated data from the toggle switch model, using the parameters value $\theta = (\mu, \sigma, \gamma, \alpha_u, \beta_u, \alpha_v, \beta_v) = (320, 0.25, 0.15, 25, 4, 15, 4)$. Subfigure (A) demonstrates a time series of $u_{c,t}$ for a single cell c , over $T = 300$ time steps. Subfigure (B) shows the histogram of the synthetic dataset used in this example.

$$y_c = u_{c,T} + \mu + \mu\sigma\eta_c/u_{c,T}^\gamma, \quad \text{for } c = 1, \dots, 2000, \quad (10)$$

where the measurement errors η are drawn from a standard normal distribution. The model consists of seven unknown parameters, which are denoted as $\theta = (\mu, \sigma, \gamma, \alpha_u, \beta_u, \alpha_v, \beta_v)$. For all model simulations, we use $h = 1$ and $T = 300$ time steps as suggested in Bonassi et al. (2011). An example of the time series data for $u_{c,t}$ over 300 time steps is shown in Figure 3A.

In this example, we illustrate our collection of methods using a synthetic dataset where $\theta = (320, 0.25, 0.15, 25, 4, 15, 4)$. The histogram of the synthetic dataset, which consists of the value of $u_{c,T}$ for $c = 1, \dots, 2000$ cells, is given in Figure 3B. A characteristic of the histogram is that the data appear to be well represented by a mixture of two unimodal densities. For the ABC analysis, we use the same uniform priors as in Bonassi et al. (2011): $\mu \sim U(250, 400)$; $\sigma \sim U(0.05, 0.5)$; $\gamma \sim U(0.05, 0.35)$; $\alpha_u \sim U(0, 50)$; $\beta_u \sim U(0, 7)$; $\alpha_v \sim U(0, 50)$; $\beta_v \sim U(0, 7)$.

6.2 Results

For the pilot run, we use the SMC ABC in Algorithm 2 with $N = 200$ particles, the minimum probability of acceptance at 20% and a 19-dimensional summary statistic $\{L_i\}_{i=1}^{19}$, where L_i is the $(0.05 \times i)$ th quantile, for $i = 1, \dots, 19$. To perform the regression procedure, we simulate $M=10,000$ datasets from the fitted multivariate normal distribution of the resulting posterior distributions from the pilot run. For each simulated dataset, we fit a Gaussian mixture model with three components and use the estimated means, standard deviations and the mixture proportions (after accounting

for label switching by ordering the means) for the regression input. This provides eight summary statistics, together with L_i and L_i^2 , for $i = 1, \dots, 19$, producing a vector of 46 variables which are then used as explanatory variables in the regression stage. We use a bidirectional stepwise regression and fit each parameter in turn.

The point estimate $\hat{\theta}$ obtained from the regression is $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma}, \hat{\alpha}_u, \hat{\beta}_u, \hat{\alpha}_v, \hat{\beta}_v) = (317.18, 0.25, 0.14, 24.86, 4.17, 22.90, 4.33)$. We found that the set of 46 explanatory variables above is only informative for μ, σ, γ and α_u , explaining about 90.5%, 91.9%, 96.9% and 94.9% of the variability of these parameters, respectively. However, for β_u, α_v and β_v , the point estimates from the regression models are very similar to the posterior means from the ABC pilot run and using the same explanatory variables only accounts for 34.2%, 33.7% and 30.7% of the variability in these three parameters, respectively. The low values of the adjusted R-square for these parameters has a strong effect in the Bayesian PB procedure. In particular, the PB distributions for these parameters have an extremely small variance. Thus, we suggest that the PB method is more suitable when the regression adjusted R-square is high, roughly 80% or above. In this case, for the PB procedure, we use the point estimate obtained from the regression for the first four parameters, and for the last three parameters, we discard the regression models and use the posterior means from the pilot run instead.

The resulting PB distributions (dashed red) with $B = 1,000$ samples are presented in Figure 4 together with the final ABC posterior distributions resulting from the PB SMC ABC algorithm (solid green) and the standard SMC ABC algorithm (dotted blue). It should be noted that, in the PB SMC ABC algorithm, we use the PB samples to form an importance distribution $g(\theta)$ for $\mu, \sigma, \gamma, \alpha_u$, while for the parameters β_u, α_v and β_v , the proposal is formed from the pilot run instead.

