Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

On selection of statistics for approximate Bayesian computing (or the method of simulated moments)

Q1 Michael Creel [a,*], Dennis Kristensen [b,1]

[a] Universitat Autònoma de Barcelona, Barcelona Graduate School of Economics, MOVE, Spain

[b] University College London, Institute of Fiscal Studies, CREATES, United Kingdom

ARTICLE INFO

Article history:
Received 6 November 2014
Received in revised form 10 May 2015
Accepted 11 May 2015
Available online xxxx

Keywords:
Approximate Bayesian computation
Likelihood-free methods
Selection of statistics
Method of simulated moments

ABSTRACT

A cross validation method for selection of statistics for Approximate Bayesian Computing, and for related estimation methods such as the Method of Simulated Moments, is presented. The method uses simulated annealing to minimize the cross validation criterion over a combinatorial search space that may contain an extremely large number of elements. A first simple example, for which optimal statistics are known from theory, shows that the method is able to select these optimal statistics out of a large set of candidate statistics. A second example of selection of statistics for a stochastic volatility model illustrates the method in a more complex case. Code to replicate the results, or to use the method for other applications, is provided at http://www.runmycode.org/companion/view/1116.

1. Introduction

Bayesian analysis centers attention on the posterior density, $f(\theta|y)$, where y is the sample, that is, a realization of a random vector Y, and θ is the parameter vector. As is well known, the posterior is proportional to the product of the likelihood of the sample, $f(y|\theta)$, and the prior, $\pi(\theta)$, so a Bayesian analysis requires the likelihood function. Classical statisticians often use the maximum likelihood (ML) estimator, because of its desirable property of asymptotic efficiency. For both approaches, the likelihood function is needed. However, there are many areas of research involving complex models where computation of the likelihood function is not possible, or is extremely expensive, effectively ruling out ordinary Bayesian methods and ML estimation.

Much research in the last several decades, both classical and Bayesian, has been focused on providing alternatives. In the classical framework, and among other approaches, the Generalized Method of Moments (GMM; Hansen, 1982) is based on a vector of moment conditions $m(\theta, y)$ that are assumed to have expectation equal to zero under the true model. A common way to define moments first defines a vector of statistics Z, a mapping operating on Y, $Z = Z(Y)$, with sample realization $z = Z(y)$. Let $E_\theta[Z]$ be the expectation of Z under the model and $m(\theta, y) = z - E_\theta[Z]$ be the moment restrictions implied by the model; by construction, these moments have expectation zero when $\theta = \theta_0$, where θ_0 is the true parameter value. Approximate Bayesian Computing (ABC) is a set of methods that attempts to use Bayesian ideas when the likelihood function of the sample is not available, but the model is simulable. An important part of this literature (e.g., Beaumont et al., 2009; Fearnhead and Prangle, 2012; Blum et al., 2013) focuses on the posterior of the parameter conditional on a statistic,

* Correspondence to: Department of Economics and Economic History, Universitat Autònoma de Barcelona, Bellaterra (Barcelona) 08193, Spain. Tel.: +34 935811696; fax: +34 935812012.

E-mail addresses: michael.creel@uab.es (M. Creel), d.kristensen@ucl.ac.uk (D. Kristensen).

[1] Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom.

http://dx.doi.org/10.1016/j.csda.2015.05.005
0167-9473/

Please cite this article in press as: Creel, M., Kristensen, D., On selection of statistics for approximate Bayesian computing (or the method of simulated moments). Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.05.005

$f(\theta|z)$. In both areas, use of a statistic offers the simple advantage of feasibility, as well as computational advantages, due to the reduction of the dimensionality of the problem from dim Y to dim Z. The cost is a possible loss of statistical efficiency, compared to what would obtain were the likelihood function available.

It is well known that the asymptotic distributions of GMM estimators depend crucially on the specific Z that is chosen. We know that adding statistics to a baseline set will not cause the asymptotic variance of the GMM estimator to increase, which might cause one to think that selection of statistics is not important, and that when in doubt, additional statistics should be included. However, use of weakly informative moment conditions can lead to very poor small sample performance of the GMM estimator (e.g., Tauchen, 1986; Stock et al., 2002). In some cases, the model can give strong guidance regarding what moments to use for estimation, but in others, the choice is not clear, and a researcher may need to select which moments to use from a set that may include weakly informative and uninformative moments. These issues are also relevant for ABC with the posterior distribution $f(\theta|z)$ in general being sensitive to the choice of the summary statistics. From a frequentist point of view, ABC is first order asymptotic equivalent to the efficient GMM estimator based on $m(\theta, y) = z - E_\theta[Z]$ (see Creel and Kristensen, 2013 and Forneron and Ng, 2015), so the above cited results for moment selection also apply to ABC. Gao and Hong (2015) show how ABC-type ideas can be employed to compute Bayesian versions of GMM estimators in a general setting.

