L_p-*Testing*

Sofya Raskhodnikova

Penn State University,

visiting Boston University and Harvard

Joint work with

Piotr Berman (Penn State),

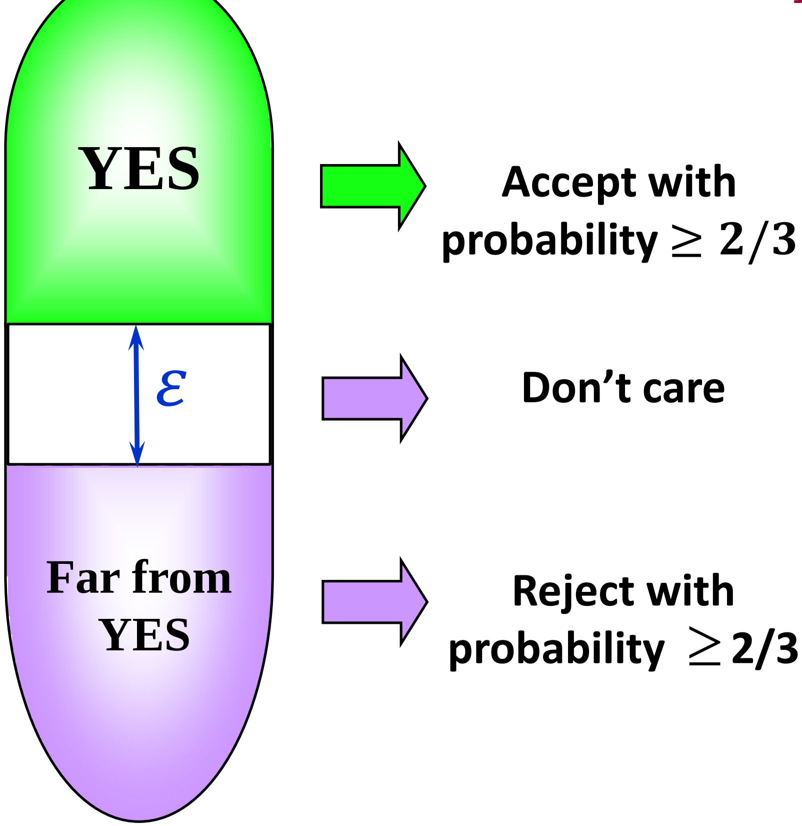
Grigory Yaroslavtsev (Penn State → Brown)

PENNSTATE

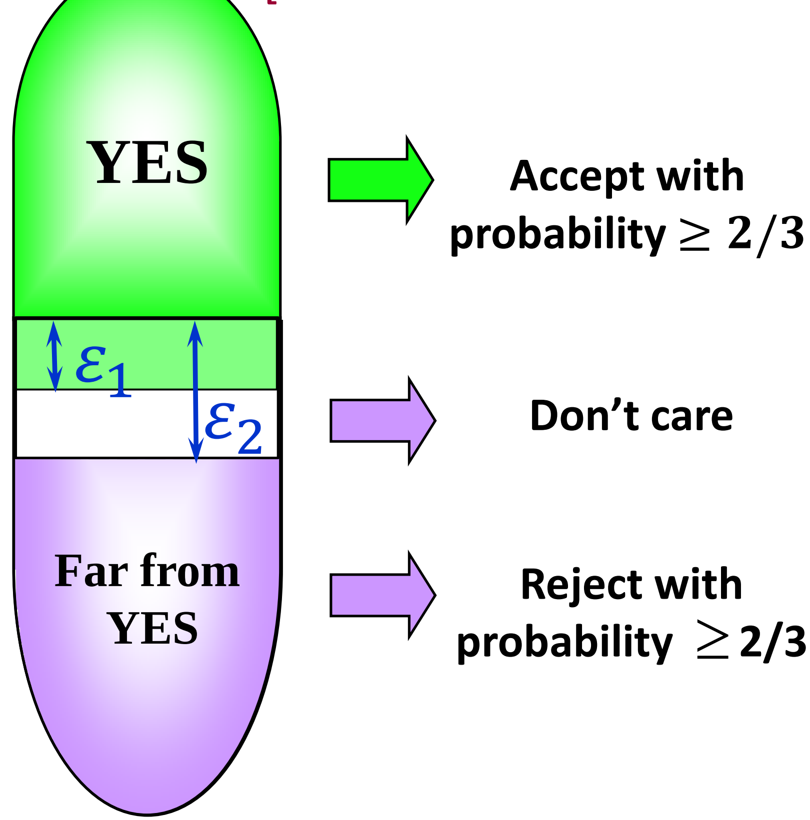
1 8 5 5

Property Testing Models

Property Tester [Rubinfeld Sudan 96, Goldreich Goldwasser Ron 98]