For the parameters μ, σ, γ and α_u , it can be seen that the ABC posterior distributions are well-defined and have much smaller variance compared to the results from Bonassi et al. (2011). The PB distributions also provide a good approximation to the marginal ABC posterior distributions, given a very small number, 18,068, of model simulations. Furthermore, the posteriors resulting from the standard SMC ABC and the PB SMC ABC algorithms are similar, as expected. Both the PB SMC ABC and the standard SMC ABC algorithms use the same summary statistics, final tolerance $\epsilon = 0.7$ and $N = 2,000$ particles. However, standard SMC ABC required 220,320 model simulations, while PB SMC ABC only used 165,210 model simulations, which is about 25% less model simulations. The reason is that, for four parameters μ, σ, γ and α_u , PB SMC ABC starts from the importance distribution $g(\theta)$ which is very close to the actual ABC posteriors, whereas standard SMC ABC starts from an importance distribution formed from the pilot run.

The results in Figure 4(e)–(g) indicate that we obtain far less information for β_u, α_v and β_v regardless of the tolerance values and the summary statistics used in the ABC analysis, indicating that the data may be not informative for these parameters. This is consistent with the results previously shown in Bonassi et al. (2011). To overcome this, one could incorporate additional information from the network such as including the information from both genes at the steady-state time T .

