Robust Bayesian synthetic likelihood via a semi-parametric approach

Ziwen An1,3 · David J. Nott2,4 · Christopher Drovandi1,3

Received: 21 December 2018 / Accepted: 9 September 2019 / Published online: 4 October 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Bayesian synthetic likelihood (BSL) is now a well-established method for performing approximate Bayesian parameter estimation for simulation-based models that do not possess a tractable likelihood function. BSL approximates an intractable likelihood function of a carefully chosen summary statistic at a parameter value with a multivariate normal distribution. The mean and covariance matrix of this normal distribution are estimated from independent simulations of the model. Due to the parametric assumption implicit in BSL, it can be preferred to its nonparametric competitor, approximate Bayesian computation, in certain applications where a high-dimensional summary statistic is of interest. However, despite several successful applications of BSL, its widespread use in scientific fields may be hindered by the strong normality assumption. In this paper, we develop a semi-parametric approach to relax this assumption to an extent and maintain the computational advantages of BSL without any additional tuning. We test our new method, semiBSL, on several challenging examples involving simulated and real data and demonstrate that semiBSL can be significantly more robust than BSL and another approach in the literature.

Keywords Likelihood-free inference · Approximate Bayesian computation (ABC) · Copula · Nonparanormal distribution · Kernel density estimation · Robust estimation

1 Introduction

Approximate Bayesian computation (ABC) is now a well-known method for conducting Bayesian statistical inference for stochastic simulation models that do not possess a computationally tractable likelihood function (Sisson et al. 2018). ABC bypasses likelihood evaluations by preferring parameter configurations that generate simulated data close to the observed data, typically on the basis of a carefully chosen summarisation of the data.

Although ABC has allowed practitioners in many diverse scientific fields to consider more realistic models, it still has several drawbacks. ABC is effectively a nonparametric procedure and scales poorly with summary statistic dimension. Consequently, most ABC analyses resort to a low-dimensional summary statistic to maintain a manageable level of computation, which could lead to significant information loss. Secondly, ABC requires the user to select values for various tuning parameters, which can impact on the approximation.

An alternative approach, called synthetic likelihood (SL, Wood 2010; Price et al. 2018), assumes that the model statistic for a given parameter value has a multivariate normal distribution. The mean and covariance matrix of this distribution are estimated via independent simulations of the statistic, which is used to approximate the summary statistic likelihood at the observed statistic. Price et al. (2018) developed the first Bayesian approach for SL, called Bayesian synthetic likelihood (BSL). The parametric assumption made by the SL allows it to scale more efficiently to increasing dimension of the summary statistic (Price et al. 2018). Further, the BSL approach requires very little tuning and is ideal for exploiting parallel computing architectures compared to ABC.
The SL method has been tested successfully and shows great potential in a wide range of application areas such as epidemiology and ecology (Barbu et al. 2017; Price et al. 2018). Further, Price et al. (2018) and Everitt (2017) demonstrate that BSL exhibits some robustness to a departure from normality. However, a barrier to the ubiquitous use of SL is its strong normality assumption. We seek an approach that maintains the efficiency gains of BSL but enhances its robustness.

The main aim of this paper is to develop a more robust approach to BSL. We do this by developing a semi-parametric method for approximating the summary statistic likelihood, called semiBSL, which involves using kernel density estimates for the marginals and combining them with a Gaussian copula. An incidental contribution of this paper is a thorough empirical investigation into the robustness of BSL. We demonstrate through several simulated and real examples that semiBSL can offer significantly improved posterior approximations compared to BSL. Furthermore, we find that the number of model simulations required for semiBSL does not appear to increase significantly compared to BSL and does not require any additional tuning parameters. However, we also find that the standard BSL approach can be remarkably robust on occasions. We also identify the limits of semiBSL by considering an example with nonlinear dependence structures between summaries.

There have been other approaches developed for robustifying the synthetic likelihood. Fasiolo et al. (2018) developed an extended empirical saddlepoint (EES) estimation method, which involves shrinking the empirical saddlepoint approximation towards the multivariate normal distribution. The regularisation helps ensure that their likelihood estimator is well defined even when the observed statistic lies in the tail of the model summary statistic distribution. The shrinkage parameter (called the “decay”) needs to be selected by the user and effectively represents a bias/variance trade-off. More shrinkage towards the Gaussian decreases variance but increases bias. Although Fasiolo et al. (2018) consider frequentist estimation only, we demonstrate that it appears to increase significantly compared to BSL and does not require any additional tuning parameters. However, we also find that the standard BSL approach can be remarkably robust on occasions. We also identify the limits of semiBSL by considering an example with nonlinear dependence structures between summaries.

2 Bayesian synthetic likelihood

In ABC and BSL, the objective is to simulate from the summary statistic posterior given by

\[
p(\theta|s) \propto p(s|\theta) p(\theta),
\]

where \( \theta \in \Theta \subseteq \mathbb{R}^p \) is the parameter that requires estimation with corresponding prior distribution \( p(\theta) \). Here, \( y \in Y \) is the observed data that are subsequently reduced to a summary statistic \( s_y = S(y) \), where \( S(\cdot) : Y \to \mathbb{R}^d \) is the summary statistic function. The dimension of the statistic \( d \) must be at least the same size as the parameter dimension, i.e. \( d \geq p \).

