Testing with Alternative Distances

Gautam "G" Kamath

FOCS 2017 Workshop: Frontiers in Distribution Testing

October 14, 2017

Based on joint works with

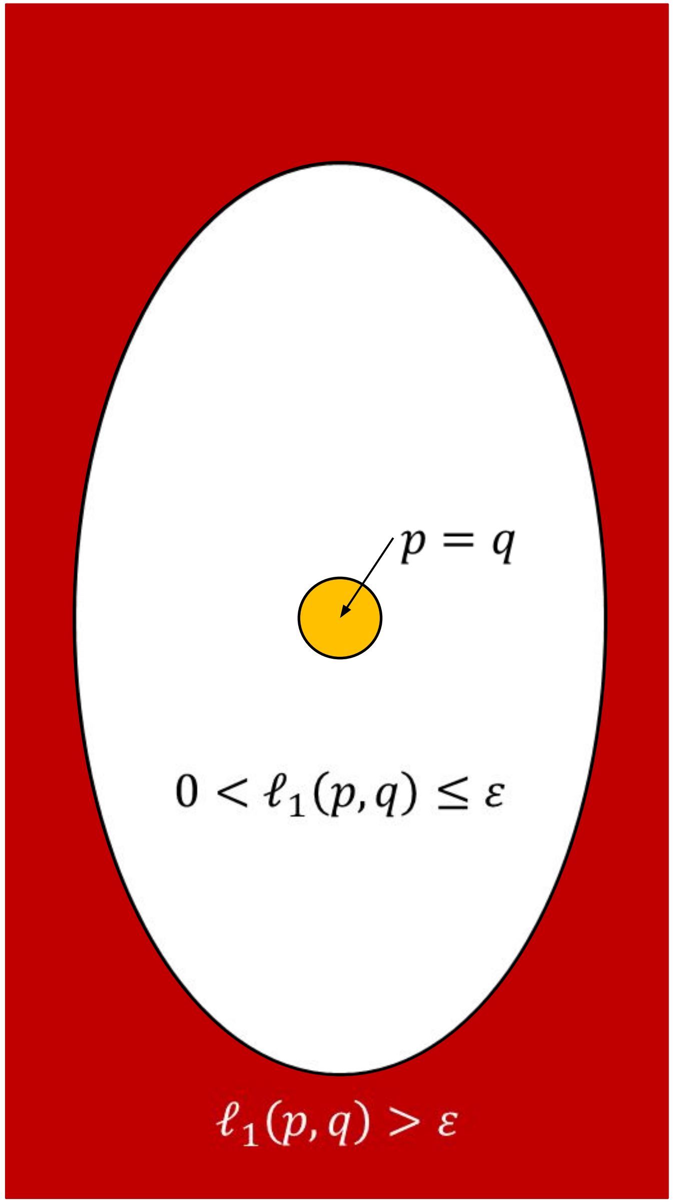
Jayadev Acharya
Cornell

Constantinos Daskalakis
MIT

John Wright
MIT

The story so far…

- Test whether $p = q$ versus $\ell_1(p, q) \geq \varepsilon$
  - Domain of $[n]$
  - Success probability $\geq 2/3$
  - Goal: Strongly sublinear sample complexity
    - $O(n^{1-\gamma})$ for some $\gamma > 0$
- Identity testing (samples from p, known q)
  - $\Theta(\sqrt{n}/\varepsilon^2)$ samples
  - [BFFKR'01, P'08, VV'14]
- Closeness testing (samples from p, q)
  - $\Theta(\max\{n^{2/3}/\varepsilon^{4/3}, \sqrt{n}/\varepsilon^2\})$ samples
  - [BFRSW'00, V'11, CDVV'14]



Generalize: Different Distances

- $p = q$ or $\ell_1(p, q) \geq \varepsilon$?

- $d_1(p, q) \leq \varepsilon_1$ or $d_2(p, q) \geq \varepsilon_2$?

