Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2020, 2020:46
https://doi.org/10.28919/cmbn/4568
ISSN: 2052-2541

THE DISCREPANCY MEASURES IN APPROXIMATE BAYESIAN COMPUTATION AND APPLICATIONS IN BIOSCIENCE

CHUANG XU[1], YONGZHEN PEI[2,*], CHANGGUO LI[3]

[1]School of Computer Science and Technology, Tiangong University, Tianjin, 300387, China

[2]School of Mathematical Sciences, Tiangong University, Tianjin, 300387, China

[3]Department of Basic Science, Army Military Transportation University, Tianjin, 300161, China

Abstract. Background and Objective: Approximate Bayesian computation (ABC), identifying the parameters that yield simulated data resembling the observed data, is a powerful likelihood-free inference framework, and has been widely applied in bioscience including population model, epidemic model and so on. A major difficulty in ABC is how to accurately determine the level of the discrepancy for it has a crucial impact on the inference results of algorithms. In this paper, our aim is to propose a novel and valid discrepancy measure approach. Methods: By analyzing and comparing existing discrepancy measure methods, one finds they have obvious shortcomings including narrow adaptability and high computational cost. To overcome these deficiencies, an improved cosine similarity to assess the discrepancy in ABC is designed. First, both simulated data and observed data are converted into vectors, and then in virtue of the angle of them and the modulus of vectors, the similarity between the two data sets are measure. Results: The proposed discrepancy measure method achieves a comparable or higher posterior quality compared with the existing methods through four examples including Gaussian, Gaussian mixture, predator-prey and epidemic models. Conclusions: The statistical inference of complex biological models is a challenging task, and ABC algorithm can solve this problem well, but it needs to choose appropriate discrepancy measures. The results show that the improved cosine similarity is a extremely efficient discrepancy measurement method. Moreover, our work facilitates researchers to choose an appropriate discrepancy measure in practice.

[*]Corresponding author

E-mail address: yzhpei@tiangong.edu.cn

Received March 6, 2020

Keywords: likelihood-free inference; approximate Bayesian computation; KL divergence; Wasserstein distance; cosine similarity.

2010 AMS Subject Classification: 92B15.

1. INTRODUCTION

With the increasing computing power of computers in recent years, approximate Bayesian computation (ABC) has played a prominent role in statistical inference during the past two decades. ABC is widely applied into evolutionary biology, ecology, epidemiology, economics, bioinformatics and other disciplines [1–6]. Suppose that there is a model M and a observed data D_{obs} determined by the parameters θ. The posterior distribution of the parameters θ can be calculated by Bayes' rule:

$$\pi(\theta|D_{obs}) = \frac{L(D_{obs}|\theta)\pi(\theta)}{\int L(D_{obs}|\theta)\pi(\theta)} \propto L(D_{obs}|\theta)\pi(\theta), \tag{1}$$

where $\pi(\theta)$ and $\pi(\theta|D_{obs})$ are the prior and posterior respectively, and $L(D_{obs}|\theta)$ denotes the likelihood function-the probability of the observed data given some parameter value [7]. The likelihood function is crucial in statistical inference [8], which affects the evaluation of posterior distribution. Likelihood functions can be deduced in some simple models, but it is often computationally intractable or too costly to evaluate for more complex models, and standard methods of Bayesian estimation can not obtain the posterior distribution of the particular parameters in this situation. Based on Bayes' theorem, Simon Tavare et al. [9] introduced the Approximate Bayesian computation (ABC) method for the first time. By comparing the number of segregating sites in the simulated and real DNA sequence data, they determined the acceptance parameters to infer the posterior distribution of the time to the most recent common ancestor of the sampled individuals. Their pioneering work laid the foundation for likelihood-free inference.

ABC bypasses the calculation of the likelihood by identifying the parameters in parameter space that can generate data very similar to the observed data, and the similarity between the two data sets is evaluated by the discrepancy measure. In this paper, we focus on how to assess the similarity, because it has a crucial impact on the inference results of ABC algorithms [10]. At first, the distance function (such as the L1 distance, Euclidean distance) was often employed to measure the discrepancy in ABC. For example, Yan Wang et al. [11] took advantage of Prairie

Grass field observations to evaluate Bayesian statistical methods for source term estimation, and compared the performances of six different distance measures. ABC usually makes use of summary statistics (such as sample moments) to collect information from data sets, then calculates the distance in the summary statistics space, and the quality of the inference depends on the selected summary statistics. Recently, Michael U. Gutmann et al. [12] found that classification methods can be used to assess the similarity: Two data sets are judged maximally similar if their classification accuracy is close to 50%. But, if the classification accuracy is close to 100%, it means that they are far from each other. They refered to this approach as classifier ABC, and classifier ABC provides a new idea for how to compare the simulated data with the observed data. Moreover, it is a fundamental difficulty to construct effective summary statistics, then Michael A. Irvine et al. [13] proposed KDE ABC method which utilizes kernel-density estimation(KDE) to approximate the probability distributions of the simulated and observed data, and measures the distance between the two approximated distributions by Kullback-Leibler(KL) divergence. This method avoids the construction of summary statistics. They demonstrated that KDE ABC is a potentially powerful tool in model fitting for epidemiological data. Similar to the KL divergence, Espen Bernton et al. [14] utilized the Wasserstein distance between the empirical distribution of the two data sets as a difference measure in ABC. However, these discrepancy measures may have drawbacks. The distance function may not accurately measure the distance between the two data sets because of the use of insufficiency summary statistics. Although Classifier ABC and KDE ABC avoid the use of summary statistics and the loss of information, they admit the disadvantages of high computational cost and narrow application range. In addition, the dependent data is common in the application of ABC methods, such as time series. The usual approach is to transform the time series so that empirical distributions can be defined, but these methods may also lead to a poor quality of the inference.

In this paper, we make use of an improved cosine similarity to assess the discrepancy. This method needs to convert the data sets into vectors, and considering both the angle and the modulus of vectors to measure the similarity between the two data sets. Experiments show that this method has good inference results. In order to verify the accuracy of the estimation results of the discrepancy measures, the observed data are usually replaced by the data generated by known parameter values, and then the inferred results are compared with the known parameter values [15]. We also employ two biological models to compare the performance of the improved cosine similarity and other discrepancy measurement methods. Further, we intuitively show the performance of five discrepancy measures by establishing the correlation between the parameters and the discrepancy (the Euclidean distance, classification accuracy, KL divergence, Wasserstein distance and improved cosine similarity). That is, the discrepancy measures chosen should be able to capture the small variations of parameters more effectively, and these correlations might be used to choose the most appropriate discrepancy metric [16]. In order to facilitate researchers to choose appropriate metrics in practice, this paper also compares the accuracy, stability and efficiency of five discrepancy measures through four examples.