The SL (Wood 2010) involves approximating \( p(s|\theta) \) with

\[
p(s|\theta) \approx p_A(s|\theta) = \mathcal{N}(s_y|\mu(\theta), \Sigma(\theta)).
\] (1)

The mean and covariance \( \mu(\theta) \) and \( \Sigma(\theta) \) are not available in closed form but can be estimated via independent model simulations at \( \theta \). The procedure involves drawing \( x_{1:n} = (x_1, \ldots, x_n) \), where \( x_i \sim p(\cdot|\theta) \) for \( i = 1, \ldots, n \), and calculating the summary statistic for each dataset, \( s_{1:n} = (s_1, \ldots, s_n) \), where \( s_i \) is the summary statistic for \( x_i \), \( i = 1, \ldots, n \). These simulations can be used to estimate \( \mu \) and \( \Sigma \) unbiasedly.
\[ \mu_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} s_i, \]
\[ \Sigma_n(\theta) = \frac{1}{n-1} \sum_{i=1}^{n} (s_i - \mu_n(\theta))(s_i - \mu_n(\theta))^\top. \] (2)

The estimates in (2) can be substituted into the SL in (1) to estimate the SL as \( N(s_y|\mu_n(\theta), \Sigma_n(\theta)) \). We can sample from the approximate posterior using MCMC, see Algorithm 1. Theoretically, the corresponding MCMC algorithm targets the approximate posterior \( p_{A,n}(\theta|s_y) \propto E[N(s_y|\mu_n(\theta), \Sigma_n(\theta))]p(\theta) \). For finite \( n \), \( N(s_y|\mu_n(\theta), \Sigma_n(\theta)) \) is a biased estimate of \( N(s_y|\mu(\theta), \Sigma(\theta)) \), thus resulting in \( p_{A,n}(\theta|s_y) \) being theoretically dependent on \( n \) (Andrieu and Roberts 2009). However, Price et al. (2018) demonstrate empirically that the BSL posterior \( p_{A,n}(\theta|s_y) \) is remarkably insensitive to its only tuning parameter, \( n \). Thus, we can choose \( n \) to maximise computational efficiency.

**Input**: Summary statistic of the data, \( s_y \), the prior distribution, \( p(\theta) \), the proposal distribution, \( q \), the number of iterations, \( T \), and the initial value of the chain, \( \theta^0 \).

**Output**: MCMC sample \( (\theta^0, \theta^1, \ldots, \theta^T) \) from the BSL posterior, \( p_{A,n}(\theta|s_y) \). Some samples can be discarded as burn-in if required.

1. Simulate \( x_{1:n} \) i.i.d. \( p(\theta^0) \) and compute \( s_{1:n} \).
2. Compute \( \theta^0 = (\mu_n(\theta^0), \Sigma_n(\theta^0)) \) using (2).
3. for \( i = 1 \) to \( T \) do
   4. Draw \( \theta^* \sim q(\theta^{i-1}) \).
   5. Simulate \( x_{1:n}^{*} \) i.i.d. \( p(\theta^*) \) and compute \( s_{1:n}^{*} \).
   6. Compute \( \phi^* = (\mu_n(\theta^*), \Sigma_n(\theta^*)) \) using (2).
   7. Compute \( r = \min\left(1, \frac{N(s_y|\mu_n(\theta^*), \Sigma_n(\theta^*))p(\theta^*)}{N(s_y|\mu_n(\theta^i), \Sigma_n(\theta^i))p(\theta^i)} \right) \).
   8. if \( r < 1 \) then
      9. Set \( \theta^i = \theta^* \) and \( \phi^i = \phi^* \).
   10. else
      11. Set \( \theta^i = \theta^{i-1} \) and \( \phi^i = \phi^{i-1} \).
   12. end
13. end

**Algorithm 1**: MCMC BSL algorithm

The computational efficiency of BSL over ABC stems from the normality assumption. This approximation can work well in many applications. However, there are also several reasons to suggest why it may not be appropriate. For a given parameter value, there may be model summary statistics with non-normal marginal distributions and nonlinear dependencies between marginals. Further, the normality assumption may not be appropriate over all the non-negligible posterior mass. Finally, as the dimension of the summary statistic grows it is likely that the multivariate normal assumption becomes less reasonable.

In the next section, we develop a method to increase the robustness of BSL, while largely retaining its computationally convenient properties.

### 3 Semi-parametric approach to Bayesian synthetic likelihood

Here we propose to use copula models to approximate the joint distribution of the model summary statistic. The main appeal of copula models is the ability to model marginal distributions and joint dependence structures separately. In this paper, we use kernel density estimates (KDEs, Rosenblatt 1956; Parzen 1962) for modelling each marginal distribution. KDEs are useful as they can be flexible and are nonparametric, eliminating the need for the user to select specific parametric forms for the marginals. However, we do note that other univariate density estimators could be used in our approach. After transforming the marginals based on the fitted KDEs, we attempt to capture the dependence structure between summaries by using a Gaussian copula. The appeal of the Gaussian copula is its tractability, which makes it ideal for its repeated use within a Bayesian algorithm. The density estimator we use is similar to the approach of Liu et al. (2009).

Of course, the standard SL approach is a special case of ours when the marginal distributions are assumed normal. Given the semi-parametric nature of the density estimator we use, we refer to our approach as semiBSL. We now describe the technical details of our method.

