Differentially Private Testing of Properties of Distributions

Maryam Aliakbarpour
MIT

Joint work with Ilias Diakonikolas (USC) and Ronitt Rubinfeld (MIT, TAU)
To appear in ICML 2018
Property testing of distributions

The Model:

- **Unknown Distribution** \( D \)
- **iid samples** \( x_1, x_2, ..., x_s \)
- **The Tester**
  - Accept
  - Reject

[Rubinfeld and Sudan’96, Goldreich and Ron’00, Batu, Fortnow, Rubinfeld, Smith, and White’00, ...]
Differential privacy

- Any possible output $O$
- Two neighboring data set $X$, $X'$ s.t. $|X - X'| = 1$

Main Question:
Can we test properties of distribution with respect to differential privacy? optimal sample complexity?

[Diakonikolas, Hardt, and Schmidt’15, Cai, Daskalakis, and Kamath’16, ...] [In an independent work: Acharya, Sun, Zhang’17]
Problems: Testing uniformity

 iid samples $x_1, x_2, \ldots, x_s$

Is $D$ Uniform, or $\epsilon$-far from being uniform?

Accept or Reject?

Sample Complexity:

- **When** $\epsilon = \Omega(n^{1/4})$: $O(\sqrt{n}/\epsilon^2 + \sqrt{n}/(\epsilon \xi))$
- **General case**: $\tilde{O}(\sqrt{n}/\epsilon^2 + \sqrt{n}/(\epsilon \xi) + 1/\epsilon^2 \xi)$

[Paninski’08, Batu, Fortnow, Rubinfeld, Smith, and White’13, Valiant and Valiant’14, Chan, Diakonikolas, Valiant, and Vaient’14, Diakonikolas, Gouleakis, Peebles, and E. Price’16, …]
Problems: Testing Identity (Goodness of Fit)

iid samples $x_1, x_2, ..., x_s$ from distribution $D$

Explicitly given Distribution $q$

Is $D$ equal to $q$, or $\epsilon$-far from it?

Accept or Reject

Sample Complexity:

- When $\epsilon = \Omega(n^{1/4})$: $O(\sqrt{n}/\epsilon^2 + \sqrt{n}/(\epsilon \sqrt{\xi}))$
- General case: $\tilde{O}(\sqrt{n}/\epsilon^2 + \sqrt{n}/(\epsilon \xi) + 1/\epsilon^2 \xi)$
Problems: Testing Closeness (Equivalence)

- iid samples $x_1, x_2, \ldots, x_s$ from distribution $p$
- iid samples $y_1, y_2, \ldots, y_s$ from distribution $q$

Is $p$ equal to $q$, or $\epsilon$-far from it?

Accept or Reject

Sample Complexity: $O(n^{2/3} \epsilon^{4/3} + \sqrt{n}/\epsilon^2 + \sqrt{n}/(\epsilon \sqrt{\xi}) + 1/\epsilon^2 \xi)$
General Framework

$X \xrightarrow{} Z \xrightarrow{\text{Laplace noise}} \tilde{Z}$

Laplace noise

$L[\tilde{Z}_{\text{accept}}] \quad L[\tilde{Z}_{\text{reject}}]$

Lipschitz statistics?  Multiple statistics?
Thank you