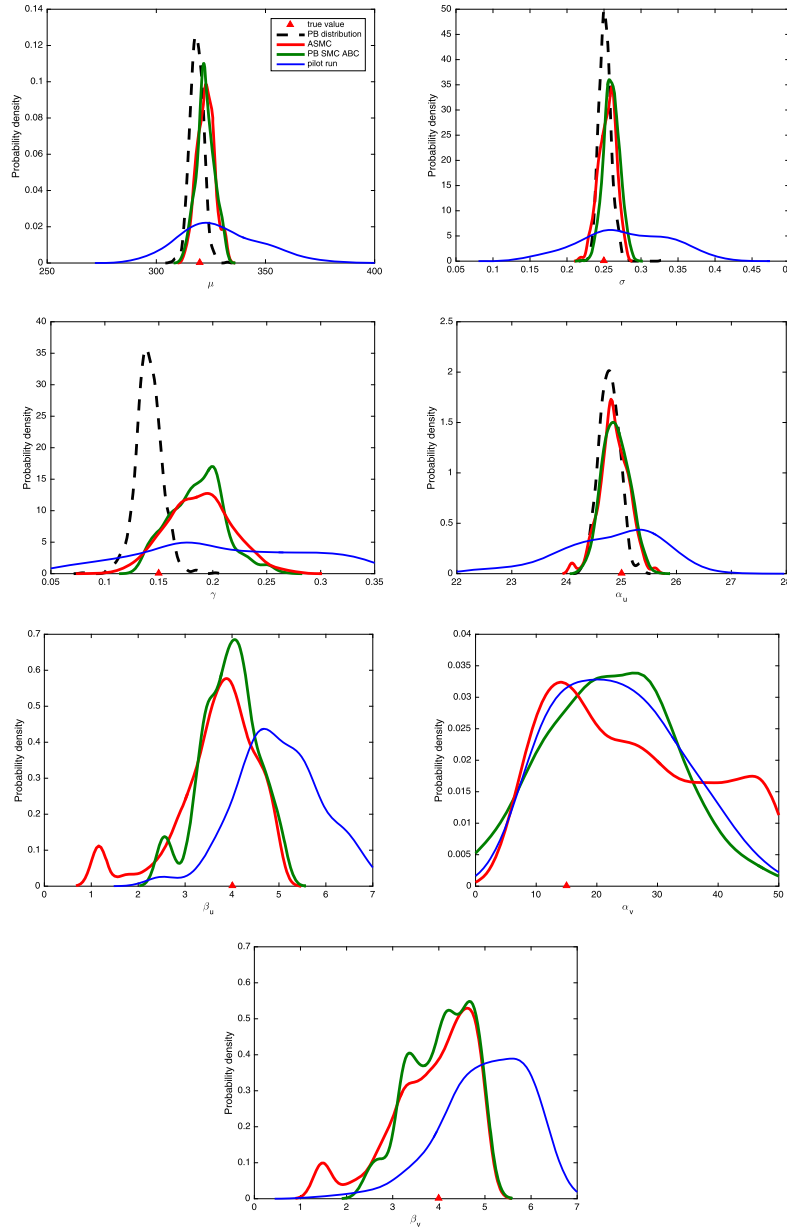


Figure 4: Posterior distributions for the seven parameters in the toggle switch model. Shown are the results from the PB distribution (dashed red), the SMC ABC algorithm of Vo et al. (2015a) (dotted blue) and the PB SMC ABC algorithm (solid green). The true values are plotted as red upper triangles. For the parameters β_u , α_v and β_v , the PB approximations have a very small variance since the regression models do not explain much of the variation in these parameter values. Because of this, only the point estimates obtained from the fitted regressions are shown as vertical lines for these parameters.

7 Application to a collective cell spreading model

We now present our main application involving a discrete stochastic model describing the expansion of melanoma cell populations (Vo et al., 2015a). Melanoma is a cancer that begins in the melanocytes and is the most dangerous form of skin cancer (Garbe et al., 2012). Melanoma is less common, approximately 5% of all skin cancer occurrences, but accounts for approximately 75% of skin cancer deaths (Australian Institute of Health and Welfare and Australasian Association of Cancer Registries, 2012). The spatial expansion of melanoma cells is governed by various mechanisms including cell motility, cell proliferation and cell-to-cell adhesion. Estimating these mechanisms can improve our understanding of melanoma biology and its response to treatment.

7.1 Data

We applied the new ABC algorithms to analyse an experiment of human malignant melanoma cells (MM127) (Pope et al., 1979; Whitehead and Little, 1973) in a circular barrier assay. Details of the experimental protocol were described in Treloar et al. (2013). In brief, the experiment was conducted using a 24-well tissue culture plate, where each well has a diameter of 15.6 mm. Initially, 20,000 cells were evenly distributed within a metal-silicone barrier, of a diameter 6.0 mm, which was located in the centre of the well. The tissue culture plate was kept for one hour to allow the cells to attach to the surface. Subsequently, the barrier was lifted and the plate was incubated for two time durations of 24 or 48 hours. The experiment was repeated in triplicate. For each experiment, we obtained two types of images: (i) a population-scale image which shows the entire melanoma cell colony and (ii) individual-scale images which show the location of each cell in the population. For the application in this paper, we only analyse the experiments that were terminated at 24 hours. Details of the ABC analyses for experiments at 48 hours and experiments with different initial cell densities can be found in Vo et al. (2015a).

Initially, we summarise the experimental data using a high dimensional summary statistics including three radii of the entire expanding melanoma colonies, $\{r_i\}_{i=1}^3$, the total number of cells, $\{c_i\}_{i=1}^6$, and the number of isolated cells, $\{p_i\}_{i=1}^6$, in six subregions of the cell population. We compute $\{r_i\}_{i=1}^3$ by locating the position of the leading edge, measuring the area of the spreading cell population and converting this area into an equivalent circular radius. We average the $\{c_i\}_{i=1}^6$ and $\{p_i\}_{i=1}^6$ over three replicates, to produce a total of 15 summary statistics. These processes were performed using a segmentation algorithm written with the Matlab Image Processing Toolbox (Vo et al., 2015a) and were repeated for images that were produced by the discrete model described in Section 7.2. For more details on the image analysis and how the summary statistics were obtained see Vo et al. (2015a). Table 1 shows 15 observed summary statistics that were used for the ABC analysis in this section.

7.2 Model

To describe the spatial expansion of the melanoma cell population, we use a discrete stochastic model that incorporates cell motility (unbiased random walk), cell prolif-

Summary statistics						
r_i (mm)	3.3136	3.3185	3.3265			
p_i (%)	12.26	11.79	12.65	11.20	11.04	10.17
c_i (cells)	446	435	410	429	444	438

Table 1: Initial summary statistics for the experimental data. Results shown include three radii, $\{r_i\}_{i=1}^3$, the percentage of isolated cells, $\{p_i\}_{i=1}^6$, and the total number of cells, $\{c_i\}_{i=1}^6$, in six subregions of the cell population (average over three replicates).

eration and cell-to-cell adhesion on a two-dimensional square lattice with spacing Δ (Treloar et al., 2013; Vo et al., 2015a,b). Each lattice site can be occupied by at most one agent. This model is also referred to as an agent based model or individual based model.