#### 3.1 Kernel density estimation

We model each univariate marginal distribution of the joint statistic using a KDE. A KDE of an unknown distribution \( f_X(x) \) for a continuous random variable \( X \in \mathcal{X} \) can be defined based on \( n \) independent and identically drawn samples \( x_1, \ldots, x_n \in \mathcal{X} \) as

\[
\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - x_i),
\]

where \( K_h(\cdot) \) is the kernel function, which has non-negative support over its domain and integrates to 1, i.e. \( \int_{\mathcal{X}} K_h(x)dx = 1 \). There are a large variety of kernels to choose from in KDE, and here we present results using a Gaussian kernel because it is nonzero everywhere. Note that we also tested the Epanechnikov kernel (Epanechnikov 1969), which is optimal in terms of asymptotic mean integrated squared error. The posterior distributions produced by both kernels are visually very similar in several of our test examples.
We generally assume that each summary statistic is continuous and unbounded. Large discrete summary statistics (e.g. counts) can be treated as continuous. Various transformations can be used if summary statistics are bounded (e.g. log transform for statistics taking positive values only).

The smoothness of the KDE is controlled by the bandwidth parameter $h$. Unlike standard ABC which pre-specifies the bandwidth parameter to use in approximating the joint summary statistic likelihood, we use the bandwidth suggested by Silverman (2018), i.e. $h = 0.9n^{-0.2} \min(\sigma, \text{interquartile range}/1.34)$, where $\sigma$ is the standard deviation of the distribution and can be estimated empirically. This is also the adopted implementation in the density function in the statistical software R. We use the same bandwidth for cumulative distribution function (CDF) estimation later, but it is possible to use a different bandwidth for the PDF as desired.

### 3.2 Copula

Suppose a random vector $X = (X_1, \ldots, X_d)^\top$ has continuous marginals. Sklar’s theorem (Sklar 1959) states that the multivariate CDF of $X$ can be uniquely defined by its marginal CDFs, denoted, $u_j = F_j(x_j)$, $j = 1, \ldots, d$, and the dependence structure is captured with a multivariate copula function, where $u_j$ is uniformly distributed. A convenient choice of copula in many applications is the Gaussian copula, whose probability density function (PDF) can be written as

$$c_R(u) = \frac{1}{\sqrt{|R|}} \exp \left\{ -\frac{1}{2} u^\top (R^{-1} - I_d) u \right\},$$

where $I_d$ is a $d$-dimensional identity matrix and $u_j = \Phi^{-1}(u_j)$, $j = 1, \ldots, d$, in which $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard normal distribution. The sole parameter $R$ of the Gaussian copula is a correlation matrix, estimation of which will be discussed shortly. Once an estimated correlation matrix $\hat{R}$ is obtained, semiBSL estimates the summary statistic likelihood $p(s^\prime_j | \theta)$ by

$$g(s^\prime_j | \theta) = \frac{1}{\sqrt{|\hat{R}|}} \exp \left\{ -\frac{1}{2} \tilde{\eta}^\top \hat{R}^{-1} \tilde{\eta} \right\} \prod_{j=1}^d \tilde{f}_j(s^\prime_j),$$  \hspace{1cm} (3)

where $s^\prime_j$ is the $j$th component of $s^\prime$, $\tilde{f}_j(\cdot)$ is the estimated KDE of the $j$th marginal probability density, $\tilde{\eta}_j = (\tilde{\eta}_{i,j}, \ldots, \tilde{\eta}_{d,j})^\top$, $\tilde{\eta}_{i,j} = \Phi^{-1}(\hat{u}_j)$, $j = 1, \ldots, d$, and $\hat{u}_j = \hat{F}_j(s^\prime_j)$.

It is critical that the semi-parametric density estimator is fast to fit here, since it must be applied at every iteration of semiBSL. The major appeal of the Gaussian copula here is its estimation tractability. We discuss some other possible extension to the copula model in Sect. 5.

### 3.3 Gaussian rank correlation

The standard approach to estimate the Gaussian copula parameter $R$ is to compute the sample correlation matrix based on the collection $\{\eta_k\}_{k=1}^n$ where $\eta_k = (\eta_{i,k}, \ldots, \eta_{d,k})^\top$ and $\eta_{i,j} = \Phi^{-1}(\hat{F}_j(s^\prime_j))$. However, this approach for estimating $R$ relies on each of the KDEs providing a very good fit to the actual distribution of the corresponding marginal. Although KDE models are flexible, large sample sizes are often required to capture strong irregularities, even in univariate distributions.

To obtain a procedure that is more robust to a potential lack of fit of the KDEs, we consider a nonparametric estimator of $R$ using the Gaussian rank correlation (GRC, Boudt et al. 2012). Using the notation defined earlier, the $(i, j)$th component of the GRC estimate is given by

$$\hat{\rho}^{\text{GRC}}_{i,j} = \frac{\sum_{k=1}^n \Phi^{-1} \left( \frac{r(s^\prime_k)}{n+1} \right) \Phi^{-1} \left( \frac{r(s^\prime_k)}{n+1} \right)} {\sum_{k=1}^n \Phi^{-1} \left( \frac{k}{n+1} \right)^2},$$  \hspace{1cm} (4)

where $r(\cdot) : \mathbb{R} \to \mathcal{A}$, where $\mathcal{A} \equiv \{1, \ldots, n\}$, is the rank function. The GRC is not expensive to compute and has two favourable properties. Boudt et al. (2012) show that the GRC is robust to a small number of outliers. In fact, it requires at least 12.4% of data contamination to revert a positive linear correlation. In addition, the GRC estimator is always semi-positive definite when data are multivariate normal.

The full procedure to estimate the summary statistic likelihood at the point $s^\prime_j$ using our semiBSL approach is shown in Algorithm 2. Importantly, our method does not require any additional tuning parameters compared with BSL. Replacing the Gaussian estimate $\mathcal{N}(s^\prime_j | \mu_{s^\prime_j}(\theta^*), \Sigma_{s^\prime_j}(\theta^*))$ with the semi-parametric estimate $g(s^\prime_j | \theta)$ of equation (3) in Algorithm 1 gives MCMC semiBSL.

