Review report for “Approximate confidence distribution

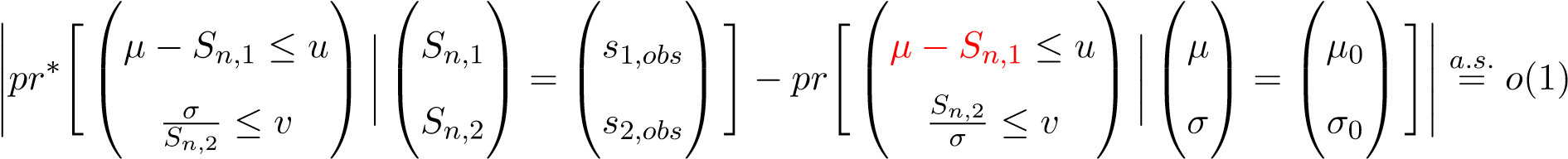
# computing”

This paper investigates the Approximate Confidence Distribution Computing (ACDC) as a likelihood-free inference procedure, and establishes its frequentist guarantees through solid theoretical arguments, especially a matching condition (Condition 1). Unlike the Approximate Bayesian Computation (ABC), ACDC is advantageous for potential utility since the sufficiency of the summary statistic is not essential for the valid inference. Besides, this paper proposes the Minibatch scheme for *rn* as a practical contribution. However, it would be good if the authors could address the following issues.

1. Theoretical perspective
   1. Although there are some practical contributions in this paper, it seems that the theoretical contributions are valued more in this paper. In Section 1.2, the authors claim that their large sample theoretical results are similar to those in [1] but with a focus on ACDC instead of the approximation to the posterior distribution. However, it still seems that the large sample theoretical results in this paper rely on the results from [1] and [2] to some degree. It would be great if authors could illustrate more on the technically challenging/interesting part for obtaining the large sample theoretical results in this paper, given the conclusions/results from [1] and [2].
   2. The main theoretical contribution of this work is the identification of a matching condition (Condition 1) for the valid frequentist inference from ACDC methods. It is good to see the authors have established theorems and a corollary to check Condition 1. Specifically, under the finite sample case, Corollary 1 is proposed to help check Theorem 1 and thus Condition 1 for the location and scale families. However, in Corollary 1, the authors assume the data-dependent *rn* not depend on data (i.e., *rn*(*µ*) ∝ 1 in (a), *rn*(*σ*) ∝ 1*/σ* in (b), and *rn*(*µ,σ*) ∝ 1*/σ* in (c)), which seems to make ACDC degenerate to ABC. It will be great if authors could discuss more on how to check Theorem 1 and Condition 1 for a true ACDC under the finite sample case (i.e., with a data-dependent *rn*, for example, obtained from the developed Minibatch scheme).
   3. Also in Corollary 1(a), the authors conclude

|*pr*∗(*µ* − *Sn* ≤ *u* | *Sn* = *sobs*) − *pr*(*Sn* − *µ* ≤ *u* | *µ* = *µ*0)| *a.s.*= *o*(1)*,*

but in Corollary 1(c), they conclude

*.*

Since Corollary 1(a) should be considered as the first marginal of the scenario considered in Corollary 1(c), it is surprising that the expressions in red from above two equations are in the opposite signs of each other. Please check that whether this inconsistency is due to a typo. If not, could the authors have some justifications/intuitions on this discrepancy?

1. Methodological perspective
   1. Could authors give some suggestions (or perform some simulations) on the choice of the kernel function *K*(·) for possibly better efficiency/performance of Algorithm 1 and the Minibatch scheme in practice, even though the choices may not affect the validity of the frequentist inference from ACDC.
   2. Nowadays, the number of parameters is usually large. However, only 2 and 3 parameters are considered in the two numerical examples of this paper respectively. It would be interesting if the authors could discuss the utility of ACDC/Algorithm 1 and the challenging part of the associated theoretical works, or perform the numerical experiments under the higher dimensional case, as long as *p* ≤ *d < n*. For example, will the kernel function based Algorithm 1 still work?
   3. In the last paragraph of Section 3.2, the authors claim that ACDC is preferred over the nonparametric bootstrap because performance of the latter one heavily depend on the quality of the estimator, and the computational cost of the estimator *θ*ˆ*S* might be higher than that of the summary statistic. Indeed, the Cauchy example (Section 4.1) and the Ricker model (Section 4.2) in this paper are good examples for illustrating the utility of ACDC. However, both of these two examples are rather restricted. It would be great if the authors could investigate/discuss some broader settings where ACDC would be a good fit, compared with some other likelihood-free inference procedures. Meanwhile, the authors say that the nonparametric bootstrap will be investigated in Section 4.2, but only parametric bootstrap is briefly mentioned in Section 4.2. Could the authors give some justifications on this discrepancy?
2. Numerical examples
   * 1. In Section 2.2, the authors claim that the Cauchy example is covered by Corollary 1. However, a data-dependent Minibatch scheme is accepted to obtain *rn* in Section 4.1, which is inconsistent with the requirements for *rn* in Corollary 1 (see the comment 1(b)). Could the authors give some discussions on this discrepancy?
     2. In the numerical example in Section 4.1, the authors claim that the summary statistics in settings one, three, and five are “informative”, whereas those in settings two and four are “less informative”. Could authors give some explanations for the criterion/definition on how to determine a summary statistics is “infomrative” or not, especially in this example the authors claim “only the entire data vector itself is sufficient” (in Section 1.2).
     3. In the numerical example in Section 4.1, the confidence regions from ACDC are tighter than those from IS-ABC by choosing “less informative” summary statistics. However, in Section 4.2, the confidence regions from ACDC are significantly tighter than those from IS-ABC, and the authors say “The summary statistics in this example were carefully selected to be informative based on domain knowledge.”. It is confusing whether “informative” or “less informative” summary statistics are preferable for the application of ACDC on statistical inference. Could authors explain more on this discrepancy, and give some guidance on the choice of the summary statistics for better efficiency, even though its choice will not affect the validity of the inference from ACDC?
3. There are several typos in the paper. Please carefully check and make corrections.

## References

1. Li, W. and Fearnhead, P. (2018a). Convergence of regression-adjusted approximate Bayesian computation. *Biometrika*, 105(2):301–318.
2. Li, W. and Fearnhead, P. (2018b). On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika*, 105(2):285–299.

# Reviewer 2

this paper provides an example of how a modern understanding of confidence

distribution theory can be used to connect Bayesian and frequentist

inferential paradigms, it presents a case to expand the current scope of

approximate Bayesian inference to include non-Bayesian inference by targeting

a confidence distribution rather than a posterior. It has developed a

data-driven approach to drive ACDC in both Bayesian or frequentist contexts.

Some theoretical results are established, which include both asymptotic and

finite sample properties. Some practical issues are also discussed.

This is a nice paper with some interesting ideas. However, I have some minor

questions: **the sample is set to be 400 in the simulation studies, what happen**

**when the sample size is set to be smaller than 400? How does the Monte Carlo**

**sample size affect the final results?**  **For the simulation study conducted in**

**Section 4.1, what is the standard deviation of the ratio of the sizes of**

**confidence sets from accept-reject ACDC divided by those from IS-ABC? What**

**happen if resampling from $M\_{\theta\_i}$ is difficult to conduct?** Some

comments on this effect would be welcome.

This paper doesn't have any real data analysis example, it would be nice if the

authors could **provide a real data analysis example**.