TUTORIAL:
"Statistical Inference in Distributed or Constrained Settings: Techniques and Recipes"

Jayadev Acharya* Clément Canonne† Himanshu Tyagi‡

July, 2021

Topic. The goal of this tutorial is to provide the attendees with an overview of techniques and recipes for distributed estimation and testing *under constraints*. Over the recent years, many papers have obtained both upper and lower bounds for statistical estimation under communication, local privacy, and memory constraints (see, e.g., [10, 11, 8, 6, 2, 9, 5, 4, 7, 3, 1], and more): these questions are motivated by applications in machine learning and distributed computing, and are at the intersection of theoretical computer science, machine learning, statistics, and information theory.

Our goal is to provide a primer of those techniques, aiming to provide both an understanding of the underlying challenges and ideas, and "plug-and-play" general recipes the attendees could then apply to the problems of their choice. Our focus will be on establishing lower bounds for statistical estimation, in particular for parameter estimation (single- and high-dimensional) and testing.

The tutorial will cover various models: nonadaptive, sequentially adaptive, blackboard model, and memory-constrained settings; with applications to high-dimensional parameter estimation and testing. (If time allows, we will briefly cover extensions to nonparametric density estimation.)

[15 min Tutorial: Part 0] Setting the stage: scope, models

[45 min Tutorial: Part 1] Lower bounds for learning/estimation problems

[45 min Tutorial: Part 2] Lower bounds for learning/estimation problems

[45 min Recitation-style practice/clarification session] Applications to mean estimation under local privacy or communication constraints; derivation of optimal bounds, comparison between existing methods.

References

[1] Jayadev Acharya, Clément L. Canonne, Ziteng Sun, and Himanshu Tyagi. Unified lower bounds for interactive high-dimensional estimation under information constraints. *CoRR*, abs/2010.06562, 2020.

[2] Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints: Lower bounds from chi-square contraction. In *COLT*, volume 99 of *Proceedings of Machine Learning Research*, pages 3–17. PMLR, 2019.

[3] Jayadev Acharya, Clément L. Canonne, Yuhan Liu, Ziteng Sun, and Himanshu Tyagi. Interactive inference under information constraints. In *ISIT*, 2021.

[4] Kareem Amin, Matthew Joseph, and Jieming Mao. Pan-private uniformity testing. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 183–218. PMLR, 09–12 Jul 2020.

[5] Leighton P. Barnes, Wei-Ning Chen, and Ayfer Özgür. Fisher information under local differential privacy. *CoRR*, abs/2005.10783, 2020.

[6] Leighton P. Barnes, Yanjun Han, and Ayfer Özgür. Fisher information for distributed estimation under a blackboard communication protocol. In *ISIT*, pages 2704–2708. IEEE, 2019.

[7] Thomas B. Berrett and Cristina Butucea. Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. In *NeurIPS*, 2020.

[8] Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Symposium on Theory of Computing Conference, STOC'16*, pages 1011–1020. ACM, 2016.

[9] Ilias Diakonikolas, Themis Gouleakis, Daniel M. Kane, and Sankeerth Rao. Communication and memory efficient testing of discrete distributions. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1070–1106, Phoenix, USA, jun 2019. PMLR.

[10] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013*, pages 429–438. IEEE Computer Society, 2013.

[11] Ohad Shamir. Fundamental limits of online and distributed algorithms for statistical learning and estimation. In *Advances in Neural Information Processing Systems 27*, pages 163–171, 2014.

*Email: acharya@cornell.edu
†Email: clement.canonne@sydney.edu.au
‡Email: htyagi@iisc.ac.in