Generalize: Different Distances

- $d_1(p,q) \leq \varepsilon_1$ or $d_2(p,q) \geq \varepsilon_2$?
  - Are p and q ε_1-close in $d_1(.,.)$, or ε_2-far in $d_2(.,.)$?
  - Distances of interest: ℓ_1, ℓ_2, χ^2, KL, Hellinger
- Classic identity testing: $\varepsilon_1 = 0, d_2 = \ell_1$
- Can we characterize sample complexity for each pair of distances?
  - Which distribution distances are sublinearly testable? [DKW'18]
- Wait, but… why?

Wait, but… why?

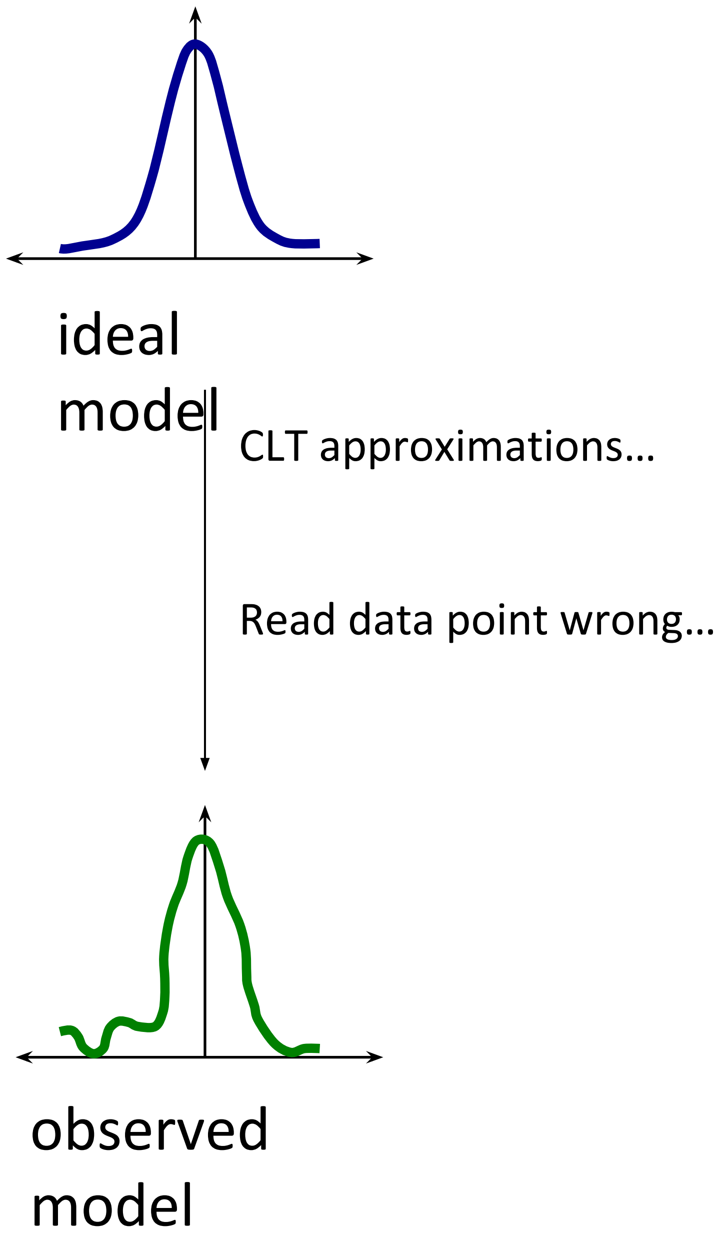
1. Tolerance for model misspecification
2. Useful as a proxy in classical testing problems
   - d_1 as χ^2 distance is useful for composite hypothesis testing
   - Monotonicity, independence, etc. [ADK'15]
3. Other distances are natural in certain testing settings
   - d_2 as Hellinger distance is sometimes natural in multivariate settings
   - Bayes networks, Markov chains [DP'17,DDG'17]
   - Costis' talk

Wait, but… why?