To simulate the experiments, we use a two-dimensional square lattice of size 867×867 , with lattice spacing $\Delta = 18 \mu\text{m}$, so that the width of the lattice corresponds to the diameter of the well, 15.6 mm ($15,600 \mu\text{m} / 18 \mu\text{m} = 867$). Let $C(t)$ be the number of agents in the discrete model at time t , $P_m \in [0, 1]$ be the probability that an isolated agent will attempt to step a distance Δ within a time step of duration τ , and let $P_p \in [0, 1]$ represent the probability that an agent will attempt to proliferate and deposit a daughter within a time step of duration τ . The strength of cell-to-cell adhesion is represented by $q \in [0, 1]$. To step from time t to time $t + \tau$, $C(t)$ agents are sampled, with replacement, and given the opportunity to move with probability $P_m \times (1 - q)^n$, where $0 \leq n \leq 4$ is the number of occupied nearest neighbour sites. If an agent is at position (x, y) and has an opportunity to move, it will attempt to step to either $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with each target site chosen with equal probability. For increasing values of q , neighbour agents adhere more tightly to each other and it is difficult for an agent to move away from its neighbours. A similar mechanism is employed for proliferation events. A proliferative agent at position (x, y) will attempt to deposit a daughter agent at $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with each target site chosen with equal probability.

In this model, the cell motility rate is quantified in terms of the cell diffusivity, D , $D = P_m \Delta^2 / 4\tau$, and the cell proliferation rate, λ , is related by $\lambda = P_p / \tau$ (Simpson et al., 2010). A uniform prior $U(0, 1)$ is used for all three parameters (P_m, q, P_p) . For all model simulations, we use a time step duration τ as 0.04 h (Vo et al., 2015a). We apply ABC algorithms to obtain joint posterior distributions for (P_m, q, P_p) , then use the values of Δ and τ to rescale posterior distributions of P_m and P_p into posterior distributions of the biological parameters of interested D and λ , respectively.

7.3 Parameter inferences

A pilot run was conducted using the SMC ABC algorithm of Vo et al. (2015a), incorporating all 15 summary statistics (Table 1), $\mathbf{s} = (\{r_i\}_{i=1}^3, \{p_i\}_{i=1}^6, \{c_i\}_{i=1}^6)$, and using the Mahalanobis distance to compute the distance between the observed and the simulated summary statistics. We set $N = 500$ particles and $p_{acc_{min}} = 0.2$ as the stopping criterion. We obtain the training regions for P_m , q and P_p as $(0.07, 0.14)$, $(0.14, 0.43)$ and $(0.0010, 0.0018)$, respectively, using only 13,925 model simulations.

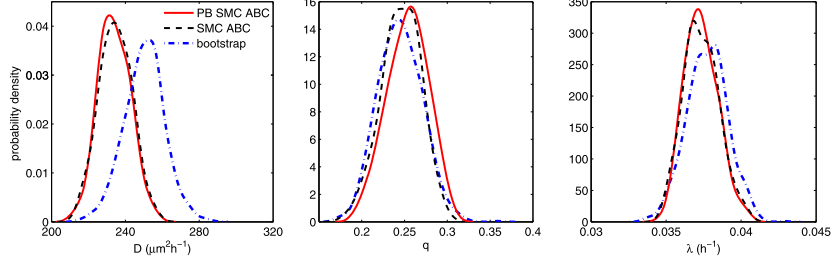


Figure 5: ABC posterior distributions for D , q and λ resulting from PB SMC ABC (solid green), standard SMC ABC of Vo et al. (2015a) (dotted blue) and the Bayesian PB distribution (dashed red).