Similar to BSL, the number of model simulations $n$ used in semiBSL represents a trade-off between the cost per iteration and the overall acceptance rate in MCMC. In “Appendix A” (Online Resource 1), we show that semiBSL does not require a larger $n$ compared to BSL in an illustrative toy example. This is important, since we find in this example and the applications below that semiBSL can provide robustness without incurring additional model simulations.

### 4 Applications

In this section, we investigate the performance of semiBSL on five examples. All the examples have some noticeable non-normal marginal summary statistic regardless of the joint...
distribution, see “Appendix B.1” (Online Resource 1). We also look into an example with strong nonlinear dependencies between summaries, which challenges the Gaussian copula assumption in semiBSL. In addition to the results presented in this section, estimates of univariate posterior distributions are provided in “Appendix B.2” (Online Resource 1).

We compare semiBSL with BSL and also a Bayesian version of the EES approach of Fasiolo et al. (2018). The EES method requires specification of an additional tuning parameter (the “decay” parameter), which mutates the EES between an empirical saddlepoint approximation (decay equals to 0) and a multivariate normal distribution (decay approaches infinity). We use the cross-validation method suggested by Fasiolo et al. (2018) to select the decay, which is also available in the associated R package.

As shown in Price et al. (2018) and An et al. (2019), BSL is remarkably insensitive to n in a variety of applications. We tested several different values of n for each example and found that semiBSL appears to inherit this useful property (see “Appendix C” (Online Resource 1) for more results). Therefore, the number of simulations per iteration n in all examples is chosen to maximise computational efficiency (measured by scaled effective sample size). The total number of simulations (the number of simulations per iteration times the total number of iterations) is fixed for BSL and semiBSL in all following examples. However, since the performance of EES is not as stable, we increase the total number of simulations accordingly.

We use MCMC with a multivariate normal random walk proposal to sample from the approximate posterior distribution in all methods. Since the different approaches may result in different posterior approximations, we tune the covariance matrix of the random walk separately. We use pilot runs until we believe we have a reasonable estimate of the correspond-

Algorithm 2: The semi-parametric procedure to estimate the summary statistic likelihood \( g(s_y|\theta) \) in semiBSL.

\begin{verbatim}
1 for j = 1 to d do
2   Estimate \( \tilde{f}_j(s_{y_j}) \) and \( \hat{f}_j(s_{y_j}) \) based on \( \{S_i^j\}_{i=1}^n \) using KDE.
3 end
4 for i = 2 to d do
5   for j = 1 to d - 1 do
6     Compute ranks \( r(s_{y_j}^i) \) and \( r(s_{y_j}^i) \) for \( k = 1, \ldots, n \).
7     Estimate \( \hat{p}_{ij}^{(k)} \) using the ranks above with equation (4).
8 end
9 end
10 Construct \( \hat{R} \) and compute \( g(s_y|\theta) \) with equation (3).
\end{verbatim}

We tested several different values of \( n \) (see “Appendix C” (Online Resource 1) for more results). The results are shown for four different transformations and EES. The contour plots represent the “true” posterior (obtained from a long run of MCMC using the exact likelihood). For the posterior approximations, bivariate scatterplots of the posterior samples are shown. We use the total variation distance to quantify the disparity between the
Fig. 1  Bivariate scatter plots of the posterior distributions of the MA(2) example using different pairs of transformation parameters for the summary statistic. The overlaid contour plots represent the true posterior. The number of simulations per iteration $n$ and the total variation distance compared to the gold standard are shown in the corner of each plot.
approximated posterior distribution and the true posterior, i.e.
\[ tv(f_1, f_2) = \int \frac{1}{2} \int \left| f_1(\theta) - f_2(\theta) \right| d\theta. \]
This can be estimated via numerical integration with a 2-D grid covering most of the posterior mass. The choice of \( n \) and the total variation distance are shown in Fig. 1.

In order to show the stability and performance of our approximate posterior results, we also run BSL and semiBSL for additional datasets generated from the MA(2) model with the same true parameter configuration. We compute the total variation using nonparametric density estimation for each case and summarise the results in Table 1. It is apparent that semiBSL always outperforms BSL in terms of the total variation distance to the true posterior.

The bivariate posteriors and total variations indicate that the semiBSL posteriors are robust to all four transformed summary statistics, while BSL fails to get close to the true posterior. The results also suggest that the EES provides little robustness in this example. We also tested the EES method with a much larger \( n \), here \( n = 5000 \), for the randomised \( \epsilon, \delta \) dataset to see whether the posterior accuracy improves. With \( n = 5000 \), the decay parameter is reduced to 45 (roughly one-third of the value obtained with \( n = 300 \)). We find that even with significant additional computational effort, the EES posterior approximation shows little improvement.

It is important to note that the dependence structure in the data (after back-transformation) is Gaussian. Therefore, the Copula assumption made by semiBSL is correct in this example up to estimation of the marginals, which are done using KDE without knowledge of true marginals. Thus, we would expect semiBSL to provide good posterior approximations here and it is necessary to test its performance in other examples. However, the example does serve to illustrate that BSL can be impacted by non-normality and that the EES may not provide sufficient robustness to non-normality.

