eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

A Parzen-based distance between probability measures as an alternative of summary statistics in Approximate Bayesian Computation

Carlos D. Zuluaga, Edgar A. Valencia, Mauricio A. Álvarez

Faculty of Engineering, Universidad Tecnológica de Pereira, Colombia, 660003.

Abstract

Approximate Bayesian Computation (ABC) are likelihood-free Monte Carlo methods. ABC methods use a comparison between simulated data, using different parameters drew from a prior distribution, and observed data. This comparison process is based on computing a distance between the summary statistics from the simulated data and the observed data. For complex models, it is usually difficult to define a methodology for choosing or constructing the summary statistics. Recently, a nonparametric ABC has been proposed, that uses a dissimilarity measure between discrete distributions based on empirical kernel embeddings as an alternative for summary statistics. The nonparametric ABC outperforms other methods including ABC, kernel ABC or synthetic likelihood ABC. However, it assumes that the probability distributions are discrete, and it is not robust when dealing with few observations. In this paper, we propose to apply kernel embeddings using an smoother density estimator or Parzen estimator for comparing the empirical data distributions, and computing the ABC posterior. Synthetic data and real data were used to test the Bayesian inference of our method. We compare our method with respect to state-of-the-art methods, and demonstrate that our method is a robust estimator of the posterior distribution in terms of the number of observations.

1 Introduction

Many Bayesian applications use inference for nonlinear stochastic models, where it is expensive or difficult to evaluate the likelihood; or the normalization constant in Bayesian modelling is also intractable. Approximate Bayesian Computation (ABC) are likelihood-free Monte Carlo methods. ABC methods can be employed to infer posterior distributions without having to evaluate likelihood functions (Toni et al., 2009; Wilkinson, 2008). They simulate data from a model with different parameter values and compare summary statistics of the simulated data whit summary statistics of the observed data. There are many problems associated to how to choose the summary statistics with the goal of obtaining accepted samples. The choice of these summaries is frequently not obvious (Joyce and Marjoram, 2008) and in many cases it is difficult or impossible to construct a general method for finding such statistics (Fearnhead and Prangle, 2012). According to Park et al. (2015), the selection of the summary statistics is an important stage in ABC methods that is still an open question.

Different algorithms for choosing or constructing summary statistics have been proposed in the literature. In the state-of-the-art, we can find linear regression (Fearnhead and Prangle, 2012). Another option is to use a minimum entropy criterion for choosing the summary statistics (Blum et al., 2013; Blum, 2010; Nunes and Balding, 2010). Park et al. (2015) propose a fully nonparametric ABC that avoids to select manually the summary statistics, using kernel embeddings (Gretton et al., 2007a); they employ a two-step process: the first step is to compare the empirical data distributions using a dissimilarity distance based on Reproducing Kernel Hilbert Spaces (RKHS). Such a distance is termed Maximum Mean Discrepancy (MMD). In the second stage, they use another kernel that operates on probability measures. This nonparametric ABC outperforms other methods, ABC, kernel ABC or synthetic likelihood ABC, for more details see Park et al. (2015). However this nonparametric ABC assumes that the probability distributions are discrete and it is not robust when there are few observations.

In this paper, we propose a new metric for comparing two data distributions in a RKHS with application to ABC simulation. The difference with respect to Park et al. in Park et al. (2015) is the assumption that the probability density functions, for corresponding probability distributions in RKHS, are continuous probability functions, estimated using an smoother density estimator or Parzen estimator. This fact allows us to obtain a biased estimator of the dissimilarity distance between two empirical data distributions that can be written as $\lambda f + (1 - \lambda)\text{MMD}$, where $f \in \mathcal{H}$ ($\mathcal{H}$ is a Hilbert space), and

$0 \leq \lambda < 1$. It is possible to demonstrate that our estimator presents a lower root mean squared error, between the true parameters and the estimated parameter, than the MMD estimator, for any value of λ, see Muandet et al. (2013). The difficulty of calculating dissimilarity metrics for empirical distributions in RKHS is to compute integrals, these integrals are usually difficult to solve analytically. In this paper, we use Gaussian kernels for estimating these probability density functions.

We compare ABC based in our metric in RKHS with the ABC and K2ABC proposed by Park et al. (2015). We also show how to use these methods in combination with sequential Monte Carlo methods.

Experimental results obtained include the application of the different methods described above over a synthetic dataset and a real dataset. The synthetic data was created from an uniform mixture model. We demonstrate that our method is a robust estimator of the parameter vector in terms of the number of observations. The real data corresponds to the change of adult blowfly population during a period of time, as explained by Wood (2010).