2. METHODS

2.1. Overview of ABC methodology.

ABC algorithms sample from the posterior distribution by accepting candidate parameter values that can generate data sufficiently resembling the observed data [17]. The basic form of ABC methods is the ABC rejection algorithm. Specifically, let θ be the parameter vector to be estimated. A candidate parameter vector $\hat{\theta}$ is sampled from the prior $\pi(\theta)$, and a data set $\hat{D}$ is generated from the specific model described by a conditional distribution $L(D|\hat{\theta})$. We accept $\hat{\theta}$ if the observed data D_{obs} is equal to the simulated data $\hat{D}$. However, this acceptance criterion will rejects almost all parameter values because the probability that the simulated data equals the observed data is very small in practice. Therefore, a appropriate metric ρ (such as Euclidean distance) and threshold ($\varepsilon \geq 0$) are usually chosen to determine the level of discrepancy between $\hat{D}$ and D_{obs}, and $\hat{\theta}$ is accepted if $\rho(\hat{D}, D_{obs}) \leq \varepsilon$. The ABC rejection algorithm is as follows:

The outcome of the ABC rejection algorithm is a sample which actually comes from the distribution $\pi(\theta|\rho(\hat{D}, D_{obs}) \leq \varepsilon)$, not from the posterior distribution $\pi(\theta|D_{obs})$, where

$$\pi(\theta|\rho(\hat{D}, D_{obs}) \leq \varepsilon) \propto Pr(\rho(\hat{D}, D_{obs}) \leq \varepsilon)\pi(\theta). \tag{2}$$

$\pi(\theta|\rho(\hat{D}, D_{obs}) \leq \varepsilon)$ will be a good approximation of the posterior distribution $\pi(\theta|D_{obs})$ under the condition of the enough small threshold and the reasonable distance metric [17] .

Algorithm 1 ABC rejection algorithm

```
for $i = 1$ to $N$ do
    repeat
        Sample $\hat{\theta}$ from the prior $\pi(\theta)$
        Simulate a data set $\hat{D}$ from $L(D|\hat{\theta})$
    until $\rho(\hat{D}, D_{obs}) \leq \varepsilon$
    set $\theta_i = \hat{\theta}$
end for
```

This method has two-fold approximations including the distance metric and threshold. On the one hand, researchers choose a discrepancy measurement subjectively basing on expert knowledge of the observed data [18]. The discrepancy metric used is critical for the success of the statistical inference. We will focus on the influence of different discrepancy measures on the inference results of ABC methods. On the other hand, the estimation result will be very poor if the threshold ε is too large; for the very small values of ε, the acceptance rate can be dramatically low, and which will greatly reduce the computational efficiency of the algorithm [19]. Therefore, on a very essential level, ABC has a trade-off between computational efficiency and statistical efficiency [1].

Summary statistics $s(D)$ are often employed to capture the relevant information about θ when the data is continuous or high-dimensional [20], which can greatly improve the computational efficiency of ABC. A review of the selection of summary statistics is given in [21, 22]. Then, the acceptance criteria is modified as follows: $\rho(s(\hat{D}), s(D_{obs})) \leq \varepsilon$.

As mentioned above, the ABC rejection algorithm is basically an experimental scheme in which the proposal parameter values are sampled from the prior. The acceptance rate is low when there is a great difference between the prior and posterior. In order to overcome this disadvantage and improve the computational efficiency of the ABC rejection algorithm, Markov Chain Monte Carlo(MCMC) sampling has been embedded in the ABC framework (ABC-MCMC) [20, 23]. The ABC-MCMC algorithm obtains a Markov chain $\{\theta_0, \theta_1, ..., \theta_N\}$ from the stationary distribution $\pi(\theta|\rho(\hat{D}, D_{obs}) \leq \varepsilon)$. Although this algorithm improves the acceptance rate, and it brings two potential disadvantages. Firstly, ABC-MCMC algorithm generates

Algorithm 2 ABC-SMC algorithm

```
Initialize $\varepsilon_1, ..., \varepsilon_T$
for $t = 0$ to $T$ do
  for $i = 1$ to $N$ do
    repeat
      if t==0 then
        Sample $\theta^{**}$ independently from $\pi(\theta)$
      else
        repeat
          Sample $\theta^{*}$ form $\{\theta_{t-1}^{(i)}\}$ with weights $\omega_{t-1}$
          $\theta^{**} \sim K_t(\theta|\theta^{*})$ where $K_t$ is a perturbation kernel
        until $\pi(\theta^{**}) \neq 0$
      end if
      $\hat{D} \sim p(D|\theta^{**})$
    until $\rho(\hat{D}, D_{obs}) \leq \varepsilon_t$
    set $\theta_t^{(i)} = \theta^{**}$ and calculate the weight for particle $\theta_t^{(i)}$
        $w_t^{(i)} = \begin{cases} 1 & \text{if t=0} \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^{N} \omega_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}|\theta_t^{(i)})} & \text{if t>0} \end{cases}$
  end for
  Normalize the weights
end for
```

a sequence of highly dependent samples from $\pi(\theta|\rho(\hat{D}, D_{obs}) \leq \varepsilon)$ [24]. The second disadvantage is that the chain may be stuck in the low probability region for a long time if the proposal distribution is poorly chosen [25].

Sisson et al. [24] developed a new ABC method based on sequential Monte Carlo (SMC)(called ABC-SMC), which avoided the disadvantages of the ABC rejection and ABC-MCMC methods in part. In ABC-SMC, a set of tolerances $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_T\}$ is chosen such that $\varepsilon_1 > ... > \varepsilon_T \geq 0$. Firstly, N parameter values(called particles) $\{\theta_0\} = \{\theta_0^{(1)}, \theta_0^{(2)}, ..., \theta_0^{(N)}\}$ are sampled from the prior $\pi(\theta)$. ABC-SMC is an iterative algorithm which obtains weighted samples $\{\theta_t\}$ from a sequence of intermediate distributions $\pi(\theta|\rho(\hat{D},D_{obs}) \leq \varepsilon_t)$ by weighting the accepted parameters $\theta_t^{(i)}$ with

$$w_t^{(i)} = \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^{N} \omega_{t-1}^{(j)} K_t(\theta_{t-1}^{(j)}|\theta_t^{(i)})}, i = 1, ..., T-1. \tag{3}$$

As the tolerance ε_t decreases, $\pi(\theta|\rho(\hat{D},D_{obs}) \leq \varepsilon_t)$ gradually approaches the posterior. Finally, ABC-SMC samples from the distribution $\pi(\theta|\rho(\hat{D},D_{obs}) \leq \varepsilon_T)$ to approximate the posterior distribution. The process of the algorithm is shown in Algorithm 2. ABC-SMC reduces the computational time compared with the ABC rejection and ABC-MCMC methods. But these ABC algorithms are all looking for parameters whose corresponding generated data have a very small discrepancy with the observed data. Next, we discuss how to define the discrepancy between the two data sets.

2.2. **Discrepancy measures.**

It can be seen that the statistical efficiency of ABC methods depends largely on how to accurately determine the level of the discrepancy between the simulated and observed data. In this section, we introduce five discrepancy measures that can be used in ABC.