Thus, an important part of implementing GMM and ABC is the choice of the statistics Z. Systematic methods for selection of statistics have received a good amount of attention in both the ABC and GMM literatures. Within the ABC literature, Blum et al. (2013) provide a review of this work, with some new results based on use of regularization methods. The methods that have been used fall into three classes: best subset selection; projection methods that define new statistics by combining statistics in some way to reduce dimension; and regularization techniques such as ridge regression.

This paper adds to this literature by proposing a simple cross validation method for selecting statistics from a possibly large collection of candidate statistics, denoted W, assuming that the model is simulable. The method lies in the best subset selection classification of Blum et al. (2013): By searching over the candidate statistics using the simulated annealing algorithm, it provides a systematic method of dealing with the problem of a potentially large set of candidate statistics, while limiting computational cost. One of the examples presented below has 56 candidate statistics, so the number of possible combinations to search over, 2^{56}, is more than 70,000 million millions. Specifically, selection of statistics is investigated based on the idea of minimization of an integrated version of Bayes expected loss, $\int L(E[\Theta|Z = z] - \theta) f(z|\theta)\pi(\theta)dzd\theta$ for some loss function L, with respect to all subsets of statistics, Z, of W. Here, $E[\Theta|Z = z]$ is the posterior mean, $f(z|\theta)$ is the density of the statistic conditional on the parameter, and $\pi(\theta)$ is the prior. The integrated Bayes expected loss cannot be computed analytically, but can easily be approximated through simulations as demonstrated in the following. The use of Bayes expected loss in a decision theoretic framework is quite standard, but there, one usually conditions on the particular outcome z that has been observed. Here, on the other hand, we choose to integrate over all possible outcomes with weights assigned by their likelihoods; this is meant to reflect that we are interested in identifying a universal ("global") set of sufficient statistics that work well for the given model irrespectively of the particular outcome observed. Moreover, this integrated version of Bayes expected loss proves to be simpler to estimate compared to the conditional version.

The above selection rule aims at minimizing the loss of the posterior mean. If the interest lies in the whole posterior distribution, and not just its mean, we show how our method is easily adjusted so as to minimize the integrated Kullback–Leibler distance of the posterior density for a given set of summary statistics w.r.t. z. This version of our selection procedure chooses the set of summary statistics that maximizes the information content over the whole distribution and so is similar in spirit to the one of Barnes et al. (2012). However, the procedure of Barnes et al. (2012) only performs a partial search over the set of candidate statistics, while our algorithm does a full search.

From a frequentist point of view, $E[\Theta|Z = z]$ can be interpreted as a point estimator of the true data-generating value; using of the posterior mean as an estimator is defensible by the first order asymptotic equivalence of the posterior mean and the efficient GMM estimator (Chernozhukov and Hong, 2003; Creel and Kristensen, 2013). Our procedure can therefore be thought of as minimizing Bayes Risk, which again is a well-known object in decision theory, of this estimator. Thus, our procedure also applies to simulation-based GMM estimators, including the method of simulated moments (McFadden, 1989), indirect inference (Smith, 1993; Gouriéroux et al., 1993), and the efficient method of moments (Gallant and Tauchen, 1996). The specific procedure that we propose for computing the integrated Bayes expected loss computation requires that the model of interest is simulable. This, in general, requires a fully specified model. Thus, the implementation of our procedure would have to be modified in order to use it for selection of moments in GMM estimation of models that are only partially specified.

In comparison to existing methods developed in the ABC literature, our proposal has the advantage that it provides a universal search over all possible candidate sets of statistics. In contrast, most competing methods either rely on fairly complex step-up procedures, where one statistic is added at a time, or on the construction of approximately optimal statistics. These methods tend to be more complex to implement and require much computational effort to achieve a solution that is stable across repeated runs. In contrast, our method is relatively simple to implement and tends to converge quite quickly. We investigate the performance of the proposed method through two examples, and in both cases our procedure performs well. The first example takes the form of a linear regression model, where the optimal statistics are known, and thus it provides a good testing ground for our procedure. We show that our method can identify these statistics quite rapidly and does so robustly across many simulated samples. This simple test problem should be of independent interest since it might be used by proponents of other methods to evaluate performance and to allow comparison of computational demands. The second

Please cite this article in press as: Creel, M., Kristensen, D., On selection of statistics for approximate Bayesian computing (or the method of simulated moments). Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.05.005