

Computational limits for distributed estimation

Quentin Berthet

Statistical Laboratory, DPMMS
University of Cambridge

Abstract:

We study the consequences of estimating a high dimensional signal in a distributed manner, by splitting the data in blocks that are processed independently. This process is often motivated by the computational advantages of dealing with smaller datasets, as well as privacy and security concerns. We show that the statistical costs of this method can be particularly high when one is limited to algorithmically efficient estimation procedures.

This result offers a decentralized counterpart to earlier work establishing statistical costs due to resource constraints on centralized procedures. We also show how the introduction of redundancy in the distributed data can be used to overcome these limits, while still preserving some key advantages of these approaches. These results exhibit the existence of multiple and tangled trade-offs between diverse aspects of statistical inference for large datasets.

Joint work with V. Chandrasekaran