Tolerant Property Tester [Parnas Ron Rubinfeld 06]



Equivalent to tolerant testing: estimating distance to the property.

Two objects are at distance ε = they differ in an ε fraction of places

Why Hamming Distance?

- Nice probabilistic interpretation
  - probability that two functions differ on a random point in the domain
- Natural measure for
  - algebraic properties (linearity, low degree)
  - properties of graphs and other combinatorial objects
- Motivated by applications to probabilistically checkable proofs (PCPs)
- It is equivalent to other natural distances for
  - properties of Boolean functions

Which stocks grew steadily?

Microsoft

IBM®

Data from *http://finance.google.com*

L_p-Testing

for properties of real-valued data

Use L_p-metrics to Measure Distances

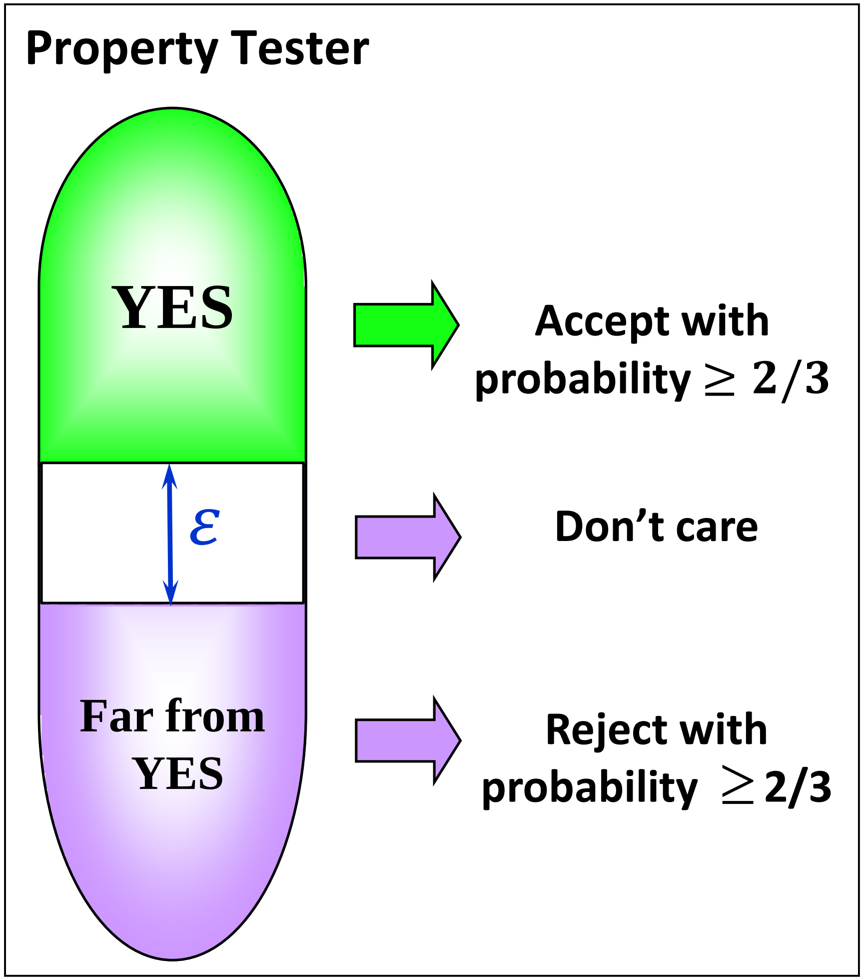
- Functions $f, g\colon\ D \to [0,1]$ over (finite) domain D
- For $p \geq 1$