### 4.3 Stereological extremes

During the process of steel production, the occurrence of microscopic particles, called inclusions, is a critical measure of the quality of steel. It is desirable that the inclusions are kept under a certain threshold, since steel fatigue is believed to start from the largest inclusion within the block. Direct observation of three-dimensional inclusions is inaccessible, so that inclusion analysis typically relies on examining two-dimensional planar slices. Anderson and Coles (2002) establish a mathematical model to formulate the relationship between observed cross-section samples, \( S \), and the real diameter of inclusions, \( V \), assuming that the inclusions are spherical. The model focuses on large inclusions, i.e. \( V > v_0 \), where \( v_0 \) is a certain threshold, which is endowed with a generalised Pareto distribution such that...
The spherical model possesses a likelihood function that is easily computable, the spherical assumption itself might be inappropriate. This leads to the ellipsoidal model proposed by Bortot et al. (2007), who use ABC for likelihood-free inference due to the intractable likelihood function that it inherits. The new model assumes that inclusions are ellipsoidal with principal diameters \((V_1, V_2, V_3)\), where \(V_3\) is the
largest diameter without loss of generality. Here, $V_1 = U_1 V_3$ and $V_2 = U_2 V_3$, where $U_1$ and $U_2$ are independent uniform $U(0, 1)$ random variables. The observed value $S$ is the largest principal diameter of an ellipse in the two-dimensional cross-section.

Here, we consider the ellipsoidal model with parameter of interest $\theta = (\lambda, \sigma, \xi)$. The prior distribution is $U(30, 200) \times U(0, 15) \times U(-3, 3)$. Denoting the observed samples as $S$, we consider four summary statistics: the number of inclusions, $\log(\min(S))$, $\log(\text{mean}(S))$, $\log(\text{max}(S))$. Figure 5 in “Appendix B.1” (Online Resource 1) shows the distribution of the chosen summary statistic simulated at a point estimate of $\theta$. The last three summary statistics have a significantly heavy right tail, strongly invalidating the normality assumption of BSL.

Figure 3 shows the bivariate scatterplot of posteriors obtained by BSL, semiBSL and EES. The number of simulations is $n = 50$ for all methods. The overlaying contour plot is drawn with an MCMC ABC result with tolerance 1. A Mahalanobis distance is used to compare summary statistics. The covariance used in the Mahalanobis distance is taken to be the sample covariance of summary statistics independently generated from the model at a point estimate of the parameter. Note that the outliers are removed before computing the Mahalanobis covariance matrix. Given that there are only four summary statistics in this example, we take ABC as the gold-standard approximation. It is apparent that both BSL and EES results are accepting parameter values in the tails that are rejected by ABC and semiBSL. We use the boxplot (Fig. 4) to explain the impact of outlier simulations in BSL. The first column uses a parameter value that has high posterior support in the semiBSL result (referred to here as “good”) and a medium number of simulations, the second column uses a parameter value that should have negligible posterior support (referred to here as “poor”) and a medium number of simulations, and the last column uses a poor parameter value and a large number of simulations. In each test, we simulate 300 independent BSL and semiBSL log-likelihood estimates. Negative infinities are ignored in the second and third tests in semiBSL. It is worth noting that semiBSL produces a lot of negative infinity log-likelihood estimates for simulations at the poor parameter value, which are ignored in the boxplot. The overestimated log-likelihoods at the “poor” parameter value by BSL are competitive to those at the “good” parameter value. Thus, when BSL grossly overestimates the likelihood at a poor parameter value, the algorithm may get stuck in the tails for long periods. Overall, semiBSL is the only method that gives an approximate posterior consistent with that of ABC.

### 4.4 Fowler’s toads

Movements of amphibian animals exhibit patterns of site fidelity and long-distance dispersal at the same time. Modelling such patterns helps in understanding amphibian’s travel behaviour and contributes to amphibian conservation. Marchand et al. (2017) develop an individual-based model for a species called Fowler’s Toads (*Anaxyrus fowleri*) and collect data via radiotracking in Ontario, Canada. The comprehensive experimental and modelling details are stated in the original paper. Here, we only present a brief description of the model.

The model assumes that a toad hides in its refuge site in the daytime and moves to a randomly chosen foraging place at night. After its geographical position is collected via a transmitter, the toad either takes refuge at the current location or returns to one of the previous sites. For simplicity, the refuge locations are projected to a single axis and thus can be represented by a single-dimensional spatial process. GPS location data are collected on $n_t$ toads for $n_d$ days, i.e., the observation matrix $Y$ is of dimension $n_d \times n_t$. For the synthetic data we use here, $n_t = 66$, $n_d = 63$, and missingness is not considered. Then, $Y$ is summarised down to four sets comprising the relative moving distances for time lags of 1, 2, 4 and 8 days. For instance, $y_1$ consists of the displacement information of lag 1 day, $y_1 = \{||\Delta y|| = |Y_{i,j} - Y_{i,j+1}|; 1 \leq i \leq n_d - 1, 1 \leq j \leq n_t\}$.

Simulation from the model involves two distinct processes. For each single toad, we first generate an overnight displacement, $\Delta Y$, and then mimic the returning behaviour with a simplified model. The overnight displacement is deemed to have significant heavy tails, assumed to belong to the Lévy-alpha stable distribution family, with stability parameter $\alpha$ and scale parameter $\gamma$. This distribution has no closed form, while simulation from it is straightforward (Chambers et al. 1976), making simulation-based approaches appealing. The original paper provided three returning models with different complexity. We adopt the random return model here as it has the best performance among the three and is the easiest for simulation. The total returning probability is a constant $p_0$, and if a return occurs on day $1 \leq i \leq m$, then the return site is the same as the refuge site on day $i$, where $i$ is selected randomly from 1, 2, $\ldots$, $m$ with equal probability. Here, we take the observed data as synthetically generated
Fig. 3  Bivariate scatter plots of the posterior distributions of the stereological extreme example. The scatter plot is generated with thinned approximate sample by BSL (run no. 2 in Fig. 11), semiBSL and EES approach. The contour plot corresponds to the gold standard in this example (MCMC ABC). The number of simulations per iteration \( n \) and the total variation distance compared to the gold standard are shown in the corner of each plot

with true parameter \( \theta = (\alpha, \gamma, p_0) = (1.7, 35, 0.6) \). We use a uniform prior over \((1, 2) \times (0, 100) \times (0, 0.9) \) here.