		Bayesian PB	Standard SMC ABC	PB SMC ABC
D	$E[D]$ ($\mu\text{m}^2 \text{h}^{-1}$)	250 (242, 278)	235 (219, 248)	234 (221, 248)
	CV (%)	4.2	3.7	3.6
q	$E[q]$	0.24 (0.20, 0.29)	0.25 (0.21, 0.28)	0.25 (0.21, 0.29)
	CV (%)	10.5	9.2	9.1
λ	$E[\lambda] \times 10^{-2}$ (h^{-1})	3.77 (3.54, 4.01)	3.73 (3.55, 3.94)	3.73 (3.57, 3.92)
	CV (%)	3.6	3.1	2.9

Table 2: ABC posterior summary for D , q and λ for two different ABC algorithms and the bootstrap distribution. Results shown include the posterior mean (the 90% CI in the parentheses) and the coefficient of variation, CV.

A regression analysis (3) was performed for each parameter in turn, using $M = 5,000$ datasets that were generated by parameters in these training regions. We found that using a vector of 30 explanatory variables $\{\mathbf{s}, \mathbf{s}^2\}$ can produce reasonable accuracy in the regression models, where the adjusted R-square values of the regression for P_m , q and P_p are 95.8%, 94.4% and 97.5%, respectively. Furthermore, we find that all elements of the initial summary statistics \mathbf{s} are informative about P_m . However, to obtain estimates for q and P_p , the two largest radii of the expanding cell colonies were not significant in the regression. From the regression analysis, we obtain the point estimate $(\hat{P}_m, \hat{q}, \hat{P}_p) = (0.1217, 0.2477, 0.0015)$. Using the values of $\Delta = 18 \mu\text{m}$ and $\tau = 0.04 \text{h}$, we obtain estimates of the biological parameters D and λ , $\hat{D} = 246 \mu\text{m}^2 \text{h}^{-1}$ and $\hat{\lambda} = 0.038 \text{h}^{-1}$. A Bayesian PB procedure with $B = 1,000$ particles is then performed with the point estimate $(\hat{P}_m, \hat{q}, \hat{P}_p)$. The resulting PB distributions are given in Figure 5 together with the posterior distributions from the PB SMC ABC and the standard SMC ABC algorithms. A numerical summary is given in Table 2.

Results in Figure 5 show that the Bayesian PB distributions (dashed red) are very close to the ABC posterior distributions for q and λ , whereas there is some deviation between the PB distribution and the ABC posterior for D . This suggests that the PB

distributions produce good approximations to the posterior distributions of q and λ , and a good enough approximation for D to produce a useful initial importance distribution for the PB SMC ABC algorithm. This PB distribution is produced using only 19,925 models simulations.

The ABC posterior distributions resulting from the two ABC algorithms, with sample size $N = 2,000$, are very similar as expected given that we use the same summary statistics \mathbf{s}^{FP} and the same final target tolerance. However, given the same target tolerance $\epsilon_T = 0.4$, the PB SMC ABC (solid green curves) only requires 135,080 model simulations, whereas the standard SMC ABC (dotted blue curves) using an initial importance distribution formulated from the pilot run needed more than 210,000 model simulations.

8 Discussion

In this paper, we have presented a novel approach to perform a Bayesian PB for models with intractable likelihoods and newly developed ABC algorithms that aim to minimise the number of model simulations. The main idea is to use the PB distribution as an initial importance distribution for SMC ABC (named PB SMC ABC), and ABC IS (named PB ABC IS) algorithms. The new algorithms were validated on two test examples, the g-and-k quantile distribution and the toggle switch model. Both of these examples have shown that the PB SMC ABC algorithm outperformed the standard SMC ABC by 25–35% in terms of the number of model simulations. We also found efficiency gains when incorporating the PB approximation in ABC IS. However, for brevity, results from the PB ABC IS and standard ABC IS algorithms were not shown.

The main application of the new method was to obtain Bayesian inference for the key parameters governing the expansion of melanoma cell colonies. The simulation procedure from the stochastic model is computationally intensive for some regions of the parameter space (a high proliferation rate). Thus, using the PB approximation as an importance distribution is efficient as it is reasonably close to the ABC posterior and does not propose additional parameter values in parameter regions where it is expensive to simulate.

