A Chasm Between Identity and Equivalence Testing with Conditional Queries

Joint work with Jayadev Acharya and Gautam Kamath
Outline of the talk

- Distribution Testing and Conditional Queries
- Our results
- Overview of the techniques and obstacles
- Open problems
Distribution Testing

Subfield of property testing:

- **Big** (Unknown) Object O
- Fixed **property** (subset of Objects) P
- Is O in P, or far from every object in P?
Distribution Testing

Big Object: probability distribution D over n elements

Type of queries: independent samples from D

Distance measure (for "far"): Total variation distance \( (L_1) \)
Distribution Testing: A Glimpse

Many results since seminal work of [BFRSW01]:

Testing uniformity: $\Theta(\sqrt{n}/\epsilon^2)$ [GR00,Pan08]
Testing identity: $\Theta(\sqrt{n}/\epsilon^2)$ [BFF+01,VV14]
Testing equivalence: $\Theta(n^{2/3}/\epsilon^{4/3})$ [BFR+10,CDVV14]

(+monotonicity, independence, membership to a class...)

Polynomial dependence on the domain size
Many results since seminal work of [BFRSW01]:

- Testing uniformity: $\Theta(\sqrt{n}/\epsilon^2)$ [GR00,Pan08]
- Testing identity: $\Theta(\sqrt{n}/\epsilon^2)$ [BFF+01,VV14]
- Testing equivalence: $\Theta(n^{2/3}/\epsilon^{4/3})$ [BFR+10,CDVV14]

(\text{+monotonicity, independence, membership to a class...})

Polynomial dependence on the domain size
Distribution Testing: A Twist

“Is $n^{\Theta(1)}$ our final answer?”

More power to the testers: conditional sampling
[Chakraborty-Fischer-Goldhirsh-Matsliah'13, C-Ron-Servedio'12]

Tester chooses $S \subseteq [n]$
Gets sample drawn from $D_S$
Distribution Testing: A Twist

- Generalizes sampling model
- Allows adaptivity
- (Several restricted variants)
- Natural in some settings

“Power of comparisons”
Distribution Testing in $O(1)$

Testing **uniformity**: $O(1)$  \[\text{[CRS15]}\]

Testing **identity**: $O(1)$  \[\text{[CRS15,FJO+15]}\]

Testing **equivalence**: $\log \log n$  \[\text{[CRS15,FJO+15]}\]

“Should it be constant-query too?”
Distribution Testing in $\Theta(1)$

Testing uniformity: $O(1)$ \cite{CRS15}

Testing identity: $O(1)$ \cite{CRS15,FJO15}

Testing equivalence: $\Omega(\sqrt{\log \log n})$ \cite{this work}

“Things are different now.”
Our main result

**Theorem.** Any (adaptive) testing algorithm for equivalence in the conditional model must make $\Omega(\sqrt{\log\log n})$ queries.

“*Know thy enemy — it helps for testing.*”
Obstacles

- **How to deal with adaptivity?**
  
  *Lack of general lower bound techniques in this model.*

- **How to deal with arbitrary queries?**
  
  *The tester can pretty much do what it wants.*

- **Cannot** rely on the “usual tricks”
  
  *Heavy elements to hide the rest, only a few types of weights...*
Ideas

- Bring in the concept of adaptive core tester of [CFGM13], + Yao's principle

- Adapt a lower bound construction of [CRS15] for a restricted conditional model (easy to beat in the full one).

- **Hide it** by scaling its support by a random factor.
Construction

$D_j(i)$

$B_1 B_2 B_3 B_4 \ldots m$
Intuition

If you don't (roughly) guess the support size, you cannot learn anything...
But we show support size estimation is hard.
Open Questions

- Can we get $\varepsilon$ in the picture?
- Does tolerant testing behave like this? (chasm!)
- Develop general techniques for proving lower bounds?
- Characterize problems that are still hard in this model?
Thank You.