2.2.1. *Basic ABC.*

The distance functions, such as the Euclidean distance and the Manhattan distance, are most often used to measure the discrepancy between the simulated and observed data in ABC. If the data set is replaced by summary statistics and the discrepancy is measured by a metric in the summary statistical space, the quality of the inference depends on the summary statistics chosen [17]. It is impossible to identify sufficient summary statistics if the likelihood function is unknown, so the use of summaries adds another approximation [10]. For some models and the observed data, the distance function may not accurately measure the discrepancy between the two data sets, because the use of non-sufficient statistics may lead to loss of information. In the following example, we refer to the ABC method, which makes use of the Euclidean distance between the summary statistics as the discrepancy measure, as Basic ABC.

2.2.2. *Classifier ABC.* Michael U. Gutmann et al. [12] found that classification methods can be applied to measure the discrepancy, and thus to perform likelihood-free inference. Intuitively, two data sets generated with two different parameter values are easier to distinguish than those generated with two similar parameter values. Classification methods are performed on feature vectors that extract information from data. Suppose that $X = \{x_1, x_2, ..., x_n\}$ is the feature vectors from the observed data and $Y = \{y_1, y_2, ..., y_n\}$ is the feature vectors from the simulated data. Each data point in X is set as a positive instance point, and each data point in Y is set as a negative instance point. Therefore, the augmented data set is $T = \{(x_1, 1), ..., (x_n, 1), (y_1, -1), ..., (y_n, -1)\}$. The augmented data set T is used to learn the classification rule R, and $R(z) \in \{1, -1\}$ ($z = x_i$ or y_i). Then the classification accuracy is

$$CA(R,T) = \frac{1}{2n}\{\sum_{i=1}^{n}(\frac{1+R(x_i)}{2} + \frac{1-R(y_i)}{2})\}. \tag{4}$$

Obviously, the closer the classification accuracy CA is to 0.5, the more difficult the two data sets are to distinguish. The closer the classification accuracy is to 1, the easier the data sets are to distinguish. That is, the closer CA is to 0.5, the smaller the discrepancy between simulated and observed data, and the closer CA is to 1, the larger the discrepancy between them. In order to improve classification accuracy, K-fold cross-validation is usually used to reduce over-fitting to some extent. Classifiability is used as a discrepancy measurement in ABC, which is a data-driven approach to assess the similarity between the two data sets [12]. However, this method has the disadvantage of high computational cost, and its inference may be poor when the size of data sets is small.

2.2.3. *KDE ABC.* When we estimate the parameters of stochastic complex models, the heterogeneity of data is an intractable problem [13]. Additionally, the selection of some summary statistics and distance functions potentially includes the assumptions of unimodality and normality [26] [27], which may lead to the inaccurate estimation. Michael A. Irvine et al. [13] used the non-parametric kernel density estimation (KDE) to directly compare the simulated and observed data to resolve this problem. Suppose that the data sets is $D = \{x_1, x_2, ..., x_n\}$, and the empirical distribution $f(x)$ of D approximated by KDE is as follows:

$$f(x|D) = \frac{1}{n}\sum_{k=1}^{n} K(x,x_i), \tag{5}$$

where $K(x,x_i)$ is a kernel function. The approximate probability distributions of the simulated data $\hat{D}$ and observed data D_{obs} are expressed in $f(x|\hat{D})$ and $f(x|D_{obs})$ respectively. The Kullback-Leibler (KL) divergence is served to calculate the distance between the two approximate distributions. It is defined as:

$$D_{kl}(f(x|D_{obs})||f(x|\hat{D})) = \int_{-\infty}^{\infty} f(x|D_{obs}) log \frac{f(x|D_{obs})}{f(x|\hat{D})} dx. \tag{6}$$

The KL divergence is zero if the two distributions are equal, and it will also increase when the difference between the them increases. The method of combining ABC with kernel density estimation is referred to as KDE ABC. The disadvantage of KDE ABC is its high computational cost, and it is difficult to calculate integrals when the data is high-dimensional.

2.2.4. *Wasserstein ABC (WABC).* Wasserstein distance can be applied to evaluate similarity between the two data sets in ABC to avoid the use of summaries and the consequent loss of information, and it defines a metric in the probability distributions space [14, 28]. Let ρ be a distance function on $\chi \subseteq \Re^d$. The q-Wasserstein distance between two distributions μ and ν is defined as

$$Dw_q(\mu,\nu) = (\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\chi\times\chi} \rho(x,y)^q d\gamma(x,y))^{\frac{1}{q}}, \tag{7}$$

where $\Gamma(\mu,\nu)$ is the set of all joint probability distribution γ on $\chi\times\chi$ with marginals μ and ν. We denote this discrepancy by $Dw_q(X,Y)$ based on the observed data $X = \{X_i\}_{i=1}^n$ and the simulated data $Y = \{Y_i\}_{i=1}^m$, and the form is

$$Dw_q(X,Y) = (\inf_{\gamma} \sum_{i=1}^{n}\sum_{j=1}^{m} \rho(X_i,Y_j)^q \gamma_{ij})^{\frac{1}{q}}, \tag{8}$$

where γ is a $n\times m$ non-negative matrix with rows summing to $\frac{1}{n}$ and rows summing to $\frac{1}{m}$. We only focus on the case $n = m$ in ABC, where per row and column of matrix γ only have one non-zero entry which equal to $\frac{1}{n}$. In particular, the 2-Wasserstein distance between X and Y takes the form $(\frac{1}{n}\sum_{i=1}^{n}|X_{(i)} - Y_{(i)}|^2)^{\frac{1}{2}}$ if $d = 1$ and $\rho(x,y) = |x-y|$.

2.2.5. *Cosine ABC.* We find that the improved cosine similarity can be served as a discrepancy measurement. Cosine similarity is to evaluate the similarity of two vectors by calculating their cosine value. The more similar the two vectors are, the closer their cosine value is to 1, and the closer their included angle is to 0 degree. In order to use cosine similarity in ABC, it is necessary to transform the simulated and observed data sets into vectors, which are represented by A and B, respectively. Then, suppose that $A = [x_1, x_2, ..., x_n]^T$, $B = [y_1, y_2, ..., y_n]^T$. The formula of cosine similarity is as follows:

$$Cos(A,B) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}. \tag{9}$$

For convenience, we make use of arccosine function $arccos(\frac{A \cdot B}{\|A\|\|B\|})$ to measure the similarity of two vectors, then the more similar two vectors are, the closer their included angle is to 0. However, cosine similarity measures the similarity of two vectors from the direction, and it is insensitive to numerical values. In order to consider the influence of numerical values, we calculate the absolute value of the difference between the modulus of two vectors, i.e. $|\|A\| - \|B\||$, and then combine $|\|A\| - \|B\||$ and $arccos(\frac{A \cdot B}{\|A\|\|B\|})$ together to form a discrepancy measurement. We know that the angle range between two vectors is $[0, \pi]$, and the angle can be ignored if the value of $|\|A\| - \|B\||$ is too large. So we narrow down $|\|A\| - \|B\||$ by dividing $\|A\|$. Therefore, the formula of the improved cosine similarity is as follows:

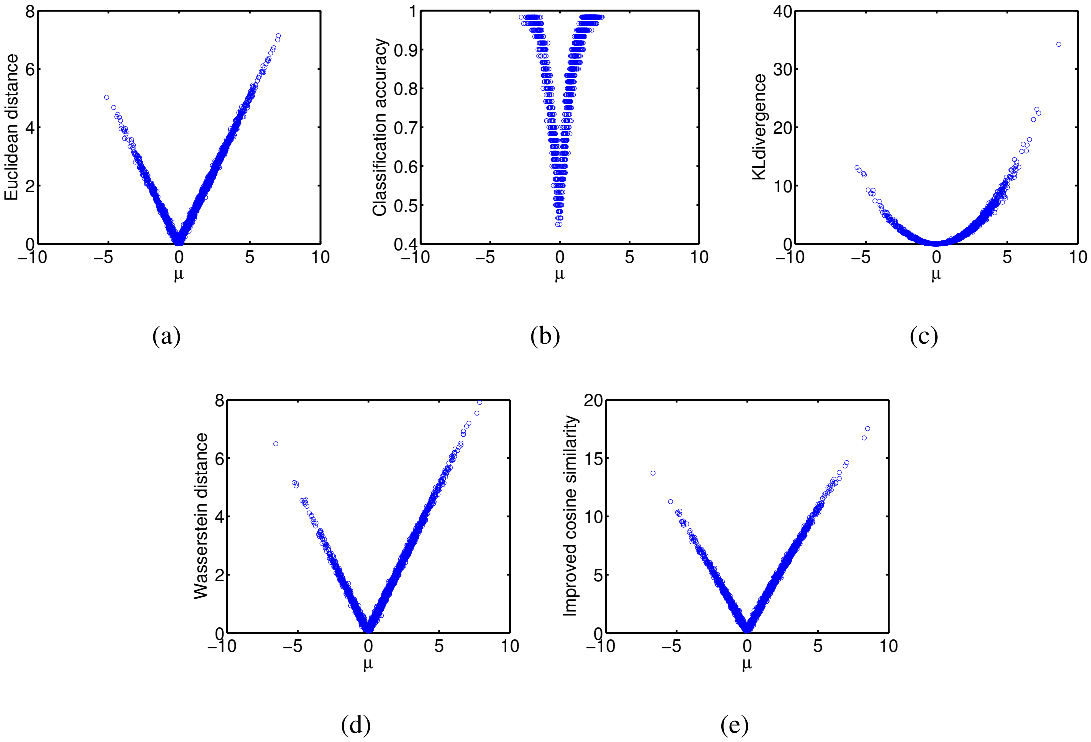
$$ImpCos(A,B) = arccos(\frac{A \cdot B}{\|A\|\|B\|}) + \frac{|\|A\| - \|B\||}{\|A\|}. \tag{10}$$

The improved cosine similarity takes into account both the direction of vectors and the modulus of vectors, and it can be served to calculate the discrepancy between the two data sets in ABC. We refer to the ABC method combining the improved cosine similarity as Cosine ABC, and experiments show that this method visibly outperforms other methods.

3. Results

The experiments contain four examples which were used to compare five discrepancy measures. The first two examples are toy models, and the observed data are generated with known parameters, and the second two examples are biological models. In this section, we describe the relation between the parameters and the discrepancy (the Euclidean distance, classification accuracy, KL divergence, Wasserstein distance and improved cosine similarity)to intuitively compare the performance of five discrepancy measures. For specific models, the chosen discrepancy measure should be able to reflect the correlation between the parameter values and the discrepancy, that is, the closer the candidate parameters extracted from the prior are to the true parameters, the smaller the discrepancy is.

3.1. Gaussian Models. Five discrepancy measures are compared by using a simple normal distribution $N(\mu,\sigma^2)$. Suppose that the parameter of interest is μ, and we set $\sigma = 0.5$. The observed data D_{obs} of size 30 are generated with $\mu = 0$. In order to calculate the true posterior distribution conveniently, the normal distribution $N(1,2^2)$ is chosen as the prior. According to the conjugate prior distribution, the posterior distribution is also normal distribution and it can be derived directly.



(a)

(b)

(c)

(d)

(e)

FIGURE 1. The relation between the parameter values and the discrepancy.

Firstly, we visualize the relation between five discrepancy measures and the parameter μ in Figure 1. Specifically, we extract 2000 candidate parameter values from the prior and simulate the corresponding data with these values, and then utilize five measures to calculate the distance. Figure 1 indicates that these five methods all can accurately measure the discrepancy in this simple toy model. That is, the corresponding discrepancy decreases as the candidate parameter values approach the true parameters. Additionally, this figure also provides guidance for choosing the appropriate tolerance.

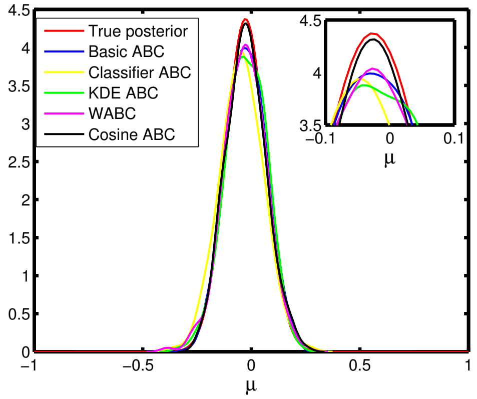


FIGURE 2. The approximate posterior distributions estimated by five algorithms and the true posterior distribution (red line).

We perform the five algorithms and retain 500 particles. The average value of reserved particles is as the result of estimating parameters (Table 1). Table 1 shows that the estimation inferred by five methods are very close to the true values. However, Basic ABC, WABC and Cosine ABC are much faster than KDE ABC and Classifier ABC, because the calculation of KL divergence and classification accuracy takes a great quantity time, so the computational cost of these two methods may be higher. For example, when we perform 100,000 simulations, Basic ABC, WABC and Cosine ABC only take about 2.11, 2.58 and 4.18 seconds respectively, while KDE ABC and Classifier ABC take about 121.13 and 478.53 seconds respectively. Approx-

TABLE 1. Inference results of five methods

methods	the results of estimating parameter
Basic ABC	-0.0206
Classifier ABC	-0.0446
KDE ABC	-0.0229
WABC	-0.0304
Cosine ABC	-0.0151

imate posterior distributions estimated by five methods and the true posterior distribution are shown in Figure 2. The red line stands for the true posterior distribution, and the blue, yellow, green, magenta and black lines are inferred by Basic ABC, Classifier ABC, KDE ABC, WABC and Cosine ABC respectively. The black lines is closer to the true posterior, which indicates that the Cosine ABC obtain comparably high quality approximate posterior in this example.