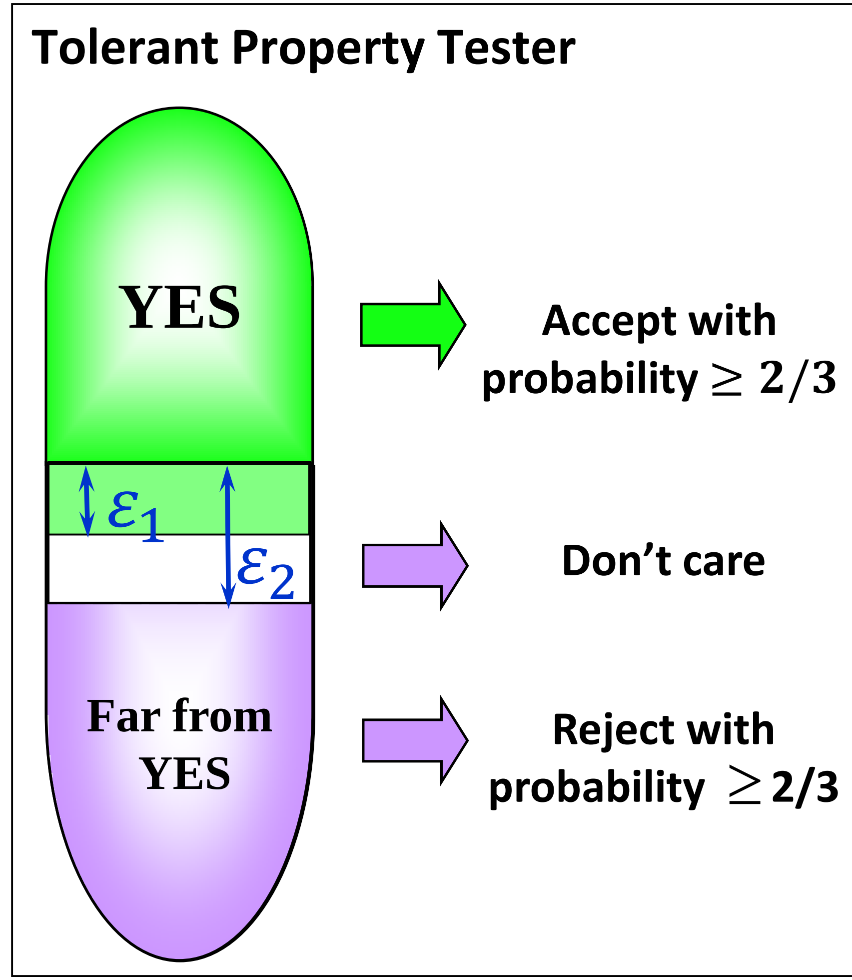
$$L_p(f,g) = \left|\left|f - g\right|\right|_p = \left(\sum_{x \in D} |f(x) - g(x)|^p\right)^{1/p}$$

$$L_0(f,g) = \left|\left|f - g\right|\right|_0 \ = \ |\{x \in D\colon f(x) \neq g(x)\}|$$

- $d_p(f,g) = \dfrac{\left|\left|\boldsymbol{f} - g\right|\right|_p}{\left|\left|1\right|\right|_{\boldsymbol{p}}}$

L_p-Testing and Tolerant L_p-Testing





Functions $f, g\colon D \to [0,1]$ are at distance ε if $d_p = \frac{\|f-g\|_p}{\|\mathbf{1}\|_p} = \varepsilon$.

New L_p-Testing Model for Real-Valued Data

- Generalizes standard L_0-testing

- For $p > 0$ still have a nice probabilistic interpretation: distance $d_p(f, g) = (\mathbf{E}[|\boldsymbol{f} - \boldsymbol{g}|^{\boldsymbol{p}}])^{1/p}$

- Compatible with existing PAC-style learning models (preprocessing for model selection)

- For Boolean functions, $d_0(f, g) = d_p(f, g)^p$.

