

氏 名 Jin Zhou

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 2074 号

学位授与の日付 平成 31年 3月 22日

学位授与の要件 複合科学研究科 統計科学専攻  
学位規則第6条第1項該当

学位論文題目 Kernel Methods in Approximate Bayesian Computation

論文審査委員 主 査 教授 栗木 哲  
教授 福水 健次  
准教授 間野 修平  
准教授 日野 英逸  
グループリーダー 赤穂 昭太郎  
産業技術総合研究所 人間情報研究部門

## 博士論文の要旨

氏名 Jin Zhou

論文題目 Kernel Methods in Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is a popular sampling method in applications involving intractable likelihood functions. Instead of evaluating the likelihood function, ABC approximates the posterior distribution by a set of accepted samples which are simulated from a generating model. Simulated samples are accepted if the distances between the samples and the observation are smaller than some thresholds. The distances are calculated in terms of summary statistics. Summary statistics are introduced to reduce the dimensionality of the samples and face the trade-off between dimensionality and information loss.

In this thesis, I focus on the applications of kernel methods to the ABC to provide an automatic algorithm which can produce low dimensional summary statistics while preserving information. As described above, summary statistics play a central role in the efficiency and accuracy of the ABC methods. It is important that the dimensional reduction algorithm can achieve the lowest dimension possible without information lost. Although a lot of dimensional reduction methods have been introduced to ABC already, an automatic algorithm with theoretically sound guarantees is still missing.

This thesis introduces a local kernel based sufficient dimensional reduction algorithm (LGKDR) to solve the above problem. Sufficient dimensional reduction (SDR) is a classic type of dimensional reduction algorithms that guarantees to find the sufficient lower dimensional subspace provided that the assumptions of the underlying spaces are met. In here sufficient means no information lost. As the assumptions of classic SDR are often too restrictive for real world problems, I instead draw the idea from the kernel dimensional reduction method.

To provide a principled way of designing the regression function, capturing the higher order non-linearity and realizing an automatic construction of summary statistics, this thesis introduces the kernel based sufficient dimension reduction method. This dimension reduction method is a localized version of gradient based kernel dimension reduction (GKDR). GKDR estimates the projection matrix onto the sufficient subspace by extracting the eigenvectors of the kernel derivatives matrices in the reproducing kernel Hilbert spaces (RKHS). In addition to GKDR, in which the estimation averages over all data points to reduce variance, a localized GKDR is proposed by averaging over a small neighborhood around the observation in ABC. Each point is weighted using a distance metric measuring the difference between the simulated data and the observation. The idea is similar to role of the distance kernel function.

Another proposal is to use different summary statistics for different parameters. Note

that sufficient subspace for different parameters can be different, depending on the particular problem. In these cases, applying separated dimension reduction procedures yield better estimations of the parameter.

Three experiments are investigated in the thesis to evaluate the proposed method against the popular dimensional reduction methods. First a M/G/1 queue model is investigated to simulate the stochastic process of serving customers in the queue, and a population genetic model is also studied to infer the parameters that effect the genetic variation within populations, at last a Ricker model that describe the population dynamics of species are investigated. Each experiment is conducted with two sampling algorithms: vanilla ABC and sequential ABC. The former provides an intuitive overall comparison and the latter is used to access the generated summary statistics in the extreme situations where the thresholds of the distance functions are pushed to as small as possible. This setting makes the latter experiments very time consuming but also provide a useful assessment on the generated summary statistics, which has not been reported before to our best knowledge.

The proposed method assumes no explicit functional forms of the regression functions nor the marginal distributions, and implicitly incorporates higher order moments up to infinity. As long as the initial summary statistics are sufficient, our method can guarantee to find a sufficient subspace with low dimensionality. While the involved computation is more expensive than the simple linear regression used in Semi-automatic ABC, the dimension reduction is conducted as the pre-processing step and the cost may not be dominant in comparison with a computationally demanding sampling procedure during ABC. Another advantage of LGKDR is the avoidance of manually designed features; only initial summary statistics are required. With the parameter selected by the cross validation, construction of low dimensional summary statistics can be performed as in a black box. For complex models in which the initial summary statistics are hard to identify, LGKDR can be applied directly to the raw data and identify the sufficient subspace. We also confirm that construction of different summary statistics for different parameter improve the accuracy significantly.

## 博士論文審査結果

Name in Full  
氏名 Jin Zhou

Title  
論文題目 Kernel Methods in Approximate Bayesian Computation

申請者は、近似ベイズ推論 (Approximate Bayesian Computation, 以下 ABC) における要約統計量の次元削減に関して研究を行った。提出論文は英文で書かれており全 6 章 77 頁からなる。

第 1 章は本論文の序章である。まず研究の背景として、尤度の値が計算できない場合の Bayes 推論において、サンプリングに基づく近似ベイズ推論 (ABC) の方法が研究されていることを紹介し、観測空間が高次元の場合にその近似精度が低くなる問題を述べている。その解決方法として本論文では、カーネル法に基づく次元削減の手法を用いて、推定するパラメータごとに異なる低次元射影を求めることにより、より効率的な ABC を構成する方法を論じることが説明されている。

第 2 章と第 3 章は準備である。第 2 章では ABC の概要といくつかの代表的な方法が紹介されている。また第 3 章では、次元削減の既存手法として、カーネル法を用いて回帰関数の勾配を求めることによって共変量の有効な低次元射影を求める Gradient Based Kernel Dimensional Reduction (GrKDR) に関して説明されている。

第 4 章では本論文における提案手法が述べられている。GrKDR を ABC の問題に有効に適用できるように、観測値の近辺だけで回帰関数の勾配を推定する Local GrKDR を提案し、この次元削減法を適用して、推定したい各パラメータに対して異なる次元削減の射影を求める方法を提案している。この方法により、ABC における要約統計量を効率的に低次元化することができる。また、カーネルのバンド幅などのハイパーパラメータの選択にはクロスバリデーションを用いることが論じられている。

第 5 章は実験的検証を行っている。提案手法と、既存手法である rejection 法、semi-automatic ABC, Sequential Monte Carlo ABC の比較を、集団遺伝、待ち行列、生態学分野の力学系の 3 種のモデルに対して行っている。実験結果によって、同程度の計算コストを用いた場合、提案した方法が既存手法と同等以上の推定精度を与えることが明らかとなっている。また、同じ目的の次元削減法として使うことができる層別逆回帰 (SIR) との数値比較もなされている。第 6 章は論文のまとめである。

尤度の値が陽に計算できないような複雑なモデルに対する Bayes 推論をシミュレーションによって近似計算する ABC の方法は、近年多くの分野で注目を集めているが、高次元の観測値に対して精度よい近似が困難な点が大きな課題となっている。本論文は、その重要な問題に対して、ノンパラメトリックな次元削減の方法による有望な方法を与えている。また第 4 章および第 5 章の内容をまとめた論文が査読付き国際学術雑誌 Open Journal of Statistics に採択されている。以上の理由により、審査委員会は、本論文が学位の授与に値すると判断した。