## Learning Discrete Distributions

**Varun Kanade (varunk@di.ens.fr)**

Consider the problem of estimating a distribution defined over the set $\{1, 2, \ldots, n\}$, when we only see samples from the distribution. When no further assumption is made, a folklore result says that we need roughly $n^2/\epsilon$ examples to output a distribution that $\epsilon$-close in variation distance. However, when one makes further assumptions about the distribution, such as unimodality/bi-modality, monotonicity, etc., much better results are possible. This has been a new active area in theoretical computer science [1], with close connections to property testing. An interesting aspect of some of the results is an apparent gap between the number of samples needed for estimation with a computationally efficient algorithm vs that with unbounded computation. The goal in this project will be to survey two to three papers from the list below, or related papers, and identify interesting open questions for future work.

# References

[1] Siu-On Chan, Ilias Diakonikolas, Rocco Servedio, and Xiaorui Sun. Learning mixtures of structured distributions over discrete domains. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013.

[2] Siu-On Chan, Ilias Diakonikolas, Rocco Servedio, and Xiaorui Sun. Efficient density estimation via piecewise polynomial approximation. In *Proceedings of the 46th Annual ACM Symposium on the Theory of Computing (STOC)*, 2014.

[3] Siu-On Chan, Ilias Diakonikolas, Rocco Servedio, and Xiaorui Sun. Near-optimal density estimation in near-linear time using variable-width histograms. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[4] Constantinos Daskalakis, Ilias Diakonikolas, Ryan O'Donnell, Rocco Servedio, and Li-Yang Tan. Learning sums of independent integer random variables. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2013.

[5] Constantinos Daskalakis, Ilias Diakonikolas, and Rocco Servedio. Learning k-modal distributions via testing. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2012.

[6] Constantinos Daskalakis, Ilias Diakonikolas, and Rocco Servedio. Learning poisson binomial distributions. In *Proceedings of the 44th Annual ACM Symposium on the Theory of Computing (STOC)*, 2012.

[1] There was a one-day workshop at STOC 2014: `www.iliasdiakonikolas.org/stoc-distribution-estimation.html`