O'Bayes 2022

Natesh S. Pillai
Department of Statistics, Harvard University

September 8, 2022

Key Idea

- Fearnhead and Prangle (2012) introduce *semi-automatic* ABC.
- Researchers quickly realize many similar algorithms are possible...
- ... and start a large search for “best" choices.

This paper (and talk by Florence Forbes) is a thoughtful addition to this search.

Two Observations in the paper

As the authors observe:

1. "Fact that is obvious in retrospect": Unlike some previous efforts, GLLiM is both (i) usually fast and (ii) able to give good answers even in some complicated/high-dimensional settings.
2. "Fact that is still a mystery": The best results come from a cheap approximation of Wasserstein distance - even though it can be *very* approximate.

What is going on with (2)? What is the cheap Wasserstein really doing here? Does its performance degrade when it looks less like exact Wasserstein? Do other Wasserstein approximations work about as well?

Does it work for ABC-MCMC

Can obviously *implement* a variant of MCMC-ABC with this paper.

- Does this implementation “just work"?
- If not, what modifications are needed?

MCMC Counterexample

One obvious problem: stability! Consider:

- Hyperparameters: $M = k = d = 1$, $f_\theta(z) = \mathcal{N}(z; \theta, 1)$.
- Estimates from data: $A_1 = 2$, $\Sigma_1 = 1$, $b_1 = 0$.
- Resulting posterior probability estimate:

$$p_G(\theta|y;\phi) = \mathcal{N}(\theta; 2y, 1).$$

- Resulting behaviour: big values of θ_t give rise to big samples $y_t \approx \theta_t$ give rise to bigger $\theta_{t+1} \approx 2y_t \approx 2\theta_t$. Explosion!

Questions: can something like the bad initial estimate of A occur? Is it easy to prevent via algorithm modification?

Speculation: One may need to show that the eigenvalues of A(the "coefficient" in GLLiM) are bounded by 1 in the posterior mixture model.

Bad Prior?

- If prior is (arbitrarily) bad, $\mathcal{D}$ will be (arbitrarily) far from neighbourhood of posterior mode...
- ... and p_G can be (arbitrarily) bad *even if the generative model is very close to true in the posterior mode.*
- Possible solution: resample $\mathcal{D}$ during run if you notice mismatch?
- What is the best diagnostic for prior-posterior mismatch in this context?

GLLiM as Model vs Subalgorithm

- Paper implicitly assumes we should treat GLLiM fitting like a "normal" statistical model (e.g. use BIC to choose K)...
- ... this is a natural idea ...
- ... but model fitting is not key here, and there are other natural criteria!
- Wild speculating on alternatives:
  1. **Idea: we can make as much "data" as we want:** Instead of fixing $|\mathcal{D}|$ then choosing best K - could increase $|\mathcal{D}|$ until your model estimate "stops improving" according to some criterion.
  2. **Idea: we don't need model accuracy:** It is well-known that estimates of GMM size K diverge with dataset size $|\mathcal{D}|$, even with BIC regularization. But this captures small "defects" in Gaussianity that are not relevant to ABC (and probably hurt it). How can we stop early with big $|\mathcal{D}|$?

Conclusion

- Interesting paper with lots of potential follow up.
- Key issues need to be understood: "why does the cheap approximation work"?
- Application of this methodology to ABC-MCMC version will be a useful addition.
- Thanks to the organizers, the authors, Christian Robert, Jean-Michel Marin and Aaron Smith.