Review for TMLR Submission 1503: Do we need to estimate the variance in robust mean estimation?

January 3, 2024

This paper proposes an ERM-style mean estimator for the class of univariate distributions with finite variance, with the goal of beating the standard median of means estimator. Theoretically, the estimator achieves an almost-sub-Gaussian rate, up to a $\sqrt{\log n}$ factor. The author(s) also show numerical experiments to justify their construction.

While the paper improves on Catoni's estimator in the sense of not having to know the variance, the strongest known result for 1-d mean estimation is in fact by Lee and Valiant (FOCS 2021, Optimal Sub-Gaussian Mean Estimation in $\mathbb{R}$), where they achieve the optimal estimation error $(1+o(1))\sigma\sqrt{2\log\frac{1}{\delta}/n}$, with the optimal multiplicative constant of $\sqrt{2}$, **without knowing the variance or any information about it other than the fact it exists**. The multiplicative $1+o(1)$ slack is also independent of the data distribution, and the Lee and Valiant estimator is also "self-tuned" in the sense of this submission.

As such, the Lee and Valiant already solves the problem set out by this work, with provable statistical optimality, while the results of this paper are much weaker, being loose to a $\sqrt{\log n}$ factor. Yet, the Lee and Valiant result was neither discussed nor cited in this paper.

Also, isn't the asymptotic performance of median-of-means also already analyzed by Minsker (2019, Electronic Journal of Statistics, Distributed statistical estimation and rates of convergence in normal approximation)?

In terms of techniques, the conceptual contribution of this paper is at a high-level also similar to the Lee-Valiant estimator yet this paper yields rather sub-optimal results. The key idea in this work is to introduce another parameter in the loss function that represents the scale, and to optimize for it directly. As Lee and Valiant discuss explicitly in their paper, their estimator already is also expressible as a 2-parameter ψ-estimator, where they add an α parameter in addition to the mean estimate $\hat{\mu}$ to handle the same scale-tuning issue. The Lee and Valiant construction and analysis yields a tight result, and an easy-to-implement and fast-to-run estimator, while the 2-parameter M-estimation approach in this paper is quite far from optimal which further requires running an optimization algorithm.