To speed up the SMC ABC method we used the PB approximation $g(\theta)$ as the initial importance distribution as opposed to the prior. Another approach would define a sequence of distributions formed by a geometric path that connects $g(\theta)$ with the ABC target of interest, $p_{\epsilon_T}(\theta|\mathbf{s}_{obs})$.

The PB distributions can also be embedded within MCMC ABC algorithms. While Fearnhead and Prangle (2012) used the results from the pilot run to choose a starting value of the chain and to form a proposal distribution for MCMC ABC algorithms, one could use an analytical approximation to the PB distribution to form a proposal distribution and use the point estimate obtained from the regression procedure as a starting value. For the tolerance value in MCMC ABC algorithms, one could use a particular quantile of the discrepancies produced from the PB replications.

It should also be noted that the quality of the PB distributions relies very much on the quality of the multiple linear regression procedure to obtain a point estimate of θ .

Investigating the output from the linear model can help to identify which parameters were poorly estimated and as such one could modify the explanatory variables or the training region of model parameters to obtain more accurate results. Here, we suggest that the PB method is more appropriate when the regression model is able to explain a large amount of the variation in the parameters, with adjusted R-square above 80%. There are also several alternative approaches to linear regression such as non-linear regression methods (Blum and François, 2010), artificial neural networks (Chen, 2012) or partial least squares (Beaumont, 2012). These more advanced techniques may be useful in obtaining better fitting regressions. Here, we demonstrate that the PB method is more suitable for low to moderate dimensional θ , where the model simulation is expensive, as in the melanoma application considered in this paper.

To examine whether the assumption for the symmetry property for the summary statistics in the Bayesian PB is reasonable, we could apply a synthetic likelihood evaluation to estimate the log-likelihood ratio in (4). If there is no extreme value in the estimated ratios, then the assumption of symmetry is reasonable. The other assumption that our method makes is that the regression produces accurate point estimates for the parameters. If a particular point estimate is biased then the corresponding posterior approximation is likely to be biased too.

We also examined the possibility of integrating a non-parametric bootstrap procedure for models with intractable likelihoods. The Bayesian version of the non-parametric bootstrap was introduced by Rubin (1981) and later was extended by Newton and Raftery (1994) who named it the weighted likelihood bootstrap (WLB). Rubin (1981) used non-parametric bootstrapping of the maximum likelihood estimate which relies on re-sampling the data, and as such this approach may be applicable for datasets of independent observations, such as the g-and-k example in this paper, but cannot be easily applied if there is a complex dependence structure in the data, such as in the melanoma example.

Instead of re-sampling the data as in Rubin (1981), the WLB randomly weights the components of a likelihood function then maximises this weighted likelihood function to provide a bootstrap replication of the parameter. For a certain weight distribution, the WLB samples can provide an approximation to the posterior distribution, and as such it can be used to form a good starting point for an adaptive importance sampling algorithm, similar in spirit to what we do in this paper. This approach is straightforward to apply, however it relies on being able to explicitly write the likelihood function as a product of components so different weights for each component can be easily applied. Thus, the WLB is not applicable for models of interest in this paper. In conclusion, we suggest that the PB approach is the only bootstrap method generally applicable for models with intractable likelihoods.

Supplementary Material

Supplement Material for “Bayesian parametric bootstrap for models with intractable likelihoods” (DOI: [10.1214/17-BA1071SUPP](https://doi.org/10.1214/17-BA1071SUPP); .pdf).

