ABC random forests for Bayesian parameter inference: review

Dennis Prangle

July 2017

This paper proposes a new method of likelihood-free parameter inference. The traditional ABC approach is to sample a "reference table" of prior parameter draws and corresponding simulated data (or summary statistics of it). The closest simulated data sets to the observations are found, and the corresponding parameters returned, possibly after some "regression correction" post-processing. This can be viewed as a nearest neighbours method of regression. However this is subject to curse of dimensionality problems.

The paper instead proposes using random forest regression. This gives a straightforward approach of producing a point estimate a single quantity of interest. The authors also propose and compare three methods of producing an estimate of the posterior variance, concluding with a definitive recommendation of one method. The random forest approach has some nice features in comparison to ABC - avoiding the ϵ tuning parameter, and reducing the need for summary statistic selection. Extensive computer simulations show also that the new approach is generally at least as accurate as ABC plus regression correction (with neural network corrections sometimes being competitive). A genetics example shows the method can be used on a substantial application. The method is available as an R package which helps reproducibility.

I think this paper presents an appealing new method and is practically acceptable for

publication as it is. I only suggest some minor revisions. The "major comments" below are suggestions for extra topics to discuss in the paper. The "minor comments" are points I thought were presented unclearly and should be changed. I don't anticipate any of these requiring substantial changes to the paper. Finally I also list a few possible typos.

1 Major comments

1. I suggest a discussion of Papamakarios and Murray (2016). This recent method also performs ABC-like inference using a machine learning regression method – neural networks – in place of nearest neighbours. It would be interesting to discuss, at least briefly, the relative pros and cons of random forests and neural networks in this setting. Two potential advantage of neural networks are that they can produce an approximation of a multivariate posterior, and that they can, in theory, work with raw data without requiring features to be proposed. (Although this seems hard to implement in the genetics application.)

2. I'd like to see some discussion of the time required to fit the random forests compared to ABC regression-adjustment methods.

3. Section 3.3 discusses two tuning choices: number of simulations and number of trees. Another choice is the minimum leaf size - do you have any comments on how this might affect the method? One might think this in some sense controls the level of approximation.

4. The paper focuses on approximating univariate posteriors. Is there any prospect of achieving multivariate posteriors – e.g. using multiple objective random forests (Kocev et al., 2007)?

5. The paper nicely illustrates that the random forest posterior variance estimates tend to be biased upwards. I wonder if this is due to the random selection of features

for decision trees. That is, some trees will select less informative features and so produce estimates biased towards the prior. There might be some scope to avoid this using alternatives to random feature selection e.g. Bayesian additive regression trees, or boosted regression trees. This could be worth discussing (I'll leave this to the discretion of the authors.)

2 Minor comments

1. Section 2.3.4: "This representation remains valid since the weights are equal to zero when $\eta(y) \neq \eta(y^{(t)})$ in the limiting case of exact ABC, namely when only accepting parameter values for which the summary statistics of simulated data are identical to the summary statistics of the observed data."

   I didn't understand this sentence - can you elaborate? In particular:

   - What does "valid" refer to? (consistency as $N \to \infty$?)
   - What's the relevance of exact ABC to the random forest method?

2. In Section 3, could you briefly mention the definition of normalised mean absolute error. I'm not completely sure what normalisation would be used.

3. In Section 3.2 you describe the demographic model. I suggest briefly mentioning the genetic model as well.

4. Section 4. "The performances for covariance approximation are quite encouraging as well...". I suggest emphasising that all the details of this are in the supplementary material.

5. Pg 18 - "quantile estimation is not uniformly optimal". Does this mean that regression correction is sometimes better? Which table/figure is this referring to?

6. Supplementary material pg 1: I didn't understand how "method 2" would work. I suggest adding a brief description of it.

7. Supplementary material pg 4: "maximum node size equal to 10". I think this should say "minimum node size". Also on page 7 of the main paper minimum node size was 5 – why change to 10 here?

8. Supplementary material pg 4: Is multivariate ABC regression correction used here, or is a single scalar parameter targeted as for the random forest approach?

3 Typos etc

1. There are some backward quotation marks in the text e.g. "...".

2. Pg 2 - "a mean to deliver" $\rightarrow$ "a means to deliver"?

3. Pg 3 - "the simulations $y^{(t)}$'s" $\rightarrow$ "the simulated $y^{(t)}$'s"?

4. Pg 3 - There are some references to "Fearnhead and Prangle 2015" which I think should say "Li and Fearnhead 2015".

5. Pg 4 - "it does not request" $\rightarrow$ "it does not require"?

6. Pg 11 - "deducted" $\rightarrow$ "deduced".

7. Pg 18 - "Using simulated reference table" $\rightarrow$ "Using a simulated reference table".

8. Pg 1 (supplementary): "expansive" $\rightarrow$ "expensive"?

References

Kocev, D., Vens, C., Struyf, J., and Džeroski, S. (2007). Ensembles of multi-objective decision trees. *Machine Learning: ECML 2007*, pages 624–631.

Papamakarios, G. and Murray, I. (2016). Fast ε-free inference of simulation models with Bayesian conditional density estimation. In *Advances in Neural Information Processing Systems*, pages 1028–1036.