2 Approximate Bayesian Computation based on Kernel Embeddings

In this section, we briefly expose the different ABC methods. We then define the metrics based on kernel embeddings employed with the ABC methods, including the MMD metric and the metric that we propose in this paper.

2.1 Short summary on ABC methods

Approximate Bayesian Computation (ABC) are likelihood-free Monte Carlo methods. In practical Bayesian models, exact inference may be intractable due to different reasons: the likelihood function is expensive or intractable; or the normalization constant in Bayesian modelling is also intractable. ABC methods can be employed to infer posterior distributions without having to evaluate likelihood functions (Toni et al., 2009; Wilkinson, 2008). These likelihood-free algorithms simulate data using different parameters drew from prior distribution and compare summary statistics of the simulated data ($s(\mathcal{D}')$) whit summary statistics of the observed data ($s(\mathcal{D})$). For this comparison it is necessary to define a tolerance threshold that determines the accuracy of the algorithm, and a distance measure $\rho(s(\mathcal{D}), s(\mathcal{D}'))$, for example Euclidean distance, etc. If $\rho(s(\mathcal{D}), s(\mathcal{D}')) \leq \epsilon$, we then accept the parameters θ drew from $p(\theta)$, otherwise, it is rejected.

According to Wilkinson (2008), there are different open questions around the ABC algorithm including how to choose the measure function ρ; what should be the value of ϵ?; how to select the summary statistics s? ρ and ϵ are experimental and implementation issues. With respect to s, it is difficult to define a methodology for choosing or constructing the summary statistics (Park et al., 2015).

2.2 Metrics between probability measures by using kernels

The embedding of distributions in a Reproducing Kernel Hilbert Space (RKHS) is a methodology that allows us to compute distances between distributions by using kernel functions (Berlinet and Christine, 2004). According to Sriperumbudur et al. (2010), a metric $\gamma_k(P, Q)$ over the probability measures P and Q can be defined through a characteristic kernel[1] $k(x, x')$ as,

$$\gamma_k(P, Q) = \left\| \int_M k(\cdot, x) \, dP(x) - \int_M k(\cdot, y) \, dQ(y) \right\|_{\mathcal{H}}^2.$$

If the distributions $P(x)$, and $Q(y)$ admit a density, then we have $dP(x) = p(x)dx$, and $dQ(y) = q(y)dy$, and an alternative expression for $\gamma_k(P, Q)$ can be written as

$$\begin{aligned} \gamma_k(P, Q) = &\int_M \int_M k(x, z) \, p(x) \, q(z) \, dxdz + \int_M \int_M k(z, y) \, p(z) \, q(y) \, dzdy \\ &- 2 \int_M \int_M k(x, y) p(x) q(y) dxdy. \end{aligned} \tag{1}$$

---

[1] A characteristic kernel is a reproducing kernel for which $\gamma_k(P, Q) = 0 \iff P = Q, P, Q \in \mathcal{P}$, where $\mathcal{P}$ denotes the set of all Borel probability measures on a topological space $(M, \mathcal{A})$.

2.3 Kernel Embeddings as summary statistics for ABC

Let us assume that we have two random samples $X = \{x_i\}_{i=1}^{N_x}$, and $Y = \{y_j\}_{j=1}^{N_y}$. In ABC, one of those samples would correspond to the real data ($\mathcal{D}$), whereas the other sample would correspond to simulated data ($\mathcal{D}'$) from the model we wish to estimate. As mentioned before, we need a way to define if accepting the simulated data. A key idea introduced by Park et al. (2015) was to assume that the random samples X, and Y are drawn from probability measures P, and Q, respectively, and instead of comparing the distance between samples X, and Y, they propose to compare the distance between the probability measures P and Q.

The authors in Park et al. (2015) assume empirical distributions for P, and Q, this is $p(x) = \frac{1}{N_x} \sum_{i=1}^{N_x} \delta(x - x_i)$, and $q(y) = \frac{1}{N_y} \sum_{j=1}^{N_y} \delta(y - y_j)$. With these expressions for $p(x)$, and $q(y)$, the distance $\gamma_k(P, Q)$ is given by

$$\gamma_k (P,Q) = \frac{1}{N_x^2} \sum_{i,j=1}^{N_x} k(x_i, x_j) + \frac{1}{N_y^2} \sum_{i,j=1}^{N_y} k(y_i, y_j) - \frac{2}{N_x N_y} \sum_{i,j=1}^{N_x, N_y} k(x_i, y_j). \tag{2}$$