We fit a four-component Gaussian mixture model to each set of \( \log(|\Delta y|) \). Figure 5 shows the distributions of \( \log(|\Delta y|) \) for lags of 1, 2, 4, 8. As the summary statistic we use the 11-dimensional score of this fitted auxiliary model (corresponding to three component weights, four means and four standard deviations). This corresponds to the indirect inference approach for selecting summary statistics (see Drovandi et al. 2011; Gleim and Pigorsch 2013; Drovandi et al. 2015). Accommodating the four different lags, there are 44 summary statistics in total. The scores do not seem
Fig. 4  Boxplot of BSL and semiBSL log-likelihoods of the stereological extreme example. Negative infinite semiBSL log-likelihood values are ignored. The figure at top is the whole view of the boxplot. The figure at the bottom is zoomed in vertically to $-40$ to $7$ to show more clearly highlight the outliers for BSL.

Fig. 5  Density plot of $\log(|\Delta y|)$ for lags of 1, 2, 4 and 8 days of the Fowler’s toads example to depart a large amount from normality (see Fig. 6 in “Appendix B.1”) (Online Resource 1), and thus, standard BSL may be suitable for this model. To further explore the robustness of BSL and test our semiBSL approach, we include a power transformation of $s_x$ to push the irregularity of the summary statistic further. The transformation function is given by $f_p(x) = \text{sgn}(x) \times (|x|)^p$. It retains the sign of the input and creates a sharp peak near 0. The distribution of the transformed summary statistics using $p = 1.5$ is also included in “Appendix B.1” (Online Resource 1).

In Fig. 6, we show the posterior approximations produced by different approaches and transformation parameters. The posterior distributions obtained by different approaches are reasonably close to each other using the original score summary statistics. The BSL marginal posterior shows significant shift horizontally as the transformation power grows, while semiBSL and EES are more robust to the change.

4.5 Simple recruitment, boom and bust

Here, we consider an example that tests the limits of our semiBSL method. The simple recruitment, boom and bust model was used in Fasiolo et al. (2018) to investigate the performance of the saddlepoint approximation to a non-normal summary statistic. This is a discrete stochastic temporal model that can be used to represent the fluctuation of the population size of a certain group over time. Given the population size $N_t$ and parameter $\theta = (r, \kappa, \alpha, \beta)$, the next value $N_{t+1}$ follows the following distribution

$$N_{t+1} \sim \begin{cases} \text{Poisson}(N_t (1 + r)) + \epsilon_t, & \text{if } N_t \leq \kappa \\ \text{Binom}(N_t, \alpha) + \epsilon_t, & \text{if } N_t > \kappa, \end{cases}$$

where $\epsilon_t \sim \text{Pois}(\beta)$ is a stochastic term. The population oscillates between high- and low-level population sizes for several cycles. The true parameters are $r = 0.4, \kappa = 50, \alpha = 0.09$ and $\beta = 0.05$, and the prior distribution is $U(0, 1) \times U(10, 80) \times U(0, 1) \times U(0, 1)$. There are 250 values in the observed data. We use 50 burn-in values to remove the transient phase of the process.

We construct the summary statistics as follows. Consider a dataset $x$, define the differences and ratios as $d_x = \{x_i - x_{i-1}; i = 2, \ldots, 250\}$ and $r_x = \{x_i / x_{i-1}; i = 2, \ldots, 250\}$, respectively. We use the sample mean, variance, skewness and kurtosis of $x$, $d_x$ and $r_x$ as our summary statistic, $s_x$. We also tested the statistics used in Fasiolo et al. (2018), but we found our choice to be more informative about the model parameters. The parameter $\beta$ seems to have a strong impact on the model statistic distribution. Small values of $\beta$ tend to generate statistics that are highly non-normal so we consider such a case here.

Distributions of the 12-dimensional summary statistic (based on the true parameter value) are shown in Fig. 8 in “Appendix B.1” (Online Resource 1). None of the chosen summary statistics are close to normal. Marginal posterior distributions by BSL, semiBSL, EES and ABC are shown in Fig. 7. The values of $n$ are given in the legend. With only four parameters and 12 summary statistics in this example, ABC can perform well and is treated as the gold standard. In the MCMC ABC result, we manage to
Fig. 6  Comparing the approximate posterior distributions for the BSL, semiBSL and EES approaches at different transformation powers of the Fowler’s toads example. The vertical line indicates the true parameter value.
get 14.5 thousand accepted samples out of 18 million iterations at tolerance = 2. A Mahalanobis distance is used to compare summary statistics. The covariance used in the Mahalanobis distance is taken to be the sample covariance of summary statistics independently generated from the model at \( \theta = (0.4, 50, 0.09, 0.1) \). Figure 8 shows the bivariate scatterplot with overlaying contour plot (ABC result). It is evident that the semiBSL procedure is producing an approximation that is closer to ABC compared to BSL. It does suggest that semiBSL is providing some robustness. However, it is also evident that there is some difference between the ABC and semiBSL results. To gain some insights into the dependence structure between summaries, we consider bivariate scatterplots of the summaries, shown in Fig. 13 in “Appendix B.3” (Online Resource 1). It is clear that there is a high degree of nonlinear dependence between many of the summaries, which cannot be captured by our Gaussian copula. Therefore, this example is highly challenging for our semiBSL approach.