References

- Aeschbacher, S., Beaumont, M. A., and Futschik, A. (2012). “A novel approach for choosing summary statistics in approximate Bayesian computation.” *Genetics*, 192: 1027–1047. [215](#)
- Allingham, D., King, R. A. R., and Mengersen, K. L. (2009). “Bayesian estimation of quantile distributions.” *Statistics and Computing*, 19: 189–201. [MR2486231](#). doi: <https://doi.org/10.1007/s11222-008-9083-x>. [220](#)
- Australian Institute of Health and Welfare and Australasian Association of Cancer Registries (2012). *Cancer in Australia: an overview*. Cancer series no. 74. Cat. no. CAN 70. Canberra: AIHW. [226](#)
- Beaumont, M. A. (2012). “Discussion of a paper by P. Fearnhead and D. Prangle.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3): 419–474. [MR2925370](#). doi: <https://doi.org/10.1111/j.1467-9868.2011.01010.x>. [230](#)
- Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009). “Adaptive approximate Bayesian computation.” *Biometrika*, 96(4): 983–990. [MR2767283](#). doi: <https://doi.org/10.1093/biomet/asp052>. [214](#)
- Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). “Approximate Bayesian computation in population genetics.” *Genetics*, 162(4): 2025–2035. [211](#), [213](#)
- Blum, M. G. B. (2010). “Approximate Bayesian Computation: a non-parametric perspective.” *Journal of the American Statistical Association*, 105(491): 1178–1187. [MR2752613](#). doi: <https://doi.org/10.1198/jasa.2010.tm09448>. [212](#)
- Blum, M. G. B. and François, O. (2010). “Non-linear regression models for approximate Bayesian computation.” *Statistics and Computing*, 20(1): 63–73. [MR2578077](#). doi: <https://doi.org/10.1007/s11222-009-9116-0>. [230](#)
- Blum, M. G. B., Nunes, M. A., Prangle, D., and Sisson, S. A. (2013). “A comparative review of dimension reduction methods in approximate Bayesian computation.” *Statistical Science*, 28(2): 189–208. [MR3112405](#). [213](#)
- Bonassi, F. V., You, L., and West, M. (2011). “Bayesian learning from marginal data in bionetwork models.” *Statistical Applications in Genetics and Molecular Biology*, 10(1): 49. [MR2851291](#). doi: <https://doi.org/10.2202/1544-6115.1684>. [212](#), [222](#), [223](#), [224](#)
- Cameron, E. and Pettitt, A. (2012). “Approximate Bayesian computation for astronomical model analysis: A case study in galaxy demographics and morphological transformation at high Redshift.” *Monthly Notices of the Royal Astronomical Society*, 425: 44. [211](#)
- Chen, Y. (2012). “Discussion of a paper by P. Fearnhead and D. Prangle.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3): 419–474. [MR2925370](#). doi: <https://doi.org/10.1111/j.1467-9868.2011.01010.x>. [230](#)

- Del Moral, P., Doucet, A., and Jasra, A. (2012). “An Adaptive Sequential Monte Carlo Method for Approximate Bayesian Computation.” *Statistics and Computing*, 22(5): 1009–1020. MR2950081. doi: <https://doi.org/10.1007/s11222-011-9271-y>. 214
- Drovandi, C. C. and Pettitt, A. N. (2011a). “Estimation of Parameters for Macroparasite Population Evolution using Approximate Bayesian Computation.” *Biometrics*, 67(1): 225–233. MR2898834. doi: <https://doi.org/10.1111/j.1541-0420.2010.01410.x>. 214
- Drovandi, C. C. and Pettitt, A. N. (2011b). “Likelihood-free Bayesian Estimation of Multivariate Quantile Distributions.” *Computational Statistics & Data Analysis*, 55(9): 2541–2556. MR2802334. doi: <https://doi.org/10.1016/j.csda.2011.03.019>. 220
- Drovandi, C. C. and Pettitt, A. N. (2011c). “Using Approximate Bayesian Computation to Estimate Transmission Rates of Nosocomial Pathogens.” *Statistical Communications in Infectious Diseases*, 3(1): 2. MR2819387. doi: <https://doi.org/10.2202/1948-4690.1025>. 211
- Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011). “Approximate Bayesian computation using indirect inference.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3): 503–524. MR2767849. doi: <https://doi.org/10.1111/j.1467-9876.2010.00747.x>. 215
- Efron, B. (2012). “Bayesian inference and the parametric bootstrap.” *The Annals of Applied Statistics*, 6(4): 1971–1997. MR3058690. doi: <https://doi.org/10.1214/12-AOAS571>. 211, 212, 216, 217
- Fearnhead, P. and Prangle, D. (2012). “Constructing summary statistics for approximate Bayesian computation: Semi-automatic ABC (with discussion).” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3): 419–474. MR2925370. doi: <https://doi.org/10.1111/j.1467-9868.2011.01010.x>. 212, 213, 214, 215, 217, 220, 222, 229
- Garbe, C., Peris, K., Hauschild, A., Saiag, P., Middleton, M., Spatz, A., Grob, J., Malvehy, J., Newton-Bishop, J., Stratigos, A., Pehamberger, H., and Eggermont, A. (2012). “Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline – Update 2012.” *European Journal of Cancer*, 48(15): 2375–2390. 226
- Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). “Construction of a genetic toggle switch in *Escherichia coli*.” *Nature*, 403: 339–342. 222
- Joyce, P. and Marjoram, P. (2008). “Approximately sufficient statistics and Bayesian computation.” *Statistical Applications in Genetics and Molecular Biology*, 7(1): 26. MR2438407. doi: <https://doi.org/10.2202/1544-6115.1389>. 215
- Li, W. and Fearnhead, P. (2016a). “Improved Convergence of Regression Adjusted Approximate Bayesian Computation.” *arXiv:1609.07135v1*. 218
- Li, W. and Fearnhead, P. (2016b). “On the Asymptotic Efficiency of ABC Estimators.” 218