We refer to this distance as $\gamma_k^{\text{MMD}}(P, Q)$, since it is rooted in the Maximum Mean Discrepancy (MMD) concept developed in Gretton et al. (2007b, 2012). After obtaining $\gamma_k^{\text{MMD}}(P, Q)$, Park et al. in Park et al. (2015) apply a second Kernel that operates on probability measures, as follows

$$k_\epsilon (P_\mathcal{D}, Q_{\mathcal{D}'}) = \exp\left(-\frac{\gamma_k^{\text{MMD}}(P_\mathcal{D}, Q_{\mathcal{D}'})}{\varepsilon}\right), \tag{3}$$

where ε is a positive parameter for the second kernel. $P_\mathcal{D}$ is the distribution associated to $\mathcal{D}$, and $Q_{\mathcal{D}'}$ is the distribution associated to $\mathcal{D}'$. In Park et al. (2015), the authors use an unbiased estimate for $\gamma_k^{\text{MMD}}(P, Q)$ in which the factors $1/N_x$, and $1/N_y$ are replaced for $1/(N_x(N_x - 1))$, and $1/(N_y(N_y - 1))$, respectively. Also, the innermost sum in each of the first two terms does not take into account the terms for which $i = j$.

Instead of assuming a discrete distribution for $p(x)$, and $q(y)$, in this paper we propose to use a smooth estimate for both densities based on the Parzen-window density estimator. We assume that the densities $p(x)$, and $q(x)$ can be estimated using

$$\widehat{p}(x) = \frac{1}{N_x} \sum_{m=1}^{N_x} \frac{1}{\left(2\pi h_p^2\right)^{D/2}} \exp\left(-\frac{\|x - x_m\|^2}{2h_p^2}\right),$$

$$\widehat{q}(y) = \frac{1}{N_y} \sum_{n=1}^{N_y} \frac{1}{\left(2\pi h_q^2\right)^{D/2}} \exp\left(-\frac{\|y - y_n\|^2}{2h_q^2}\right),$$

where h_p and h_q are the kernel bandwidths, and D is the dimensionality of the input space.

If we use a Gaussian kernel with parameter Σ for $k(x, x')$, and the estimators $\widehat{p}(x)$, and $\widehat{q}(y)$, a new distance between the distributions P and Q is easily obtained from expression (2) as follows

$$\gamma_k (P,Q) = \frac{1}{N_x^2} \sum_{i,j=1}^{N_x} \hat{k}(x_i, x_i; 2\Sigma_p) + \frac{1}{N_y^2} \sum_{i,j=1}^{N_y} \hat{k}(y_i, y_j; 2\Sigma_q) \tag{4}$$

$$- \frac{2}{N_x N_y} \sum_{i,j=1}^{N_x, N_y} \hat{k}(x_i, y_j; \Sigma_p + \Sigma_q),$$

where

$$\hat{k}(x, x'; S) = \frac{|\Sigma|^{1/2}}{|\Sigma + S|^{1/2}} \exp\left(-\frac{(x - x')^\top (\Sigma + S)^{-1} (x - x')}{2}\right).$$

In expression (4), $\Sigma_p = h_p^2 \mathbf{I}$ and $\Sigma_q = h_q^2 \mathbf{I}$. We refer to the metric in (4) as $\gamma_k^{\text{Parzen}}(P, Q)$.

As a distance measure $\gamma_k(P_\mathcal{D}, Q_{\mathcal{D}'})$ we can use $\gamma_k^{\text{MMD}}(P_\mathcal{D}, Q_{\mathcal{D}'})$ or $\gamma_k^{\text{Parzen}}(P_\mathcal{D}, Q_{\mathcal{D}'})$. The algorithm proposed by Park et al. (2015) that employs kernel embeddings of probability measures in ABC using MMD is shown in Algorithm 1. If we use the metric $\gamma_k^{\text{MMD}}(P_\mathcal{D}, Q_{\mathcal{D}'})$ on line 4 for the algorithm, we refer to the method as K2ABC. If we use the metric $\gamma_k^{\text{Parzen}}(P_\mathcal{D}, Q_{\mathcal{D}'})$ instead, we refer to the method as PABC.