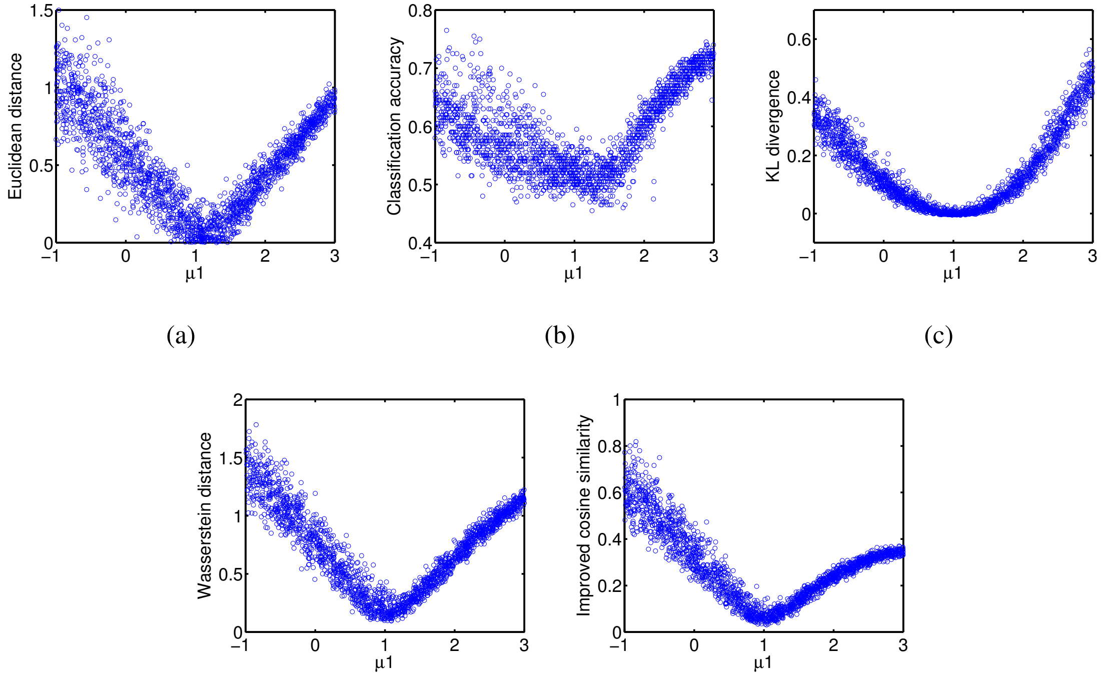
3.2. Gaussian mixture model. We make use of a Gaussian mixture model to test the performance of five methods. For simplicity, Suppose that the model is the mixture of two Gaussian distributions. The model is

$$P(y|\theta) = \alpha_1 P(y|\theta_1) + \alpha_2 P(y|\theta_2), \tag{11}$$

where $P(y|\theta_k)$ is the Gaussian density function, $\alpha_1 + \alpha_2 = 1$, $\theta_k = (\mu_k, \sigma_k^2)(k = 1, 2)$,

$$P(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} exp(-\frac{(y-\mu_k)^2}{2\sigma_k^2}). \tag{12}$$

We fixed $\alpha_1 = \alpha_2 = 0.5$, $\sigma_1 = \sigma_2 = 0.5$, $\mu_2 = 3$, and the parameter of interest is μ_1. The



(a)

(b)

(c)

(d)

(e)

FIGURE 3. The relation between the parameter values and the discrepancy.

synthetic data y_{obs} of size 100 generated with $\mu_1 = 1$ was served as observation data. The prior for μ_1 is taken to be uniform, $\mu_1 \sim U(-1,3)$. According to Bayes' rule, the posterior is proportional to the likelihood function

$$L(y_{obs}|\mu_1) = \prod_{i=1}^{n} \pi(y_i|\mu_1), \tag{13}$$

where

$$\pi(y_i|\mu_1) = \frac{1}{2\sqrt{2\pi}\sigma_1} exp(-\frac{(y_i-\mu_1)^2}{2\sigma_1^2}) + \frac{1}{2\sqrt{2\pi}\sigma_2} exp(-\frac{(y_i-\mu_2)^2}{2\sigma_2^2}) \tag{14}$$

and $y_{obs} = \{y_1, y_2, ..., y_n\}(n = 100)$. So, the posterior distribution $\pi(\mu_1|y_{obs}) = kL(y_{obs}|\mu_1)$, where k is the proportional coefficient.

In this example, Basic ABC takes the Euclidean distance as a measure of discrepancy and sample average as summary statistics. Figure 3 is obtained in the same way as the first example, and it indicates that classification methods and the Euclidean distance with insufficient statistics do not accurately determine the differences between the simulated and observed data. However, the improved cosine similarity visibly outperforms other methods, followed by Wasserstein distance and KL divergence. The approximate posteriors obtained by five methods and the true posterior distribution are shown in Figure 4. The experimental results indicate that the black line is closest to the red line, that is, Cosine ABC produces better inference, and the improved cosine similarity acted as a discrepancy measure has an overall satisfactory performance. Table 2 shows the average values of particles obtained by the five methods, and Basic ABC and classifier ABC give poor inferences.

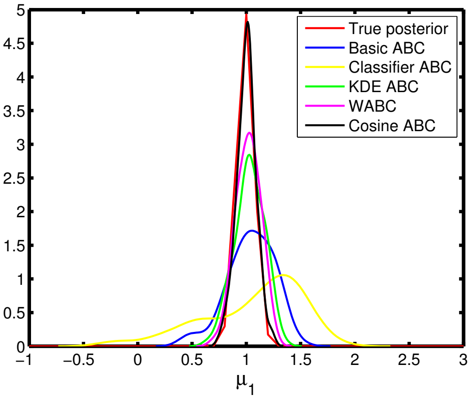


FIGURE 4. The approximate posterior distributions estimated by five algorithms and the true posterior.

TABLE 2. Inference results of five methods

methods	the results of estimating parameter
Basic ABC	1.0412
Classifier ABC	1.0645
KDE ABC	1.0283
WABC	1.0170
Cosine ABC	1.0010