- Newton, M. A. and Raftery, A. E. (1994). “Approximate Bayesian inference with the weighted likelihood bootstrap.” *Journal of the Royal Statistical Society, Series B (Methodological)*, 56(1): 3–48. [MR1257793](#). 211, 230
- Pope, J. H., Morrison, L., Moss, D. J., Parsons, P. G., and Mary, S. R. (1979). “Human malignant melanoma cell lines.” *Pathology*, 11(2): 191–195. 226
- Rayner, G. D. and MacGillivray, H. L. (2002). “Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions.” *Statistics and Computing*, 12(1): 57–75. [MR1877580](#). doi: <https://doi.org/10.1023/A:1013120305780>. 212, 220, 222
- Rubin, D. B. (1981). “The Bayesian bootstrap.” *The Annals of Statistics*, 9(1): 130–134. [MR0600538](#). 211, 230
- Simpson, M. J., Landman, K. A., and Hughes, B. D. (2010). “Cell invasion with proliferation mechanisms motivated by time-lapse data.” *Physica A: Statistical Mechanics and its Applications*, 389(18): 3779–3790. 227
- Sisson, S. A., Fan, Y., and Tanaka, M. M. (2009). “Correction for Sisson et al., Sequential Monte Carlo without likelihoods.” *Proceedings of the National Academy of Sciences*, 106(39): 16889–16890. [MR2301870](#). doi: <https://doi.org/10.1073/pnas.0607208104>. 214
- Tanaka, M. M., Francis, A. R., Luciani, F., and Sisson, S. A. (2006). “Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data.” *Genetics*, 173(3): 1511–1520. 211
- Treloar, K. K., Simpson, M. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S., and Baker, R. E. (2013). “Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies.” *BMC Systems Biology*, 7: 137. 212, 226, 227
- Vo, B. N., Drovandi, C. C., and Pettitt, A. N. (2019). “Supplement Material for “Bayesian parametric bootstrap for models with intractable likelihoods”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/17-BA1071SUPP>. 218
- Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Pettet, G. J. (2015a). “Melanoma cell colony expansion parameters revealed by approximate Bayesian computation.” *PLoS Computational Biology*, 11(12): e1004635. 211, 212, 213, 214, 215, 222, 225, 226, 227, 228
- Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Simpson, M. J. (2015b). “Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation.” *Mathematical Biosciences*, 263: 133–142. [MR3327999](#). doi: <https://doi.org/10.1016/j.mbs.2015.02.010>. 211, 227
- Wegmann, D., Leuenberger, L., and Excoffier, L. (2009). “Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood.” *Genetics*, 182: 1207–1218. 215

Whitehead, R. and Little, J. H. (1973). “Tissue culture studies on human malignant melanoma.” *Pigment Cell*, 1: 382–389. [226](#)

Acknowledgments

We thank Dr Katrina Treloar for access to the experimental data for Section 6, and appreciate the assistance of the high performance computing team at QUT. We also thank the anonymous Associate Editor and two referees for their helpful comments. ANP, CCD and BNV acknowledge support from the Australian Research Council (DP110100159). CCD was supported by an Australian Research Council’s Discovery Early Career Researcher Award funding scheme (DE160100741).