“Constructing summary for Approximate Bayesian Computation: Semi-automatic approximate Bayesian computation”, by Paul Fearnhead and Dennis Prangle: A discussion Christian P. Robert, Université Paris-Dauphine, IUF, and CREST

A discussion paper on the fast-growing technique of ABC techniques is quite timely, especially when it addresses the important issue of summary statistics used by such methods. I thus congratulate the authors on their endeavour.

While ABC has been gradually been analysed from a (mainstream) statistical perspective, this is one of the very first papers performing a decision-theoretic analysis of the factors influencing the performances of the method (along with, e.g., Dean et al., 2011). Indeed, a very interesting input of the authors is that ABC is considered there from a purely inferential viewpoint and calibrated for estimation purposes. The most important result therein is in my opinion the consistency result in Theorem 2, which shows that noisy ABC is a coherent estimation method when the number of observations grows to infinity. I however dispute the generality of the result, as explained below.

Fearnhead and Prangle do not follow the usual perspective of looking at ABC as a converging (both in N and h) approximation to the true posterior density. Instead, they consider a randomised (or noisy) version of the summary statistics

$$s_{\mathrm{obs}} = S(y_{obs}) + hx\,, \quad x \sim K(x)$$

and they derive a calibrated version of ABC, i.e. an algorithm that gives “proper” predictions, but only for the (pseudo-)posterior based upon this randomised version of the summary statistics. This randomisation however conflicts with the Bayesian paradigm in that it seems to require adding pure noise to the observation to conduct inference. Furthermore, Theorem 2 is valid for any value of h. I thus wonder at the overall statistical meaning of *calibration*, since even the prior distribution (corresponding to $h = +\infty$) is calibrated. while the most informative (or least randomised) case (ABC) is not necessarily calibrated. Nonetheless, the interesting aspect of this switch of perspective is that the kernel K used in the acceptance probability, with bandwidth h,

$$K((s - s_{\mathrm{obs}})/h)\,,$$

need not behave like an estimate of the true sampling density since it appears in the (randomised) pseudo-model.

In the authors’ setting, the Monte Carlo error that is inherent to ABC is taken into account through the average acceptance probability, which collapses to zero when h goes to zero, meaning that $h = 0$ is a suboptimal choice. This is a strong (and valid) point of the paper because this means that the “optimal” value of h is not zero, a point repeated later in this report. The later decomposition of the error into

$$\mathrm{trace}(A\Sigma) + h^2 \int x^{\mathrm{T}} A x K(x) \mathrm{d}x + \frac{C_0}{Nh^d}$$

is very similar to error decompositions found in (classical) non-parametric statistics. In this respect, I do fail to understand the argument of the authors that

Lemma 1 implies that a summary statistics with larger dimension also has larger Monte Carlo error: Given that $\pi(s_{\text{obs}})$ also depends on h, the appearance of h^d in eqn. (6) is not enough of an argument. There actually is a larger issue I also have against several recent papers on the topic, where h or ϵ is treated as a given or absolute number when it should be calibrated in terms of a number of statistical and computational factors, the number of summary statistics being one of them.

When the authors consider the errors made in using ABC, balancing the Monte Carlo error due to simulation with the ABC error due to approximation (and non-zero tolerance), they fail to account for "the third man" in the picture, namely the error made in replacing the (exact) posterior inference based on $\mathbf{y}_{\text{obs}}$ with the (exact) posterior inference based on $\mathbf{s}_{\text{obs}}$, i.e. for the loss of information due to the use of the summary statistics at the centre of the Read Paper. (As shown in Robert et al., 2011, this loss may be quite extreme as to the resulting inference to become inconsistent.) While the remarkable (and novel) result in the proof of Theorem 3 that

$$\mathbb{E}\{\theta|\mathbb{E}[\theta|\mathbf{y}_{\text{obs}}]\} = \mathbb{E}[\theta|\mathbf{y}_{\text{obs}}]$$

shows that $\mathbf{s}_{\text{obs}} = \mathbb{E}[\theta|\mathbf{y}_{\text{obs}}]$ does not loose any (first-order) information when compared with $\mathbf{y}_{\text{obs}}$, hence is "almost" sufficient in that weak sense, Theorem 3 only considers a specific estimation aspect, rather than full Bayesian inference, and is furthermore parameterisation dependent. In addition, the second part of the theorem should be formulated in terms of the above identity, as ABC plays no role when $h = 0$.

If we concentrate more specifically on the mathematical aspects of the paper, a point of the utmost importance is that Theorem 2 can only hold at best when θ is identifiable for the distribution $\mathbf{s}_{\text{obs}}$. Otherwise, some other values of θ satisfy $p(\theta|\mathbf{s}_{\text{obs}}) = p(\theta_0|\mathbf{s}_{\text{obs}})$. Considering the specific case of an ancilary statistic $\mathbf{s}_{\text{obs}}$ clearly shows the result cannot hold in full generality. Therefore, vital assumptions are clearly missing to achieve a rigorous formulation of this theorem. The call to Bernardo and Smith, 1994 is thus not really relevant in this setting as the convergence results therein require conditions on the likelihood that are not necessarily verified by the distribution of $\mathbf{s}_{\text{obs}}$. We are thus left with the open question of the asymptotic validation of the noisy ABC estimator—ABC being envisioned as an inference method *per se*—when the summary variables are not sufficient. Obtaining necessary and sufficient conditions on those statistics as done in Marin et al. (2011) for model choice is therefore paramount, the current paper obviously containing essential features to achieve this goal.

Overall, and far from mathematical reasons, I remain skeptical about the "optimality" resulting from this choice of summary statistics as (a) practice—at least in population genetics—shows that proper approximation to genuine posterior distributions stems from using a number of summary statistics that is (much) larger than the dimension of the parameter; (b) the validity of the approximation to the optimal summary statistics used as the actual summary statistics ultimately depends on the quality of the pilot run and hence on the choice of the summary statistics therein; this approximation is furthermore susceptible to deteriorate as the size of the pilot summary statistics grows; (c) important inferencial issues like model choice are not covered by this approach and recents results of ours (Marin et al., 2011) show that estimating statistics are likely to bring inconsistent solutions in this context; those results imply furthermore than a naïve duplication of Theorem 3, namely based on the Bayes factor as a candidate summary statistic, would be most likely to fail.

In conclusion, I find the paper both exciting and bringing both new questions and new perspectives to the forefront of ABC research.

References

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. John Wiley, New York.

Dean, T. A., Singh, S. S., Jasra, A. and Peters, G. W. (2011). Parameter estimation for hidden Markov models with intractable likelihoods. Tech. Rep. arXiv:1103.5399, Cambridge University Engineering Department.

Marin, J., Pillai, N., Robert, C. and Rousseau, J. (2011). Relevant statistics for Bayesian model choice. Tech. Rep. arXiv:1111.4700.

Robert, C., Cornuet, J.-M., Marin, J.-M. and Pillai, N. (2011). Lack of confidence in ABC model choice. *Proceedings of the National Academy of Sciences*, **108(37)** 15112–15117.