A Framework for Statistical Inference via Randomized Algorithms

Zhixiang Zhang

Department of Mathematics, University of Macau, Macau, P. R. China

Email: zhixzhang@um.edu.mo

We would like to present our latest findings on the following[1]. Randomized algorithms, such as randomized sketching or stochastic optimization, are a promising approach to ease the computational burden in analyzing large datasets. However, randomized algorithms also produce non-deterministic outputs, leading to the problem of evaluating their accuracy.

In this work, we develop a statistical inference framework for quantifying the uncertainty of the outputs of randomized algorithms. Our key conclusion is that one can perform statistical inference for the target of a sequence of randomized algorithms as long as in the limit their outputs fluctuate around the target according to any (possibly unknown) probability distribution. In this setting, we develop appropriate statistical inference methods—sub-randomization, multi-run plug-in and multi-run aggregation—by estimating the unknown parameters of the limiting distribution either using multiple runs of the randomized algorithm, or by tailored estimation. As illustrations, we develop methods for statistical inference for least squares parameters via random sketching (sketch-and-solve, partial and iterative sketching), by characterizing their limiting distribution in a possibly growing dimensional case. Moreover, we also apply our inference framework to stochastic optimization, including for stochastic gradient descent and stochastic optimization with momentum. The results are supported via a broad range of simulations.

References:

[1] Zhang, Z., Lee, S., & Dobriban, E. (2023). A framework for statistical inference via randomized algorithms. *arXiv:2307.11255*.