1. Tolerance for model misspecification
2. Useful as a proxy in classical testing problems
   - d_1 as χ^2 distance is useful for composite hypothesis testing
   - Monotonicity, independence, etc. [ADK'15]
3. Other distances are natural in certain testing settings
   - d_2 as Hellinger distance is sometimes natural in multivariate settings
   - Bayes networks, Markov chains [DP'17,DDG'17]
   - Costis' talk

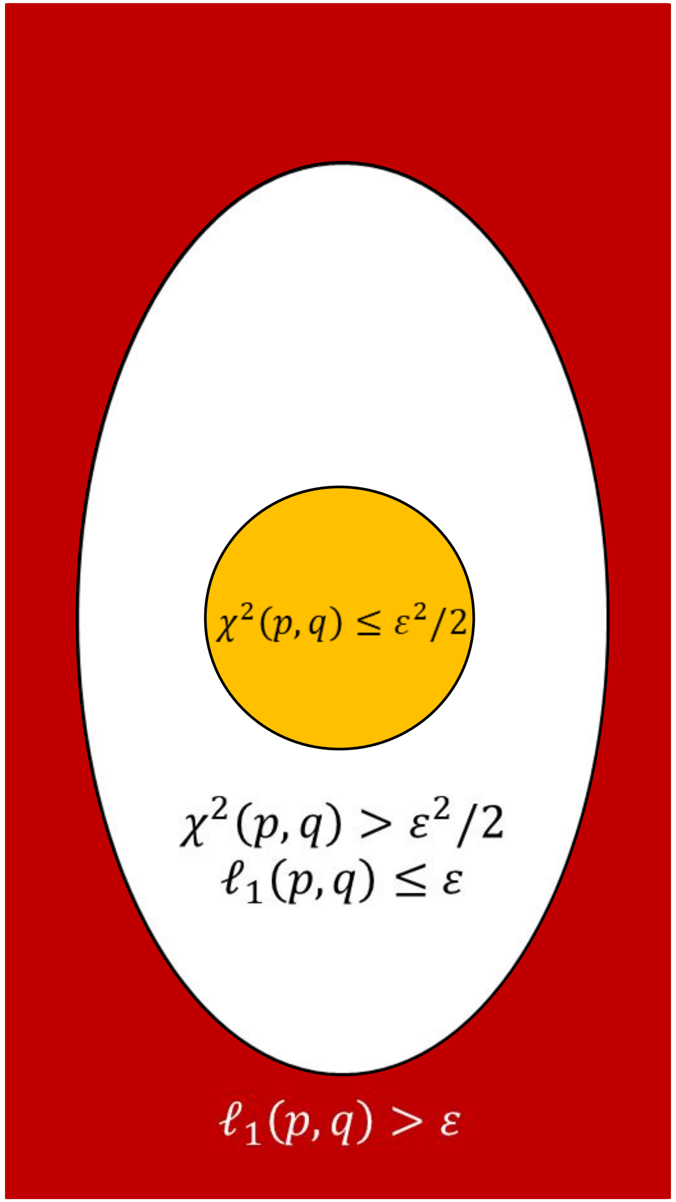
Tolerance

- Is p equal to q, or are they far from each other?
  - But why do we know q exactly?
- Models are inexact
  - Measurement errors
  - Imprecisions in nature
- p, q may be "philosophically" equal, but not literally equal
- When can we test $d_1(p,q) \leq \varepsilon_1$ versus $\ell_1(p,q) \geq \varepsilon$?



Tolerance

- $d_1(p,q) \le \varepsilon_1$ vs. $\ell_1(p,q) \ge \varepsilon$?
  - What d_1? How about ℓ_1?
- $\ell_1(p,q) \le \varepsilon/2$ vs. $\ell_1(p,q) \ge \varepsilon$?
  - No! $\Theta(n/\log n)$ samples [VV'10]
- Chill out, relax…
- χ^2-distance: $\chi^2(p,q) = \sum_{i\in\Sigma} \frac{(p_i - q_i)^2}{q_i}$
  - Cauchy-Schwarz: $\chi^2(p,q) \ge \ell_1^2(p,q)$
- $\chi^2(p,q) \le \varepsilon^2/4$ vs. $\ell_1(p,q) \ge \varepsilon$?
  - Yes! $O(\sqrt{n}/\varepsilon^2)$ samples [ADK'15]