Our Contributions

1. Relationships between L_p-testing models
2. Algorithms
   - L_p-testers for $p \geq 1$
     - monotonicity, Lipschitz, convexity
   - Tolerant L_p-tester for $p \geq 1$
     - monotonicity in 1D (aka sortedness)

   ❖ Our L_p-testers beat lower bounds for L_0-testers
   ❖ Simple algorithms backed up by involved analysis
   ❖ Uniformly sampled (or easy to sample) data suffices

3. Nearly tight lower bounds

Implications for L_0-Testing

Some techniques/observations/results carry over to L_0-testing

- Improvement on Levin's work investment strategy

  Gives improvements in run time of testers for

  - Connectivity of bounded-degree graphs [Goldreich Ron 02]
  - Properties of images [R 03]
  - Multiple-input problems [Goldreich 13]

- First example of monotonicity testing problem where adaptivity helps
- Improvements to L_0-testers for Boolean functions

Relationships between L_p-Testing Models

Relationships Between L_p-Testing Models

$C_{\boldsymbol{p}}(\boldsymbol{P}, \boldsymbol{\varepsilon})$ = complexity of $L_{\boldsymbol{p}}$-testing property $\boldsymbol{P}$ with distance parameter $\boldsymbol{\varepsilon}$

- e.g., query or time complexity
- for general or restricted (e.g., nonadaptive) tests

For all properties $\boldsymbol{P}$

- $L_{\boldsymbol{1}}$-testing is no harder than Hamming testing

$$C_{\boldsymbol{1}}(\boldsymbol{P}, \boldsymbol{\varepsilon}) \leq C_{\boldsymbol{0}}(\boldsymbol{P}, \boldsymbol{\varepsilon})$$

- $L_{\boldsymbol{p}}$-testing for $\boldsymbol{p} > 1$ is close in complexity to $L_{\boldsymbol{1}}$-testing

$$C_{\boldsymbol{1}}(\boldsymbol{P}, \boldsymbol{\varepsilon}) \leq C_{\boldsymbol{p}}(\boldsymbol{P}, \boldsymbol{\varepsilon}) \leq C_{\boldsymbol{1}}(\boldsymbol{P}, \boldsymbol{\varepsilon}^{\boldsymbol{p}})$$

Relationships Between L_p-Testing Models

$C_{\boldsymbol{p}}(\boldsymbol{P},\boldsymbol{\varepsilon})$ = complexity of $L_{\boldsymbol{p}}$-testing property $\boldsymbol{P}$
with distance parameter $\boldsymbol{\varepsilon}$

- e.g., query or time complexity
- for general or restricted (e.g., nonadaptive) tests

For properties of Boolean functions $\boldsymbol{f}: D \to \{0,1\}$

- $L_{\mathbf{1}}$-testing is equivalent to Hamming testing

$$C_{\mathbf{1}}(\boldsymbol{P},\boldsymbol{\varepsilon}) = C_{\mathbf{0}}(\boldsymbol{P},\varepsilon)$$

- $L_{\boldsymbol{p}}$-testing for $\boldsymbol{p} > 1$ is equivalent to $L_{\mathbf{1}}$-testing
with appropriate distance parameter

$$C_{\boldsymbol{p}}(\boldsymbol{P},\boldsymbol{\varepsilon}) = C_{\mathbf{1}}(\boldsymbol{P},\boldsymbol{\varepsilon}^{\boldsymbol{p}})$$

Relationships: Tolerant L_p-Testing Models

$C_p(P, \varepsilon_1, \varepsilon_2)$ = complexity of tolerant L_p-testing property P with distance parameters $\varepsilon_1, \varepsilon_2$

- E.g., query or time complexity
- for general or restricted (e.g., nonadaptive) tests

For all properties P

- No obvious relationship between tolerant L_1-testing and tolerant Hamming testing
- L_p-testing for $p > 1$ is close in complexity to L_1-testing

$$C_1(P, \varepsilon_1^p, \varepsilon_2) \leq C_p(P, \varepsilon_1, \varepsilon_2) \leq C_1(P, \varepsilon_1, \varepsilon_2^p)$$

Relationships: Tolerant L_p-Testing Models

$\boldsymbol{C_p(P,\varepsilon_1,\varepsilon_2)}$ = complexity of tolerant $\boldsymbol{L_p}$-testing property $\boldsymbol{P}$ with distance parameters $\varepsilon_1, \varepsilon_2$

- E.g., query or time complexity
- for general or restricted (e.g., nonadaptive) tests