5 Discussion

In this paper, we proposed a new method to relax the normality assumption in BSL and presented several examples of varying complexity to test the empirical performance of BSL and semiBSL. The new approach offered additional robustness in all of the considered examples. Further, given the semi-parametric nature of the method, the computational gains of the fully parametric BSL are largely retained in terms of the number of model simulations required. When model simulations per iteration required by BSL is non-negligible, the additional cost incurred by semiBSL will be small. Estimating marginal KDEs and the Gaussian rank correlation matrix is relatively straightforward.

However, we did observe situations where standard BSL was remarkably robust to lack of normality, which is consistent with some previous literature including Price et al. (2018) and Everitt (2017). Developing some theory around when we can expect standard BSL to work well would be useful.

Previous BSL research (Price et al. 2018) showed that the approximate posterior is very insensitive to the choice of \( n \). Surprisingly, we also found the semiBSL posterior to be relatively insensitive to \( n \), albeit not as insensitive as BSL. We expect semiBSL to be less insensitive to \( n \) since the choice of \( n \) is more likely to impact kernel density estimates compared to the Gaussian synthetic likelihood. The sensitivity to \( n \) results for semiBSL is presented in “Appendix C” (Online Resource 1).

The new approach was also compared with another robustified synthetic likelihood method, the EES (Fasiolo et al. 2018). Because of potential numerical issues with the standard empirical saddlepoint approximation, the EES has to resort to a tuning parameter called the decay, which shifts the estimation between a flexible saddlepoint one and a rigid normal distribution. For roughly the same number of simulations, the posterior approximation by EES generally shows some improvements over BSL but less compared to semiBSL in the examples tested.

Since we can obtain synthetic likelihood estimates of \(-\infty\) with our approach for parameter values in the far tails of the posterior, we recommend that the practitioner firstly finds and initialises the MCMC at a parameter that produces a summary statistic distribution that has reasonable support for the observed statistic. We note that standard BSL can also exhibit slow convergence when initialised in the tail of the posterior (Price et al. 2018).

Note that we use an unrestricted correlation matrix \( R \) in Eq. (3) throughout the main paper. However, it is possible to improve the computational efficiency with shrinkage estimation on \( R \). An et al. (2019) used the graphical lasso (Friedman et al. 2008) in the standard BSL algorithm and proposed a novel penalty selection method so that the Markov chain has efficient mixing. We present the results using a straightforward shrinkage estimator proposed by Warton (2008) in “Appendix D” (Online Resource 1) and show a significant computational gain in the M/G/1 example.

If the true underlying marginal distribution of a statistic is highly irregular, one limitation of our approach is that the number of simulations required for a KDE to capture this will be large. We point out that we only require estimates of the marginal distributions of the summary statistics at the observed statistic values, rather than estimating the entire marginal distribution. We note that future work could revolve around improving the performance of KDE. It could be beneficial to deliberately undersmooth in kernel estimation to reduce bias to get more accurate results in some cases.
Fig. 8  Bivariate scatter and contour plots of the posterior distributions of the simple recruitment, boom and bust example. The scatter plot is generated with thinned approximate sample obtained by BSL, semiBSL or EES methods. The contour plot is based on MCMC ABC with a tolerance of 2. Only three pairs of the parameters are shown here, $r$ versus $\alpha$, $r$ versus $\beta$ and $\kappa$ versus $\beta$. The number of simulations per iteration $n$ and the total variation distance compared to the gold standard are shown in the corner of each plot.

Another direction would be considering other adaptive kernel density estimation approaches, such as the balloon estimator and the sample point estimator (e.g. Terrell and Scott 1992), which may provide more stability and require less model simulations.

There is still research to be done in improving the robustness of BSL. Our semiBSL method relies on the Gaussian copula dependence structure. The results in the boom and bust example show a compromised performance of semiBSL when there exists strong nonlinear dependence structures between summary statistics. For future work, we plan to
investigate other more flexible copula structures such as multivariate skew normal (Sahu et al. 2008) and vine copulas (Bedford and Cooke 2002). Another direction with great potential is to incorporate the semiBSL likelihood estimator into the variational Bayes synthetic likelihood approaches (see Ong et al. 2018a, b) to speed up computation for a high-dimensional statistic and/or parameter.

Overall, we have demonstrated that our semiBSL approach can provide a significant amount of robustness relative to BSL with little or no additional computational cost in terms of the number of model simulations, while requiring no additional tuning parameters.

Acknowledgements CD was supported by an Australian Research Council’s Discovery Early Career Researcher Award funding scheme (DE160100741). ZA was supported by a scholarship under CDis Grant DE160100741 and a top-up scholarship from the Australian Research Council Centre of Excellence for Mathematical and Statistics Frontiers (ACEMS). DJN was supported by a Singapore Ministry of Education Academic Research Fund Tier 1 Grant (R-155-000-189-114). Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia. The authors thank Alex Shestopaloff for sharing his code on exact MCMC for the M/G/1 model.