Details for a χ^2-Tolerant Tester

- Goal: Distinguish (i) $\chi^2(p,q) \leq \frac{\varepsilon^2}{2}$ versus (ii) $\ell_1^2(p,q) \geq \varepsilon^2$
- Draw $Poisson(m)$ samples from p ("Poissonization")
  - N_i: number of appearances of symbol i
    - $N_i \sim Poisson(m \cdot p_i)$
  - N_i's are now independent!
- Statistic: $Z = \sum_{i \in \Sigma} \frac{(N_i - m \cdot q_i)^2 - N_i}{m \cdot q_i}$

Acharya, Daskalakis, K. Optimal Testing for Properties of Distributions. NIPS 2015

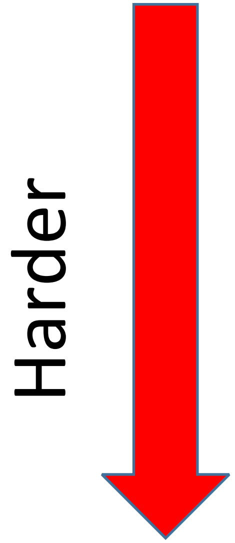
Details for a χ^2-Tolerant Tester

- Goal: Distinguish (i) $\chi^2(p,q) \le \frac{\varepsilon^2}{2}$ versus (ii) $\ell_1^2(p,q) \ge \varepsilon^2$
- Statistic: $Z = \sum_{i\in[n]} \frac{(N_i - m\cdot q_i)^2 - N_i}{m\cdot q_i}$
  - N_i: # of appearances of i; m: # of samples
  - $E[Z] = m \cdot \chi^2(p,q)$
    - (i): $E[Z] \le m \cdot \frac{\varepsilon^2}{2}$, (ii): $E[Z] \ge m \cdot \varepsilon^2$
  - Can bound variance of Z with some work
    - Need to avoid low prob. elements of q
      1. Either ignore i such that $q_i \le \frac{\varepsilon^2}{10n}$; or
      2. Mix lightly ($O(\varepsilon^2)$) with uniform distribution (also in [G'16])
  - Apply Chebyshev's inequality

Side-Note:
- Pearson's χ^2-test uses statistic $\sum_i \frac{(N_i - m\cdot q_i)^2}{m\cdot q_i}$
- Subtracting N_i in the numerator gives an unbiased estimator and importantly may hugely decrease variance
- [Zelterman'87]
- [VV'14, CDVV'14, DKN'15]

Acharya, Daskalakis, K. Optimal Testing for Properties of Distributions. NIPS 2015

Tolerant Identity Testing



	$d_{\mathrm{TV}}(p,q) \geq \varepsilon$
$p = q$	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ [Pan08]
$d_{\chi^2}(p,q) \leq \varepsilon^2/4$	$O\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ [Theorem 1]
$d_{\mathrm{KL}}(p,q) \leq \varepsilon^2/4$	$\Omega\left(\frac{n}{\log n}\right)$ [Theorem 8]
$d_{\mathrm{H}}(p,q) \leq \varepsilon/2\sqrt{2}$	
$d_{\mathrm{TV}}(p,q) \leq \varepsilon/2$ or $\varepsilon^2/4^4$	$O\left(\frac{n}{\log n}\right)$ [Corollary 3]

$d_{\ell_2}(p,q) \leq \frac{\varepsilon}{\sqrt{n}}$ vs $d_{\mathrm{TV}}(p,q) \geq \varepsilon$	$\Theta\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ [Theorem 2]

(Implicit in [DK'16])

Daskalakis, K., Wright. Which Distribution Distances are Sublinearly Testable? SODA 2018.