Algorithm 1 ABC based on Kernel Embeddings

1: **Input:** Observed data $\mathcal{D}$, prior distribution, threshold ε.
2: **for** $i = 1, \ldots, N_s$ **do**
3: Draw θ from $p(\theta)$
4: Simulate $\mathcal{D}'$ using $p(\cdot | \theta)$
5: Compute $\widetilde{w}_i = \exp\left(-\frac{\gamma_k(P_{\mathcal{D}}, Q_{\mathcal{D}'})}{\varepsilon}\right)$
6: **end for**
7: Normalize $\widetilde{w}_i$

2.4 Extension to ABC SMC

The disadvantages of ABC, suck as: the selection of ϵ, the choice of summary statistics or accepted samples with low probability, it can be avoided using an ABC algorithm based on sequential Monte Carlo methods (ABC SMC) proposed by Sisson et al. (2007). The goal of ABC SMC is to obtain an approximation of true posterior using a series of sequential steps, expressed by $p(\theta | \rho(\mathcal{D}, \mathcal{D}') \leq \epsilon_t)$, for $i = 1, \cdots, T$, where ϵ_t is a threshold that decreases in each step ($\epsilon_1, >, \ldots, \epsilon_t, >, \ldots, > \epsilon_T$), thus it refine the approximation towards the target distribution.

ABC SMC has a first stage based on ABC rejection. We can replace this stage with K2ABC or PABC, leading to what we call in the paper as the K2ABC SMC, and PABC SMC methods. Details of the ABC SMC method can be found in Toni et al. (2009).

3 Experimental Setup

We first describe the synthetic dataset and the real dataset used for our experiments. Finally, we explain the procedure for applying the ABC methods over synthetic and real datasets.

3.1 Datasets

We follow the experiments described in Park et al. (2015). Synthetic data is created from an uniform mixture model. The real dataset correspond to a time series of an adult blowfly population (Wood, 2010).

Toy Problem. The synthetic dataset was generated by an uniform mixture model, $p(\mathcal{D} | \theta) = \sum_{k=1}^{K} \pi_k \mathcal{U}(k, k-1)$, where $\mathcal{D}$ are the observed data; π_k are the mixing coefficients; and K is the number of components. The model parameters θ correspond to the mixing coefficients $\pi = [0.25, 0.04, 0.33, 0.04, 0.34]^\top$. The prior distribution for θ is a symmetric Dirichlet distribution.

Noisy nonlinear ecological dynamic system For the real dataset example, we would like to infer the parameters of a non-linear ecological dynamic system (Meeds and Welling, 2014) represented through $N_{t+1} = PN_{t-\tau} \exp\left(-\frac{N_{t-\tau}}{N_0}\right) e_t + N_t \exp(\partial \varepsilon_t)$, where N_t is the adult population at time t, and P, N_0, ∂ and τ are parameters. Variables e_t and ε_t are employed to represent noise, and they are assumed to follow Gamma distributions $e_t \sim \mathcal{G}\left(1/\sigma_p^2, 1/\sigma_p^2\right)$, and $\varepsilon_t \sim \mathcal{G}\left(1/\sigma_d^2, 1/\sigma_d^2\right)$. The parameter vector we want to infer is given by $\theta = [P, N_0, \partial, \tau, \sigma_p, \sigma_d]^\top$. The observed data that we use in the experiments correspond to a time series of the adult population of sheep blowfly, with 180 observations. The data was provided by the authors of Wood (2010).

3.2 Validation

We perform a comparison between ABC, K2ABC, PABC, SMCABC, K2ABC SMC, and PABC SMC. The comparison is made in terms of the root-mean-squared error (RMSE) between the true parameters (when known) and the estimated parameters by the different methods. For the real dataset, we compute the cross-correlation coefficient (ρ) between the real time series and the time series generated by the different simulation methods.

3.3 Procedure

For the toy problem, the observed data $\mathcal{D}$ is sample once from the mixture of uniform densities with the mixture coefficients θ described above. The observed data contains 400 observations. To generate the simulated data $\mathcal{D}'$, we sample from the Dirichlet prior, and with that sample, we then generate 400 observations from the mixture model. We then apply the ABC, K2ABC, PABC, ABC SMC, K2ABC SMC and PABC SMC to the observed data, and the simulated data, and compute the RMSE over the true and estimated parameters. For all the ABC methods, the procedure for generating simulated data is repeated 1000 times. For ABC, we use $\epsilon = 0.002$, and compute the mean and standard deviation as summary statistics. For ABC SMC, K2ABC SMC and PABC SMC, we employ $\{\epsilon_t\}_{t=1}^{T} = (0.5, 0.01, 0.005, 0.001, 0.0005)$. For the ABC SMC type of algorithms, we need to specify what is known as a perturbation kernel. For this, we use a spherical multivariate Gaussian distribution, with parameter 0.0001.