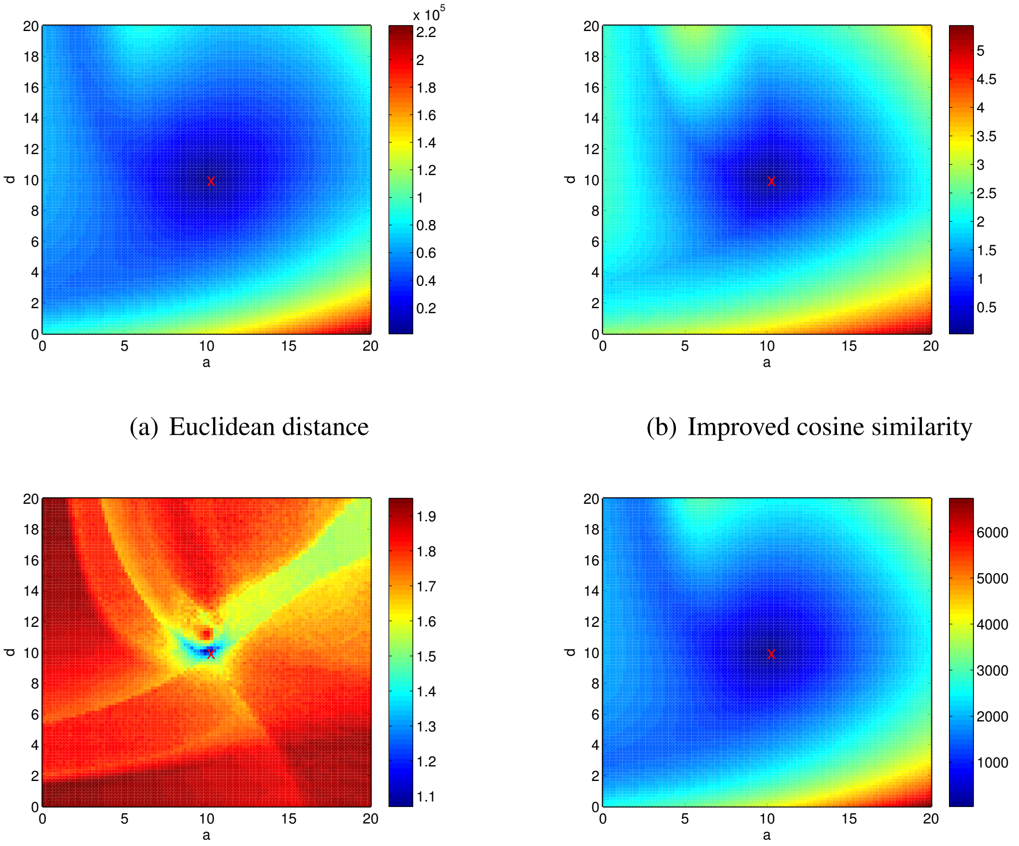
3.3. Lotka-Volterra (LV) model.

Volterra first proposed Lotka-Volterra (LV) model for the predation of one species by another to explain the oscillatory levels of certain fish catches in the Adriatic [29,30]. Now it will be recruited to verify the improved cosine similarity and compare with other discrepancy measures. The interaction between prey species(X) and predator species(Y) can be described by the following differential equations:

$$\begin{aligned} \frac{dX}{dt} &= aX - bXY, \\ \frac{dY}{dt} &= cXY - dY. \end{aligned} \tag{15}$$

where a, b, c and d are positive constants. The aX term means that the prey grows in a Malthusian way without any predation. The $-bXY$ term reflects the effect of predation. The prey's contribution to the predators' growth rate is cXY. The $-dY$ term means that the mortality rate of predators decreases exponentially in the absence of any prey.

The synthetic data are obtained by solving the equations with $(a,b,c,d) = (10, 0.01, 0.01, 10)$ at initial conditions $(X_0, Y_0) = (1100, 990)$, and Gaussian noise $N(0, 10^2)$ is added to the data points. We estimate the parameters a and d, keeping b and c fixed at the value which was generated the synthetic data. The prior is taken to be uniform, $a \sim U(0,20)$ and $d \sim U(0,20)$. In this example, the data points are large and scattered, which is not suitable for KDE ABC. We extract 10,000 candidate parameter values from the parameter space at equal intervals, and wield these values to generate corresponding simulation data from the model, then calculate the discrepancy between the simulated and observed data with the Euclidean distance, improved cosine similarity, classification accuracy and Wasserstein distance respectively. The relation between the parameters and the discrepancy is depicted in a heat map (Figure 5), which exposes that all four discrepancy measures are minimal when the parameters are close to the true parameters.



(a) Euclidean distance

(b) Improved cosine similarity

(c) Classification accuracy

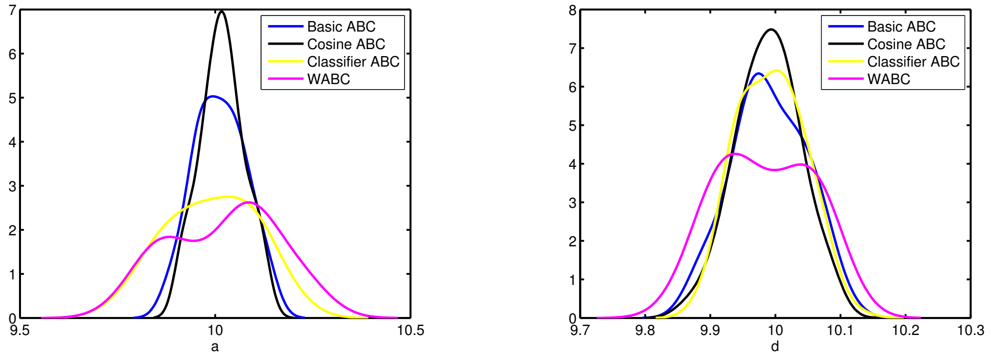
(d) Wasserstein distance

FIGURE 5. The heat map of the relation between the parameters and the discrepancy. The red crosses mark the data-generating parameter value.

TABLE 3. Inference results of three methods

methods	a	d
Basic ABC	10.0123	9.9905
Cosine ABC	10.0133	9.9987
Classifier ABC	10.0308	9.9911
WABC	10.0204	9.9844

Next the ABC-SMC algorithm was employed to estimate parameters, and the Euclidean distance, improved cosine similarity, classification accuracy and Wasserstein distance were served as the discrepancy measures respectively. We set $\varepsilon = (10000, 5000, 2000, 1500), (0.4, 0.1, 0.07, 0.04), (1.5, 1.3, 1.2, 1.1), (300, 100, 50, 27)$ respectively and retain 200 particles to obtain the approximate posteriors of parameters a and d (Figure 6). Simulations suggest that Cosine ABC



(a) (b)

FIGURE 6. The approximate posterior distributions of parameters a and d.

yields the approximate posteriors that have narrower regions and are more closely centered around the true parameter values than those produced by other methods [31]. Table 3 takes the average value of the retained particles as the estimation of the parameters, which indicates that the estimation results of the four methods are all very close to the true values.

3.4. Stochastic SEIR model.

Infectious diseases have tremendous influence on human life. Every year, millions of people die of various infectious diseases such as West Nile Virus, Dengue, Zika, Ebola and so on. Mathematical modelling is of considerable importance in the study of epidemiology because it may provide understanding of the underlying mechanisms which influence the spread of disease. Various epidemic models have been formulated and analyzed. The stochastic SEIR model describing the spread of a non-lethal disease in a large population is based on a partition of the total population into four classes: susceptible (S), exposed (infected but not yet infectious) (E), infectious (I), recovered (R) and $N = S+E+I+R$ [32]. The process can be described by the following rate equations:

$$\begin{aligned} & S+I \to E+I \; at\ rate\ \frac{\lambda SI}{N}, \\ & E \to I \; at\ rate\ \delta E, \\ & I \to R \; at\ rate\ \mu I. \end{aligned} \tag{16}$$

The parameter λ is the infection rate, μ denotes the rate for recovery, and δ is the rate at which the exposed individuals become infective.