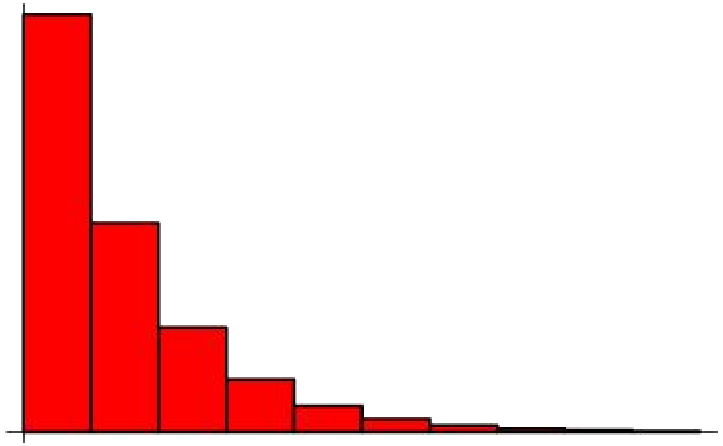
Tolerant Testing Takeaways

1. Can handle ℓ_2 or χ^2 tolerance at no additional cost
   - $\Theta(\sqrt{n}/\varepsilon^2)$ samples
2. KL, Hellinger, or ℓ_1 tolerance are expensive
   - $\Theta(n/\log n)$ samples
   - KL result based off hardness of entropy estimation
3. Closeness testing (q unknown): Even χ^2 tolerance is costly!
   - $\Theta(n/\log n)$ samples
   - Only ℓ_2 tolerance is free
   - Proven via hardness of ℓ_1-tolerant *identity* testing
   - Since q is unknown, χ^2 is no longer a polynomial

Application: Testing for Structure

- Composite hypothesis testing
- Test against a class of distributions!
  - $p \in C$ versus $\ell_1(p, C) > \varepsilon$

    $\min_{q \in C} \ell_1(p, q)$
- Example: C = all monotone distributions
  - p_i's are monotone non-increasing
  - Others: unimodality, log-concavity, monotone hazard rate, independence
  - All can be tested in $\Theta(\sqrt{n}/\varepsilon^2)$ samples [ADK'15]
    - Same complexity as vanilla uniformity testing!



Testing by Learning

- Goal: Distinguish $p \in C$ from $\ell_1(p, C) > \varepsilon$
- Learn-then-Test:
  1. Learn hypothesis $q \in C$ such that
     - $p \in C \Rightarrow \chi^2(p, q) \leq \varepsilon^2/2$ (needs cheap "proper learner" in χ^2)
     - $\ell_1(p, C) > \varepsilon \Rightarrow \ell_1(p, q) > \varepsilon$ (automatic since $q \in C$)
  2. Perform "tolerant testing"
     - Given sample access to p and description of q, distinguish $\chi^2(p, q) \leq \varepsilon^2/2$ from $\ell_1(p, q) > \varepsilon$
- Tolerant testing (step 2) is $O(\sqrt{n}/\varepsilon^2)$
  - Naïve approach (using ℓ_1 instead of χ^2) would require $\Omega(n/\log n)$
- Proper learners in χ^2 (step 1)?
  - Claim: This is cheap

Hellinger Testing

- Change d_2 instead of d_1
- Hellinger distance: $H^2(p,q) = \frac{1}{2}\sum_{i\in[n]}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2$
  - Between linear and quadratic relationship with ℓ_1
  - $H^2(p,q) \leq \ell_1(p,q) \leq H(p,q)$
- Natural distance when considering a collection of iid samples
  - Comes up in some multivariate testing problems (Costis @ 2:55)
- Testing $p = q$ vs. $H(p,q) \geq \varepsilon$?
- Trivial results via ℓ_1 testing
  - Identity: $O(\sqrt{n}/\varepsilon^4)$ samples
  - Closeness: $O(\max\{n^{2/3}/\varepsilon^{8/3}, \sqrt{n}/\varepsilon^4\})$ samples