For the noisy nonlinear ecological dynamic system, we apply all the ABC methods and calculate the cross-correlation coefficient between the observed data ($\mathcal{D}$), and the simulated data ($\mathcal{D}'$), using each method. We adopt similar priors to the ones used by Meeds and Welling (2014): $\log P \sim \mathcal{N}(3, 0.2)$, $\log \partial \sim \mathcal{N}(-1.5, 0.1)$, $\log N_0 \sim \mathcal{N}(6, 0.2)$, $\log \sigma_d \sim \mathcal{N}(-0.1, 0.01)$, $\log \sigma_p \sim \mathcal{N}(0.1, 0.01)$ and $\tau \sim \mathcal{P}(6)$.

For ABC, we use $\epsilon = 0.35$, and compute 8 summary statistics used in Meeds and Welling (2014): 4 statistics using the log of the mean of all 25% quantiles of $\left\{ \frac{N_t}{1000} \right\}_{t=1}^{180}$, 4 statistics employing the mean of 25% quantiles of the first-order differences of $\left\{ \frac{N_t}{1000} \right\}_{t=1}^{180}$, and we also compute the maximum and minimum value of $\left\{ \frac{N_t}{1000} \right\}_{t=1}^{180}$. For ABC SMC, K2ABC SMC and PABC SMC, we employ $\{\epsilon_t\}_{t=1}^{T} = (2, 1, 0.5, 0.35, 0.25, 0.2, 0.15)$; we also specify what is known as a perturbation kernel. For this, we use a spherical multivariate Gaussian distribution, with parameter 0.0001.

For K2ABC and PABC, in both experiments, we use a kernel bandwidth optimization approach based on minimizing the mean integrated square error (MISE) between the observed data and simulated data, to define the values of Σ, h_x and h_y. For more details about this kernel bandwidth optimization approach, see Shimazaki and Shinomoto (2010).

4 Results and discussion

The ABC, K2ABC, PABC, ABC SMC, K2ABC SMC, PABC SMC were evaluated and compared over the synthetic data set and the real data set described in section 3.

4.1 Results from synthetic data

All ABC methods were applied over vectors of 400 observations obtained from the uniform mixture model described in section 3. Fig. 1 shows a comparison among the evaluated methods. The figure presents the estimated $\mathbb{E}[\theta | \mathcal{D}']$ for each method.

For K2ABC, PABC, K2ABC SMC and PABC SMC the estimated parameters are close to the true posterior mean of the parameters (see section 3 for the true parameters). ABC and ABC SMC do not correctly estimate to θ_3 and θ_5. To observe the quality of the prediction using each method, we varied the number of observations and compute the RMSE between the true parameter vector and the estimated posterior mean of the parameter vector. We increase the observations from 40 to 400, in steps of 5 observations. For each step, we ran all methods and computed the RMSE. Fig. 2 presents the RMSE when the number of observations is increased.

Comparing Figs. 2(a) and 2(b) show that the RMSE obtained by PABC is the steadiest and present the lowest variability, indicating a robust estimator of the parameter vector in terms of the number of observations. The mean and one standard deviations for the RMSE for PABC was 0.0696 ± 0.0006. The PABC SMC also present a low variability for this metric, the RMSE value was 0.0716 ± 0.0022. RMSE obtained by ABC was 0.0879 ± 0.0050, for K2ABC was 0.0733 ± 0.0031, for ABCSMC was 0.0755 ± 0.0032 and for K2ABCSMC was 0.0747 ± 0.0041. For ABC, K2ABC, ABC SMC and K2ABC SMC, notice how the RMSE decreases when the number of observations increases.

4.2 Results from nonlinear ecological dynamic system

In Fig. 3, we present the posterior distribution for the parameters, obtained for the different methods when applied to the blowfly dataset of section 3. Figs. 3(a), 3(b), 3(c), 3(d), 3(e) and 3(f) illustrate the posterior of P, ∂, N_0, σ_d. σ_p and τ, respectively. In these figures, the black dashed line corresponds to the posterior of the each parameter using K2ABC, the blue dashed line is the posterior obtained by PABC, the red dashed line is the posterior employing ABC, the posterior