For properties of Boolean functions $\boldsymbol{f}: D \to \{0,1\}$

- $L_{\boldsymbol{1}}$-testing is equivalent to Hamming testing

$$\boldsymbol{C_1(P,\varepsilon_1,\varepsilon_2) = C_0(P,\varepsilon_1,\varepsilon_2)}$$

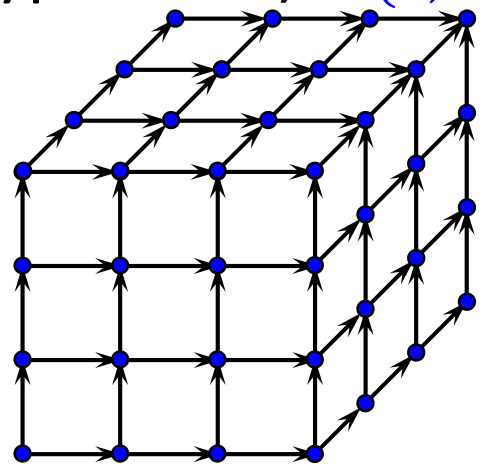
- $L_{\boldsymbol{p}}$-testing for $\boldsymbol{p} > 1$ is equivalent to $L_{\boldsymbol{1}}$-testing with appropriate distance parameters

$$\boldsymbol{C_p(P,\varepsilon_1,\varepsilon_2) = C_1(P,\varepsilon_1^p,\varepsilon_2^p)}$$

Our Results

Property: Monotonicity

Monotonicity



- Domain D=$[n]^d$ (vertices of d-dim hypercube)
- A function $f: D \to \mathbb{R}$ is monotone if increasing a coordinate of x does not decrease $f(x)$.
- Special case $d = 1$

$f: [n] \to \mathbb{R}$ is monotone $\Leftrightarrow f(1), \ldots f(n)$ is sorted.

One of the most studied properties in property testing

[Ergün Kannan Kumar Rubinfeld Viswanathan , Goldreich Goldwasser Lehman Ron, Dodis Goldreich Lehman R Ron Samorodnitsky, Batu Rubinfeld White, Fischer Lehman Newman R Rubinfeld Samorodnitsky, Fischer, Halevy Kushilevitz, Bhattacharyya Grigorescu Jung R Woodruff, ..., Chakrabarty Seshadhri, Blais R Yaroslavtsev, Chakrabarty Dixit Jha Seshadhri]

Monotonicity Testers: Running Time

f	L_0	L_p
$[n] \to [0,1]$	$\Theta\left(\frac{\log n}{\boldsymbol{\varepsilon}}\right)$ [Ergün Kannan Kumar Rubinfeld Viswanathan 00, Fischer 04]	$\Theta\left(\frac{1}{\boldsymbol{\varepsilon}^p}\right)$
$[n]^d \to [0,1]$	$\Theta\left(\frac{d \cdot \log n}{\boldsymbol{\varepsilon}}\right)$ [Chakrabarty Seshadhri 13]	$O\left(\frac{d}{\boldsymbol{\varepsilon}^p} \log \frac{d}{\boldsymbol{\varepsilon}^p}\right)$ $\Omega\left(\frac{1}{\boldsymbol{\varepsilon}^p} \log \frac{1}{\boldsymbol{\varepsilon}^p}\right)$ for $d = 2$ nonadaptive 1-sided error

Monotonicity Testers: Running Time

f	L_0	L_p
$[n] \to \{0,1\}$	$\Theta\left(\frac{1}{\varepsilon}\right)$	$\Theta\left(\frac{1}{\varepsilon^p}\right)$
$[n]^d \to \{0,1\}$	$\Theta\left(\frac{d}{\varepsilon} \cdot \log^3 \frac{d}{\varepsilon}\right)$ [Dodis Goldreich Lehman R Samorodnitsky 99]	$O\left(\frac{d}{\varepsilon^p} \log \frac{d}{\varepsilon^p}\right)$ $\Omega\left(\frac{1}{\varepsilon^p} \log \frac{1}{\varepsilon^p}\right)$ for $d = 2$ nonadaptive 1-sided error
		$\Theta\left(\frac{1}{\varepsilon^p}\right)$ for constant d adaptive 1-sided error