Gillespie algorithm was executed to simulate data under given parameter values in this model [29, 33]. In this example, 21 data points are acquired by Gillespie algorithm with $(\lambda,\delta,\mu) = (2,4,0.3)$ at initial conditions $(S_0,E_0,I_0,R_0) = (198,1,1,0)$. We infer the parameters λ, δ and μ, and uniform priors including $\lambda \sim U(0,10)$, $\delta \sim U(0,10)$ and $\mu \sim U(0,1)$ are adopted. Because of the randomness of Gillespie algorithm simulation process, the average value of multiple runs can better reflect the mean dynamic of the system than that of one run. So in lab settings, the simulated data of each run are averaged in three runs [25]. KDE ABC is not suitable for solving this problem because the data points are large and scattered, which is also the defect of this method.

Here we perform four ABC methods including Basic ABC, Cosine ABC, Classifier ABC and WABC, and set $\varepsilon = (500,300,200,100,60), (0.8,0.2,0.15,0.1,0.08), (0.7,0.6,0.55,0.5), (40, 30,20,12,8)$ respectively. We retain 200 particles to obtain the histograms of the approximate posteriors of parameters λ, δ and μ displayed in Figure 8. The inference of parameters λ and μ obtained by Cosine ABC are better than other methods. However, none of the four methods can infer parameter δ. Perhaps because δ is insensitive. The relation between the parameter and the distance is described to verify the sensitivity of δ in the case of fixing parameters $\lambda = 2$ and $\mu = 0.3$. Figure 7 indicates that parameter δ is insensitive because the distance does not change near $\delta = 4$.

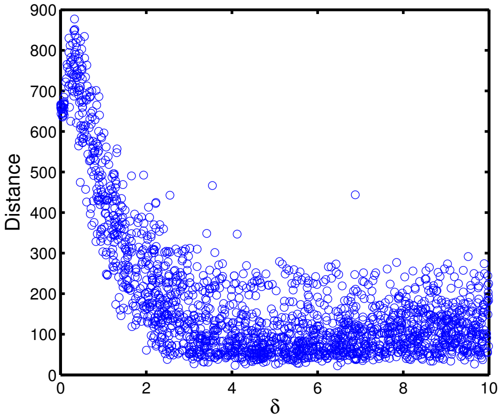
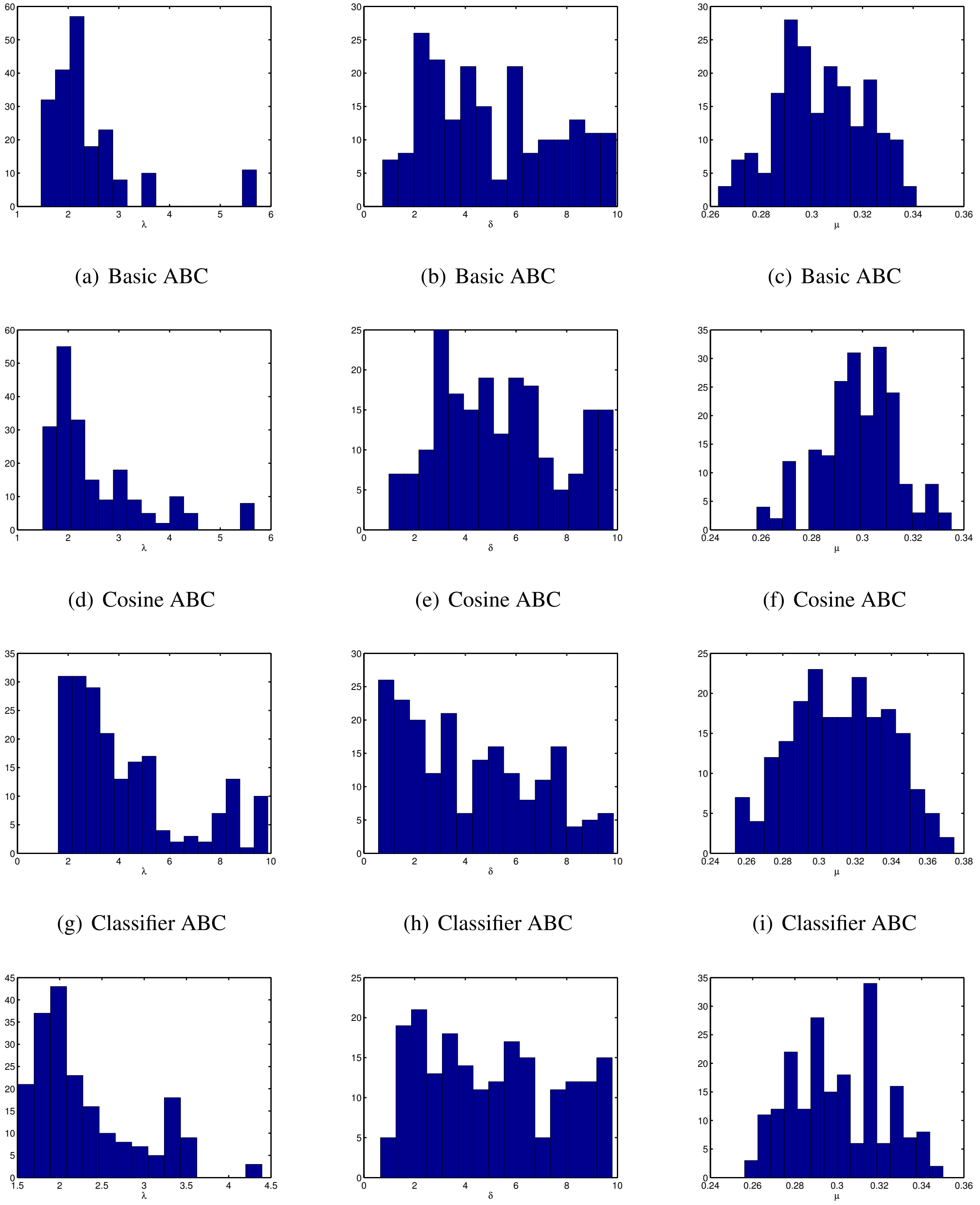


FIGURE 7. The relation between the parameter δ and the distance.



(a) Basic ABC (b) Basic ABC (c) Basic ABC

(d) Cosine ABC (e) Cosine ABC (f) Cosine ABC

(g) Classifier ABC (h) Classifier ABC (i) Classifier ABC

(j) WABC (k) WABC (l) WABC

FIGURE 8. The histograms of the approximate posterior distributions of parameters λ, δ and μ.

4. Discussion

The statistical inference of complex models is an challenging task because the likelihood functions is usually intractable. Approximate Bayesian computation (ABC) is a powerful likelihood-free inference method, and the key step of ABC is to identify parameters by calculating the discrepancy between the simulated data and observed data. In this paper, we add the improved cosine similarity to the arsenal of the discrepancy measures for ABC, and compare it with other four data discrepancy measures. We analyze the advantages and disadvantages of these five methods and check their performance through four examples.

In terms of the approximate posterior quality, the improved cosine similarity performs comparably better than other discrepancy measures methods. In the second example, classification methods and the Euclidean distance using summary statistics can not accurately measure the differences, and their inference is poor. Research suggests that the kernel density estimation and the improved cosine similarity can capture more data information, so better inference results are obtained by KDE ABC and Cosine ABC.