(a) ABC (b) K2ABC (c) PABC
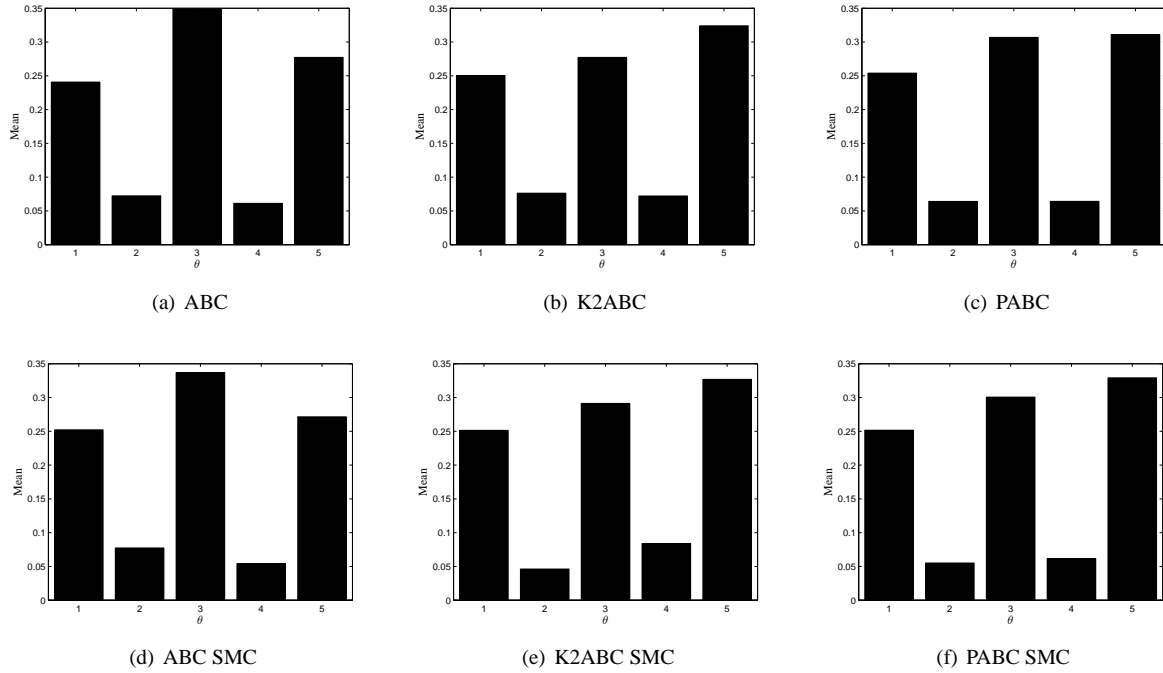
(d) ABC SMC (e) K2ABC SMC (f) PABC SMC

Figure 1: Estimated posterior mean of model parameters for the uniform mixture model, using all ABC methods. The integer number in the x axis represents the subindex i in the parameter θ_i.

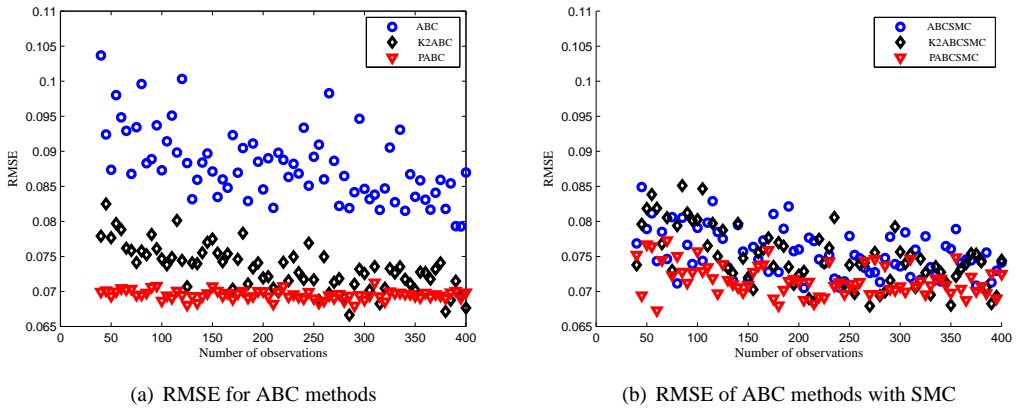(a) RMSE for ABC methods (b) RMSE of ABC methods with SMC

Figure 2: Root-mean-square error for different number of observations. We start to increase the observations from 40 to 400, in steps of 5 observations. In Fig. 2(a), the circles represent the RMSE obtained by ABC, the diamonds are the RMSE for K2ABC and the triangles are the RMSE values using PABC. In Fig. 2(b), the circles represent the RMSE obtained by ABC SMC, the diamonds are the RMSE for K2ABC SMC and the triangles are the RMSE values using PABC SMC.

obtained by K2ABC SMC is the cyan dashed line, the magenta dashed line is the posterior using PABC SMC and the green dashed line is the posterior obtained by ABC SMC. The posterior distributions obtained by using all ABC methods for ∂, σ_d. σ_p and N_0 are similar. For P and τ, the posterior obtained by K2ABC SMC, PABC SMC and ABC SMC are not similar with respect to the posterior using K2ABC, PABC and ABC; it is due to the parameter values, since this nonlinear model has a specific dynamical range from stable equilibrium to chaos.