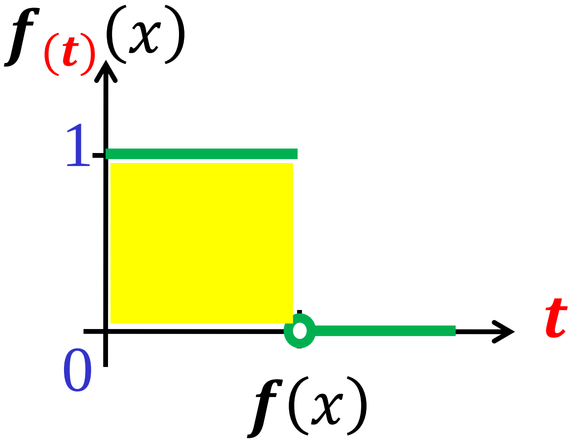
L_1-Testing of Monotonicity

Monotonicity: Reduction to Boolean Functions

Given $\boldsymbol{f}: D \to [0,1]$, a Boolean threshold function $\boldsymbol{f}_{(\boldsymbol{t})}: D \to \{0,1\}$

$$\boldsymbol{f}_{(\boldsymbol{t})}(x) = \begin{cases} 1 & \text{if } \boldsymbol{f}(x) \geq t \\ 0 & \text{otherwise} \end{cases}$$

- Decomposition: $f(x) = \int_0^1 \boldsymbol{f}_{(\boldsymbol{t})}(x)\, d\boldsymbol{t}$



- M = class of monotone functions

Characterization Theorem

$$L_1(\boldsymbol{f}, M) = \int_0^1 L_1\left(\boldsymbol{f}_{(\boldsymbol{t})}, M\right) d\boldsymbol{t}$$

Characterization Theorem: One Direction

$$\boxed{L_1(\boldsymbol{f}, M) \le \int_0^1 L_1\left(\boldsymbol{f}_{(\boldsymbol{t})}, M\right) d\boldsymbol{t}}$$

- $\forall t \in [0,1]$, let g_t=closest monotone (Boolean) function to $\boldsymbol{f}_{(\boldsymbol{t})}$.
- Let $\boldsymbol{g} = \int_0^1 g_t d\boldsymbol{t}$. Then $\boldsymbol{g}$ is monotone, since g_t are monotone.

$$
\begin{aligned}
L_1(\boldsymbol{f}, M) &\le \|f - g\|_1 && \text{Because } \boldsymbol{g} \text{ is monotone} \\
&= \left\| \int_0^1 \boldsymbol{f}_{(\boldsymbol{t})} d\boldsymbol{t} - \int_0^1 g_t d\boldsymbol{t} \right\|_1 && \text{Decomposition \& definition of } \boldsymbol{g} \\
&= \left\| \int_0^1 (\boldsymbol{f}_{(\boldsymbol{t})} - g_t) d\boldsymbol{t} \right\|_1 \\
&\le \int_0^1 \left\| \boldsymbol{f}_{(\boldsymbol{t})} - g_t \right\|_1 d\boldsymbol{t} && \text{Triangle inequality} \\
&= \int_0^1 L_1\left(\boldsymbol{f}_{(\boldsymbol{t})}, M\right) d\boldsymbol{t} && \text{Definition of } g_t
\end{aligned}
$$

Monotonicity: Using Characterization Theorem

Characterization Theorem

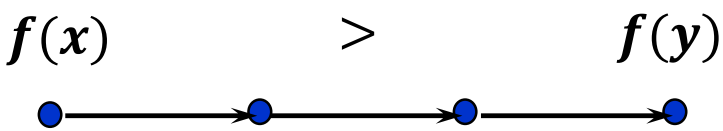
$$d_1(\boldsymbol{f}, M) = \int_0^1 d_1\left(\boldsymbol{f}_{(\boldsymbol{t})}, M\right) d\boldsymbol{t}$$

We use Characterization Theorem
to get monotonicity L_1-testers
and tolerant testers from
standard property testers for Boolean functions.

L_1-Testers from Testers for Boolean Ranges

A nonadaptive, 1-sided error L_0-test for monotonicity of $f: D \to \{0,1\}$ is also an L_1-test for monotonicity of $f: D \to [0,1]$.

