ABC random forests for Bayesian parameter inference

Louis Raynal[*†1], Jean-Michel Marin[‡1,2], Pierre Pudlo[2,3], Mathieu Ribatet[1], Christian P. Robert[4,5], and Arnaud Estoup[2,6]

[1]Institut Montpelliérain Alexander Grothendieck (IMAG) – Université de Montpellier, CNRS – France
[2]Institut de Biologie Computationnelle (IBC) – Université de Montpellier, CNRS – France
[3]Institut de Mathématiques de Marseille (I2M) – Aix-Marseille Université – France
[4]Universite Paris Dauphine – PSL Research University, Paris – France
[5]Departement of Statistics, University of Warwick – United Kingdom
[6]CBGP, INRA, CIRAD, IRD, Montpellier SupAgro – Université de Montpellier – France

Abstract

Approximate Bayesian computation (ABC) has grown into a standard methodology that manages Bayesian inference for models associated with intractable likelihood functions. Most ABC implementations require the preliminary selection of a vector of informative statistics summarizing raw data. Furthermore, in almost all existing implementations, the tolerance level that separates acceptance from rejection of simulated parameter values needs to be calibrated. We propose to conduct likelihood free Bayesian inferences about parameters with no prior selection of the relevant components of the summary statistics and bypassing the derivation of the associated tolerance level. The approach relies on the random forest methodology of Breiman (2001) applied in a (non parametric) regression setting. We advocate the derivation of a new random forest for each component of the parameter vector of interest. When compared with earlier ABC solutions, this method offers significant gains in terms of robustness to the choice of the summary statistics, does not depend on any type of tolerance level, and is a good trade-off in term of quality of point estimator precision and credible interval estimations for a given computing time. We illustrate the performance of our methodological proposal and compare it with earlier ABC methods on a Normal toy example and a population genetics example dealing with human population evolution. All methods designed here have been incorporated in the R package abcrf (version 1.7) available on CRAN.

[*]Speaker
[†]Corresponding author: louis.raynal@umontpellier.fr
[‡]Corresponding author: jean-michel.marin@umontpellier.fr