Except for the quality of approximate posterior distribution, another problem to be considered in ABC is the computational efficiency of the discrepancy function. KDE ABC and Classifier ABC have obvious shortcomings including narrow adaptability and high computational cost, that is, the calculation of KL divergence and classification accuracy takes a great quantity time. Cosine ABC has a wide range of applications and low computational costs, and our research indicates that it is a very effective discrepancy measurement method.

However, for some problems with poor estimation results, we do not know whether the error is due to the discrepancy measurement chosen or the inaccuracy of the model. Moreover, ABC is still limited by large data sets and high-dimensional parameters, which is a difficult problem to be solved.

5. Conclusions

Approximate Bayesian calculation is an effective statistical inference method, which is very suitable for parameter estimation of biological model. The discrepancy measure is an essential part of ABC algorithm, which has a great impact on the results of the inference. Various discrepancy measurement methods are discussed in this paper. Therefore, our work is of practical value, which enlightens us to flexibly choose the measure of discrepancy in practice.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] M.A. Beaumont, W. Zhang, D.J. Balding, et al. Approximate Bayesian Computation in Population Genetics. Genetics, 162 (4) (2002), 2025-2035.

[2] E. Bazin, K.J. Dawson, M.A. Beaumont, et al. Likelihood-Free Inference of Population Structure and Local Adaptation in a Bayesian Hierarchical Model. Genetics, 185 (2) (2010), 587-602.

[3] O. Ratmann, O. Jorgensen, T. Hinkley, et al. Using Likelihood-Free Inference to Compare Evolutionary Dynamics of the Protein Networks of H. pylori and P. falciparum, PLoS Comput. Biol. 3 (2007), e230.

[4] T.J. Mckinley, J.V. Ross, R. Deardon, et al. Simulation-based Bayesian inference for epidemic models. Comput. Stat. Data Anal. 71 (2014), 434-447.

[5] S. Wood, Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466 (2010), 1102-1104.

[6] T. Liu, Y. Pei, C. Li, M. Ye, Amount of Escape Estimation Based on Bayesian and MCMC Approaches for RNA Interference, Mol. Ther. Nucleic. Acids. 18 (2019), 893-902.

[7] J. Cussens, Approximate Bayesian Computation for the Parameters of PRISM Programs, in: P. Frasconi, F.A. Lisi (Eds.), Inductive Logic Programming, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 38-46.

[8] Z. Meixia, P. Yongzhen, Y. Ming, L. Changguo, Quantitative evaluation of impacts of likelihood functions on Bayesian parametric estimation of epidemic models, Stat. Interface. 12 (2019), 415-422.

[9] S. Tavare, D.J. Balding, R.C. Griffiths, et al. Inferring Coalescence Times From DNA Sequence Data. Genetics, 145 (2) (1997), 505-518.

[10] R.D. Wilkinson, Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. Stat. Appl. Genet. Mol. Biol. 12 (2) (2013), 129-141.

[11] Y. Wang, H. Huang, L. Huang, B. Ristic, Evaluation of Bayesian source estimation methods with Prairie Grass observations and Gaussian plume model: A comparison of likelihood functions and distance measures, Atmos. Environ. 152 (2017), 519-530.

[12] M.U. Gutmann, R. Dutta, S. Kaski, J. Corander, Likelihood-free inference via classification, Stat. Comput. 28 (2018), 411-425.

[13] M.A. Irvine, T.D. Hollingsworth, Kernel-density estimation and approximate Bayesian computation for flexible epidemiological model fitting in Python, Epidemics. 25 (2018), 80-88.

[14] E. Bernton, P. E. Jacob, M. Gerber, C. P. Robert, Approximate bayesian computation with the wasserstein distance. J. R. Stat. Soc., Ser. B. (Stat. Methodol.) 81 (2) (2019), 235-269.

[15] J. Lintusaari, M.U. Gutmann, S. Kaski, J. Corander, On the Identifiability of Transmission Dynamic Models for Infectious Diseases, Genetics. 202 (2016), 911-918.

[16] V.C. Sousa, M. Fritz, M.A. Beaumont, L. Chikhi, Approximate Bayesian Computation Without Summary Statistics: The Case of Admixture, Genetics. 181 (2009), 1507-1519.

[17] J. Lintusaari, M.U. Gutmann, R. Dutta, S. Kaski, J. Corander, Fundamentals and Recent Developments in Approximate Bayesian Computation, Syst Biol. 66 (2017), e66-e82.

[18] J. Marin, P. Pudlo, C.P. Robert, et al. Approximate Bayesian computational methods. Stat. Comput. 22 (6) (2012), 1167-1180.

[19] M.A. Beaumont, Approximate Bayesian Computation in Evolution and Ecology. Ann. Rev. Ecol. Evol. Systemat. 41 (1) (2010), 379-406.

[20] P. Marjoram, J. Molitor, V. Plagnol, et al. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. USA. 100 (26) (2003), 15324-15328.

[21] S. Aeschbacher, M.A. Beaumont, A. Futschik, et al. A novel approach for choosing summary statistics in approximate Bayesian computation. Genetics, 192 (3) (2012), 1027-1047.

[22] D. Prangle, Summary Statistics in Approximate Bayesian Computation. arXiv:1512.05633 [stat.CO], 2015.

[23] D. Wegmann, C. Leuenberger, L. Excoffier, et al. Efficient Approximate Bayesian Computation Coupled With Markov Chain Monte Carlo Without Likelihood. Genetics, 182 (4) (2009), 1207-1218.

[24] S.A. Sisson, Y. Fan, M.M. Tanaka, et al. Sequential Monte Carlo without likelihoods. Proc. Nat. Acad. Sci. USA, 104 (6) (2007), 1760-1765.

[25] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M.P. Stumpf, Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface, 6 (31) (2009), 187-202.

[26] D.M. Walker, D. Allingham, H.W. Lee, et al. Parameter inference in small world network disease models with approximate Bayesian Computational methods. Physica A. 389 (3) (2010), 540-548.

[27] G.D. Brown, A.T. Porter, J.J. Oleson, J.A. Hinman, Approximate Bayesian computation for spatial SEIR(S) epidemic models, Spat. Spat.Temp. Epidemiol. 24 (2018), 27-37.

[28] B. Jiang, Approximate Bayesian Computation with Kullback-Leibler Divergence as Data Discrepancy. Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, PMLR. 84 (2018), 1711-1721.

[29] D.J. Wilkinson, Stochastic Modelling for Systems Biology. Chapman and Hall/CRC, Boca Raton, 2006.

[30] J.D. Murray, Mathematical biology. 2nd ed. Springer-Verlag, New York, 2002.

[31] C. Leuenberger, D. Wegmann, Bayesian Computation and Model Selection Without Likelihoods. Genetics, 184 (1) (2009), 243-252.

[32] M.Y. Li, J.R. Graef, L. Wang, J. Karsai, Global dynamics of a SEIR model with varying total population size, Math. Biosci. 160 (1999), 191-213.

[33] D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22 (4) (1976), 403-434.