(a) P
(b) ∂
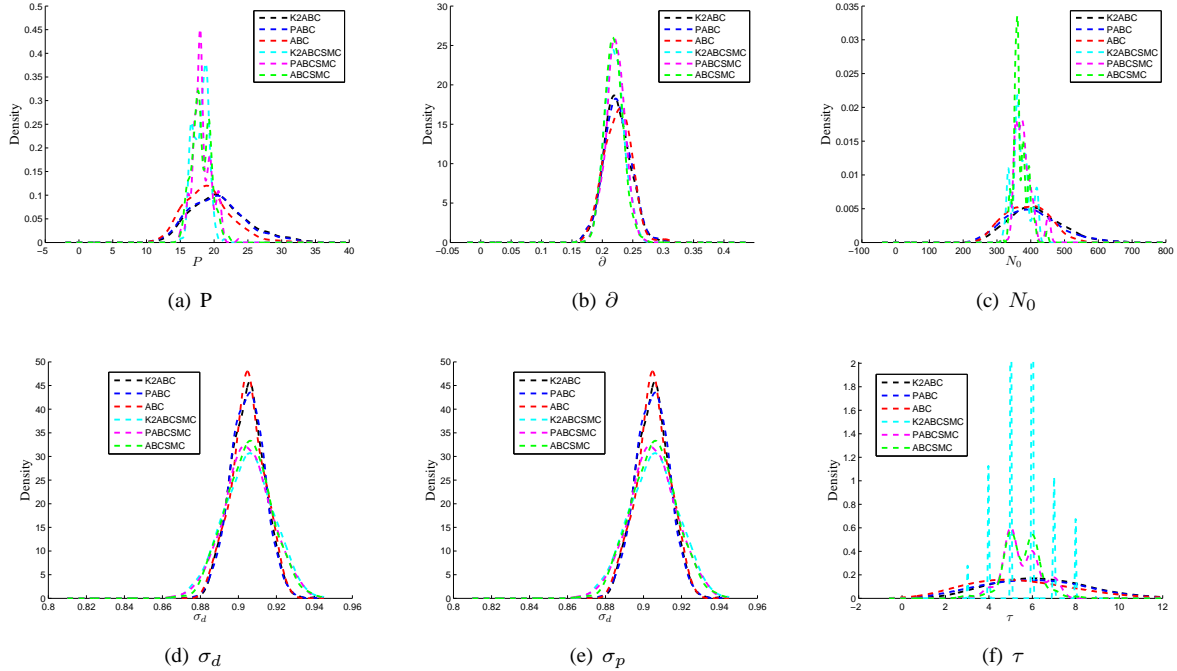(c) N_0

(d) σ_d
(e) σ_p
(f) τ

Figure 3: Posterior distribution for parameters of the nonlinear ecological dynamic system using ABC, K2ABC, PABC, ABC SMC, K2ABC SMC and PABC SMC.

To quantify the performance of all methods over the estimated parameters, we compare the time series obtained using the posterior means of the parameter and the observed data. For this comparison, we use a cross-correlation coefficient (ρ) between simulated data $\mathcal{D}'$ and observed data $\mathcal{D}$. The cross-correlation coefficient measures the similarity between two signals. We drew 100 subsets of parameters, we then apply all ABC methods, obtaining 100 estimated posterior mean for the parameters, with these estimated parameters we obtained 100 time series for these sets of signals, we compute ρ, we then sort ρ in descending order and choose the 50 first values, for all methods. Fig. 4 contains a box plot for the cross-correlation coefficient (ρ) using the different methods.

From Fig. 4, notice that PABC presents the highest median for ρ, with a value of 0.6501. We also observe that PABC SMC obtained a median of 0.6360 for ρ. K2AB CSMC presents a median of 0.6277, and for ABC SMC the obtained median for ρ was 0.6220. Finally, K2ABC and ABC obtained a median of 0.6138.