Hellinger Testing

- Testing $p = q$ vs. $H(p, q) \geq \varepsilon$?
- Trivial results via ℓ_1 testing
  - Identity: $O(\sqrt{n}/\varepsilon^4)$ samples
  - Closeness: $O(\max\{n^{2/3}/\varepsilon^{8/3}, \sqrt{n}/\varepsilon^4\})$ samples
- But you can do better!
  - Identity: $\Theta(\sqrt{n}/\varepsilon^2)$ samples
    - No extra cost for ℓ_2 or χ^2 tolerance either!
  - Closeness: $\Theta(\min\{n^{2/3}/\varepsilon^{8/3}, n^{3/4}/\varepsilon^2\})$ samples
    - LB and previous UB in [DK'16]
- Similar chi-squared statistics as [ADK'15] and [CDVV'14]
  - Some tweaks and more careful analysis to handle Hellinger

Daskalakis, K., Wright. Which Distribution Distances are Sublinearly Testable? SODA 2018.

Miscellanea

- $p = q$ vs. $KL(p, q) \geq \varepsilon$?
  - Trivially impossible, due to ratio between p_i and q_i
  - $p_i = \delta, q_i = 0, \delta \to 0$
- Upper bounds for $\Omega(n/\log n)$ testing problems?
  - i.e., $KL(p, q) \leq \varepsilon^2/4$ vs. $\ell_1(p, q) \geq \varepsilon$?
  - Use estimators mentioned in Jiantao's talk

Thanks!

	$d_{\mathrm{TV}}(p,q) \geq \varepsilon$	$d_{\mathrm{H}}(p,q) \geq \varepsilon/\sqrt{2}$	$d_{\mathrm{KL}}(p,q) \geq \varepsilon^2$	$d_{\chi^2}(p,q) \geq \varepsilon^2$
$p = q$	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ [Pan08]		Untestable [Theorem 7]	
$d_{\chi^2}(p,q) \leq \varepsilon^2/4$		$O\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ [Theorem 1]		
$d_{\mathrm{KL}}(p,q) \leq \varepsilon^2/4$	$\Omega\left(\frac{n}{\log n}\right)$ [Theorem 8]			
$d_{\mathrm{H}}(p,q) \leq \varepsilon/2\sqrt{2}$				
$d_{\mathrm{TV}}(p,q) \leq \varepsilon/2$ or $\varepsilon^2/4^4$		$O\left(\frac{n}{\log n}\right)$ [Corollary 3]		

Table 1: Identity Testing

	$d_{\mathrm{TV}}(p,q) \geq \varepsilon$	$d_{\mathrm{H}}(p,q) \geq \varepsilon/\sqrt{2}$	$d_{\mathrm{KL}}(p,q) \geq \varepsilon^2$	$d_{\chi^2}(p,q) \geq \varepsilon^2$
$p = q$	$O\left(\max\left\{\frac{n^{1/2}}{\varepsilon^2}, \frac{n^{2/3}}{\varepsilon^{4/3}}\right\}\right)$ [CDVV14] $\Omega\left(\max\left\{\frac{n^{1/2}}{\varepsilon^2}, \frac{n^{2/3}}{\varepsilon^{4/3}}\right\}\right)$ [CDVV14]	$O\left(\min\left\{\frac{n^{3/4}}{\varepsilon^2}, \frac{n^{2/3}}{\varepsilon^{8/3}}\right\}\right)$ [Theorem 5] $\Omega\left(\min\left\{\frac{n^{3/4}}{\varepsilon^2}, \frac{n^{2/3}}{\varepsilon^{8/3}}\right\}\right)$ [DK16]	Untestable [Theorem 7]	
$d_{\chi^2}(p,q) \leq \varepsilon^2/4$	$\Omega\left(\frac{n}{\log n}\right)$ [Theorem 9]			
$d_{\mathrm{KL}}(p,q) \leq \varepsilon^2/4$				
$d_{\mathrm{H}}(p,q) \leq \varepsilon/2\sqrt{2}$				
$d_{\mathrm{TV}}(p,q) \leq \varepsilon/2$ or $\varepsilon^2/4^4$		$O\left(\frac{n}{\log n}\right)$ [Corollary 3]		

Table 2: Equivalence Testing