The paper proposes and studies variants of the MCEM (Monte Carlo Expectation Maximisation) algorithm. This type of algorithm may be expensive for two reasons:

1. Large sample size: each iteration involves a complete pass on the data.
2. Large Monte Carlo sample: which may be required to make sure the expectation to be computed in the E step is approximated with sufficient accuracy.

Several papers have proposed to tackle one of these two issues through some form SA (stochastic approximation). The paper attempts to tackle both issues simultaneously, using various SA schemes that involve two timescales.

I really wanted to like this paper. The basic idea seems simple, but making it work looks non trivial. The paper is really pleasant to read until Section 2.2 (not included). And then, things get pretty confusing. In particular, in Table 1, three different updates are proposed for the $S$ statistic, however:

- The acronyms iSAEM, fiTTEM and vrTTEM are not spelled out!
- These updates may be taken from, or inspired from previous papers, but they are not cited; e.g. I guess iSAEM is a SA version of the iEM (incremental EM) algorithm of Neal and Heaton (1998)?
- There is very little explanation on why these updates should work, why do they differ, which one may be expected to work better... It looks like the authors are trying to come up with a general framework to discuss these three updates?

The theoretical part is also quite unclear. I had to look at several papers (including Fort et al, 2020, which looks relevant and is note cited?) to understand what is going on:

- The fact that the stopping time $K$ is random is hardly explained. Again, I had to look at previous papers to find some explanation, and the fact that this idea is usually attributed to Ghadimi and Lan (2013).
- I also had to look at e.g. Fort et al (2020) to understand the motivation of this type of non-asymptotic result; basically, one bounds the error by some term that scales like $n^{2/3}/K_{\max}$ (where $K_{\max}$ is the maximum number of iterations), and thus, to get an error smaller than some $\varepsilon$, one needs to take $K_{\max} \propto n^{2/3}$.
- In fact, even after reading previous papers, I am still not entirely clear on what to do with the proposed results. In particular, it involves several terms related to the Monte Carlo. What are the implications in terms of choosing the $M_k$'s (number of Monte Carlo sample size at iteration $k$)?
- The paper does not seem to say anything about which update should work better than the others.

By and large, I suspect that the paper may be a bit incremental relative to Karimi et al (2019), at least by Bernoulli standards. I might be wrong, but, again, I am sorry to say that the authors did not make the best job at presenting their ideas. A complete rewrite that addresses the issues raised above should be considered before resubmitting to any journal.

I