5   Conclusions

We introduced a new metric for comparing two data distributions in a RKHS, using smoother density estimators to compare empirical data distributions, and then highlight the accepted samples by employing ABC methods. We demonstrated that our method is a robust estimator of the parameter vector in terms of the number of observations. Finally, we showed for a real application that our method obtained the best similarity with respect to the observed data, in an application involving time-series. As future work, it would be possible to propose a new dissimilarity distance using RKHS for different applications like electrical networks analysis.
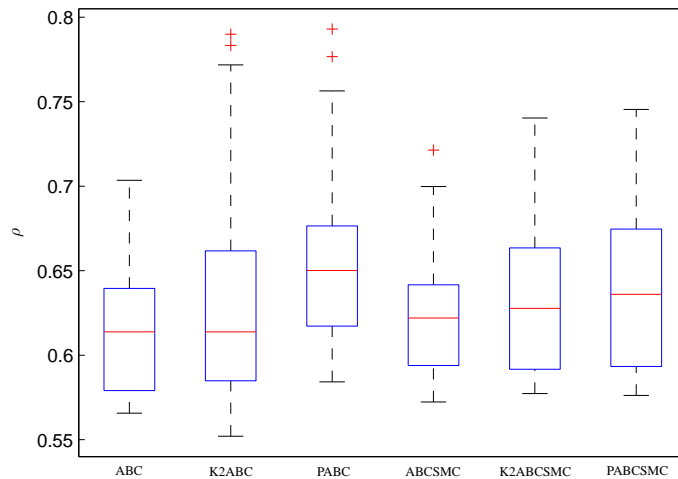
Figure 4: Boxplots of the 25th and 75th percentiles of the cross-correlation coefficient, using all ABC methods. We drew 100 subsets of parameters, we then apply all ABC methods, obtaining 100 estimated posterior mean for the parameters, with these estimated parameters we obtained 100 time series for these sets of signals, we compute ρ, we then sort ρ in descending order and choose the 50 first values, for all methods.

Acknowledgements

References

Alain Berlinet and Thomas-Agnan Christine. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science+Business Media, LLC, 2004.

M. G. B. Blum, M. A. Nunes, D. Prangle, and S. A. Sisson. A comparative review of dimension reduction methods in Approximate Bayesian Computation. *Stat. Sc.*, 28(2):189–208, 2013. doi: 10.1214/12-STS406.

Michael G.B. Blum. Choosing the summary statistics and the acceptance rate in Approximate Bayesian Computation. In Yves Lechevallier and Gilbert Saporta, editors, *Proceedings of COMPSTAT'2010*, pages 47–56. Physica-Verlag HD, 2010. ISBN 978-3-7908-2603-6.

Paul Fearnhead and Dennis Prangle. Constructing summary statistics for Approximate Bayesian Computation: semi-automatic Approximate Bayesian Computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2011.01010.x.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rash, Bernhard Schölkopf, and Alexander Smola. A kernel approach to comparing distributions. In *22nd AAAI Conference on Artificial Intelligence*, pages 1637–1641. AAAI Press, 2007a.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rash, Bernhard Schölkopf, and Alexander Smola. A kernel method for the two-sample problem. In *Advances in Neural Information Processing Systems 15*, pages 513–520. MIT Press, 2007b.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rash, Bernhard Schölkopf, and Alexander Smola. A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Paul Joyce and Paul Marjoram. Approximately Sufficient Statistics and Bayesian Computation. *Statistical Applications in Genetics and Molecular Biology*, 7(1), August 2008. ISSN 1544-6115. doi: 10.2202/1544-6115.1389.

W. Meeds and M. Welling. GPS-ABC: Gaussian Process Surrogate Approximate Bayesian Computation. arXiv:1401.2838, 2014.

Krikamol Muandet, Kenji Fukumizu, Bharath K. Sriperumbudur, Arthur Gretton, and Bernhard Schlkopf. Kernel mean estimation and Stein's effect. *CoRR*, abs/1306.0842, 2013.

M. A. Nunes and D. J. Balding. On optimal selection of summary statistics for Approximate Bayesian Computation. *Statistical Applications in Genetics and Molecular Biology*, 9(1), 2010. ISSN 1544-6115.

M. Park, W. Jitkrittum, and D. Sedjdinovic. K2-ABC: Approximate Bayesian Computation with infinite dimensional summary statistics via kernel embeddings. arXiv:1502.02558, 2015.

Hideaki Shimazaki and Shigeru Shinomoto. Kernel bandwidth optimization in spike rate estimation. *J. Comput. Neurosci.*, 29(1-2):171–182, August 2010. ISSN 0929-5313.

S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.

BK. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and GRG. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11:1517–1561, April 2010.

Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P. H. Stumpf. Approximate Bayesian Computation scheme for parameter inference and model selection in dynamical systems, 2009.

Richard D. Wilkinson. Approximate Bayesian Computation (ABC) gives exact results under the assumption of model error. arXiv/0811.3355, 2008.

S. N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(5):1102–1104, 2010. ISSN 1476-4687. doi: 10.1038/nature09319.