Proof:



- A violation (x, y):
- A nonadaptive, 1-sided error test queries a random set $Q \subseteq D$ and rejects iff Q contains a violation.
- If $f: D \to [0,1]$ is monotone, Q will not contain a violation.
- If $d_1(f, M) \geq \varepsilon$ then $\exists \boldsymbol{t}^*: d_0\left(\boldsymbol{f}_{(\boldsymbol{t}^*)}, M\right) \geq \boldsymbol{\varepsilon}$
- W.p. $\geq 2/3$, set Q contains a violation (x, y) for $\boldsymbol{f}_{(\boldsymbol{t}^*)}$

$$\boldsymbol{f}_{(\boldsymbol{t}^*)}(x) = 1, \boldsymbol{f}_{(\boldsymbol{t}^*)}(y) = 0$$

$$\Downarrow$$

$$\boldsymbol{f}(x) > \boldsymbol{f}(y)$$

Monotonicity Testers: Running Time

f	L_0	L_p
$[n] \to [0,1]$	$\Theta\left(\frac{\log n}{\boldsymbol{\varepsilon}}\right)$ [Ergün Kannan Kumar Rubinfeld Viswanathan 00, Fischer 04]	$\Theta\left(\frac{1}{\boldsymbol{\varepsilon}^p}\right)$
$[n]^d \to [0,1]$	$\Theta\left(\frac{d \cdot \log n}{\boldsymbol{\varepsilon}}\right)$ [Chakrabarty Seshadhri 13]	$O\left(\frac{d}{\boldsymbol{\varepsilon}^p} \log \frac{d}{\boldsymbol{\varepsilon}^p}\right)$ $\Omega\left(\frac{1}{\boldsymbol{\varepsilon}^p} \log \frac{1}{\boldsymbol{\varepsilon}^p}\right)$ for $d = 2$ nonadaptive 1-sided error

Distance Approximation and Tolerant Testing

Approximating L_1-distance to monotonicity $\pm\delta\ w.p. \geq 2/3$

f	L_0	L_1
$[n] \to [0,1]$	polylog $n \cdot \left(\frac{1}{\delta}\right)^{O(1/\delta)}$ [Saks Seshadhri 10]	$\Theta\left(\frac{1}{\delta^2}\right)$

- Time complexity of tolerant L_1-testing for monotonicity is

$$O\left(\frac{\varepsilon_2}{(\varepsilon_2 - \varepsilon_1)^2}\right).$$

L_1-Testers for Other Properties

Via combinatorial characterization of L_1-distance to the property

- Lipschitz property $\boldsymbol{f}: [\boldsymbol{n}]^{\boldsymbol{d}} \to [0,1]$:

$$\Theta\left(\frac{\boldsymbol{d}}{\epsilon}\right) \text{ (tight)}$$

Via (implicit) proper learning: approximate in L_1 up to error $\boldsymbol{\epsilon}$, test approximation on a random $O(1/\epsilon)$-sample

- Convexity $\boldsymbol{f}: [\boldsymbol{n}]^{\boldsymbol{d}} \to [0,1]$:

$$\mathrm{O}\left(\boldsymbol{\epsilon}^{-\frac{\boldsymbol{d}}{2}} + \frac{1}{\boldsymbol{\epsilon}}\right) \text{ (tight for } \boldsymbol{d} \leq 2\text{)}$$

- Submodularity $\boldsymbol{f}: \{0,1\}^{\boldsymbol{d}} \to [0,1]$

$$2^{\tilde{O}\left(\frac{1}{\epsilon}\right)} + poly\left(\frac{1}{\boldsymbol{\epsilon}}\right) \log \boldsymbol{d}$$ [Feldman Vondrak 13]

Open Problems

- Our L_1-tester for monotonicity is nonadaptive, but we show that adaptivity helps for Boolean range.

Is there a better adaptive tester?

- All our algorithms for L_p-testing for $p \geq 1$ were obtained directly from L_1-testers.

Can one design better algorithms by working directly with L_p-distances?

- We designed tolerant tester only for monotonicity (d=1,2).

Tolerant testers for higher dimensions?

Other properties?