References

An, Z., South, L.F., Nott, D.J., Drovandi, C.C.: Accelerating Bayesian synthetic likelihood with the graphical lasso. J. Comput. Graph. Stat. 28(2), 471–475 (2019). https://doi.org/10.1080/10618600.2018.1537928

Anderson, C., Coles, S.: The largest inclusions in a piece of steel. Extremes 5(3), 237–252 (2002)

Andrieu, C., Roberts, G.O.: The pseudo-marginal approach for efficient Monte Carlo computations. Ann. Stat. 37(2), 697–725 (2009)

Barbu, C.M., Sethuraman, K., Billig, E.M.W., Levy, M.Z.: Two-scale dispersal estimation for biological invasions via synthetic likelihood. Ecography 41(4), 661–672 (2017)

Bedford, T., Cooke, R.M.: Vines-a new graphical model for dependent random variables. Ann. Stat. 30(4), 1031–1066 (2002)

Blum, M.G.B., François, O.: Non-linear regression models for approximate Bayesian computation. Stat. Comput. 20(1), 63–73 (2010)

Bortot, P., Coles, S.G., Sisson, S.A.: Inference for stereological extremes. J. Am. Stat. Assoc. 102(477), 84–92 (2007)

Boudt, K., Cornelissen, J., Croux, C.: The Gaussian rank correlation estimator: robustness properties. Stat. Comput. 22(2), 471–483 (2012)

Chambers, J.M., Mallows, C.L., Stuck, B.W.: A method for simulating stable random variables. J. Am. Stat. Assoc. 71(354), 340–344 (1976)

Chiachio, M., Beck, J., Chiachio, J., Rus, G.: Approximate Bayesian computation by subset simulation. SIAM J. Sci. Comput. 36(3), A1339–A1358 (2014)

Drovandi, C.C., Pettitt, A.N., Faddy, M.J.: Approximate Bayesian computation using indirect inference. J. R. Stat. Soc. Ser. C (Appl. Stat.) 60(3), 317–337 (2011)

Drovandi, C.C., Pettitt, A.N., Lee, A.: Bayesian indirect inference using a parametric auxiliary model. Stat. Sci. 30(1), 72–95 (2015)

Dutta, R., Corander, J., Kaski, S., Gutmann, M.U.: Likelihood-free inference by ratio estimation. ArXiv preprint arXiv:1611.10242v3 (2017)

Epanechnikov, V.A.: Non-parametric estimation of a multivariate probability density. Theory Probab. Appl. 14(1), 153–158 (1969)

Everitt, RG.: Bootstrapped synthetic likelihood. ArXiv preprint arXiv:1711.05825v2 (2017)

Fasliolo, M., Wood, S.N., Hartig, F., Bravington, M.V.: An extended empirical saddlepoint approximation for intractable likelihoods. Electron. J. Stat. 12(1), 1544–1578 (2018). https://doi.org/10.1214/18-EJS1433

Fearnhead, P., Prangle, D.: Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 74(3), 419–474 (2012)

Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3), 432–441 (2008)

Gleim, A., Pigorsch, C.: Approximate Bayesian computation with indirect summary statistics. Draft paper: http://www.zet-pigorschmeuuni-bonnde/data/research/papers (2013)

Jones, M.C., Pewsey, A.: Sinh-arcsinh distributions. Biometrika 96(4), 761–780 (2009). https://doi.org/10.1093/biomet/asp053

Li, J., Nott, D., Fan, Y., Sisson, S.: Extending approximate Bayesian computation methods to high dimensions via a Gaussian copula model. Comput. Stat. Data Anal. 106, 77–89 (2017)

Liu, J., Lafferty, J., Wasserman, L.: The nonparamoral: semiparameteric estimation of high dimensional undirected graphs. J. Mach. Learn. Res. 10, 2295–2328 (2009)

Marchand, P., Boenke, M., Green, D.M.: A stochastic movement model reproduces patterns of site fidelity and long-distance dispersal in a population of Fowler’s toads (Anaxyrus fowleri). Ecol. Model. 360, 63–69 (2017)

Ong, V.M.H., Nott, D.J., Tran, M.N., Sisson, S.A., Drovandi, C.C.: Likelihood-free inference in high dimensions with synthetic likelihood. Comput. Stat. Data Anal. 128, 271–291 (2018a)

Ong, V.M.H., Nott, D.J., Tran, M.N., Sisson, S.A., Drovandi, C.C.: Variational Bayes with synthetic likelihood. Stat. Comput. 28(4), 971–988 (2018b)

Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. 33(3), 1065–1076 (1962)

Price, I.F., Drovandi, C.C., Lee, A., Nott, D.J.: Bayesian synthetic likelihood. J. Comput. Graph. Stat. 27, 1–11 (2018)

Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. 27(3), 832–837 (1956)

Sahu, S.K., Dey, D.K., Branco, M.D.: A new class of multivariate skew distributions with applications to Bayesian regression models. Can. J. Stat. 31(2), 129–150 (2008)

Shestopaloff, A.Y., Neal, R.M.: On Bayesian inference for the M/G/1 queue with efficient MCMC sampling. ArXiv preprint arXiv:1401.5548 (2014)

Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Routledge, Abingdon (2018)

Sisson, S.A., Fan, Y., Beaumont, M.: Handbook of Approximate Bayesian Computation, 1st edn. Chapman and Hall/CRC, Boca Raton (2018)

Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. Inst. Univ. Paris 8, 229–231 (1959)

Terrell, G.R., Scott, D.W.: Variable kernel density estimation. Ann. Stat. 20(3), 1236–1265 (1992)

Warton, D.I.: Penalized normal likelihood and ridge regularization of correlation and covariance matrices. J. Am. Stat. Assoc. 103(481), 340–349 (2008)

Wood, S.N.: Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466, 1102–1107 (2010)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.