Algorithms and Limits in Statistical Inference

Jayadev Acharya

Massachusetts Institute of Technology

Statistical Inference

Given samples from an unknown source P

Learn P?

mixture of Gaussians, Log-concave, etc

- Pearson (1894), …, Redner ,Walker (1984), …, Dasgupta (1999), …, Moitra, Valiant (2010),…
- Devroye, Lugosi (2001), Bagnoli, Bergstrom (2005), …, Wellner, Samworth et al

Density estimation of mixture of Gaussians with information theoretically optimal samples, and linear run time?

Test if P has a property $\mathcal{P}$?

Is P monotone, product distribution, etc

Traditional Statistics: samples $\rightarrow \infty$

- Pearson's chi-squared tests, Hoeffding's test, GLRT, …
  error rates
- Batu et al (2000, 01, 04), Paninski (2008), …,
  sample and computational efficiency

Sample optimal and efficient testers for monotonicity, and independence over $[k]\times[k]\times[k]$?

Illustrative Results: Learning [Acharya-Diakonikolas-Li-Schmidt'15]

Agnostic univariate density estimation with t-piece d-degree polynomial

$$\mathrm{O}\left(\frac{\mathrm{t}(\mathrm{d}+1)}{\varepsilon^2}\right) \text{ samples, } \widetilde{\mathrm{O}}\left(\frac{\mathrm{t}\cdot\mathrm{poly}(\mathrm{d})}{\varepsilon^2}\right) \text{ run time}$$

First near sample-optimal, linear-time algorithms for learning:

- Piecewise flat distributions
- Mixtures of Gaussians
- Mixtures of log-concave distributions
- Densities in Besov spaces, …

Illustrative Results: Testing

[Acharya-Daskalakis-Kamath'15]

Sample complexity to test if $P \in \mathcal{P}$, or $d_{TV}(P, \mathcal{P}) > \varepsilon$,

For many classes, optimal complexity: $\sqrt{|domain|}$

- Applications:
  - Independence, monotonicity over $[k]^d$: $\Theta(\frac{k^{d/2}}{\varepsilon^2})$
  - Log-concavity, unimodality over $[k]$: $\Theta(\frac{\sqrt{k}}{\varepsilon^2})$
- Based on:
  - a new χ^2-ℓ_1 test
  - a modified Pearson's chi-squared statistic