

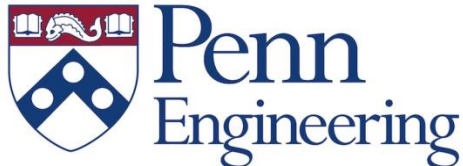
# $L_p$ -Testing

With P. Berman and S. Raskhodnikova (STOC'14+).

**Grigory Yaroslavtsev**

Warren Center for Network and Data Sciences

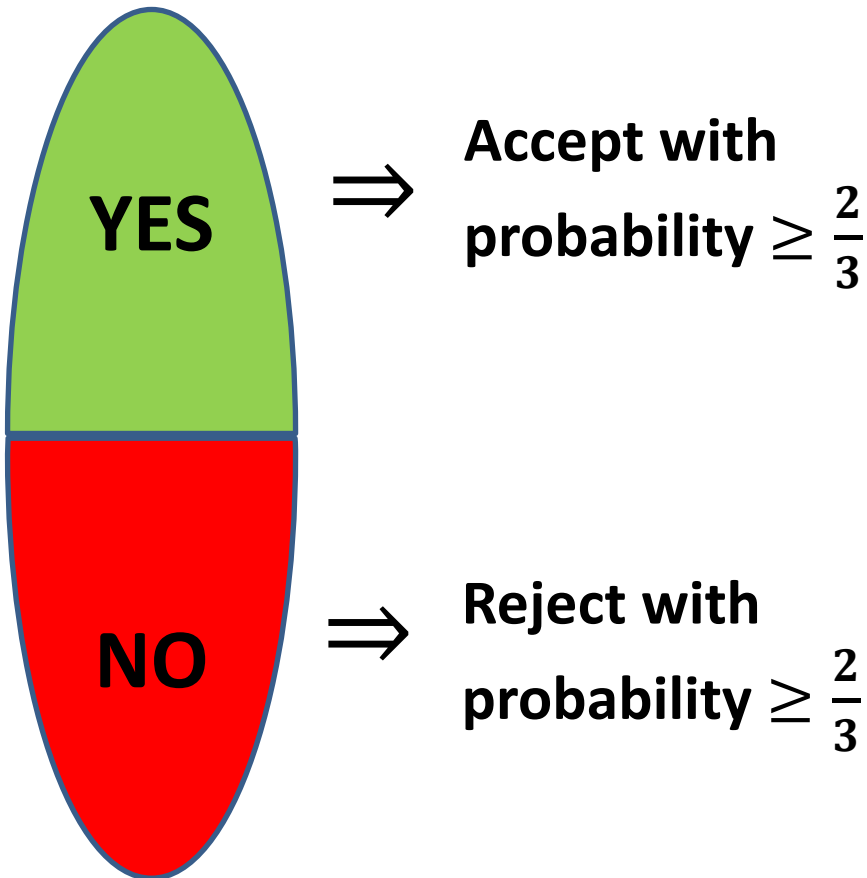
<http://grigory.us>



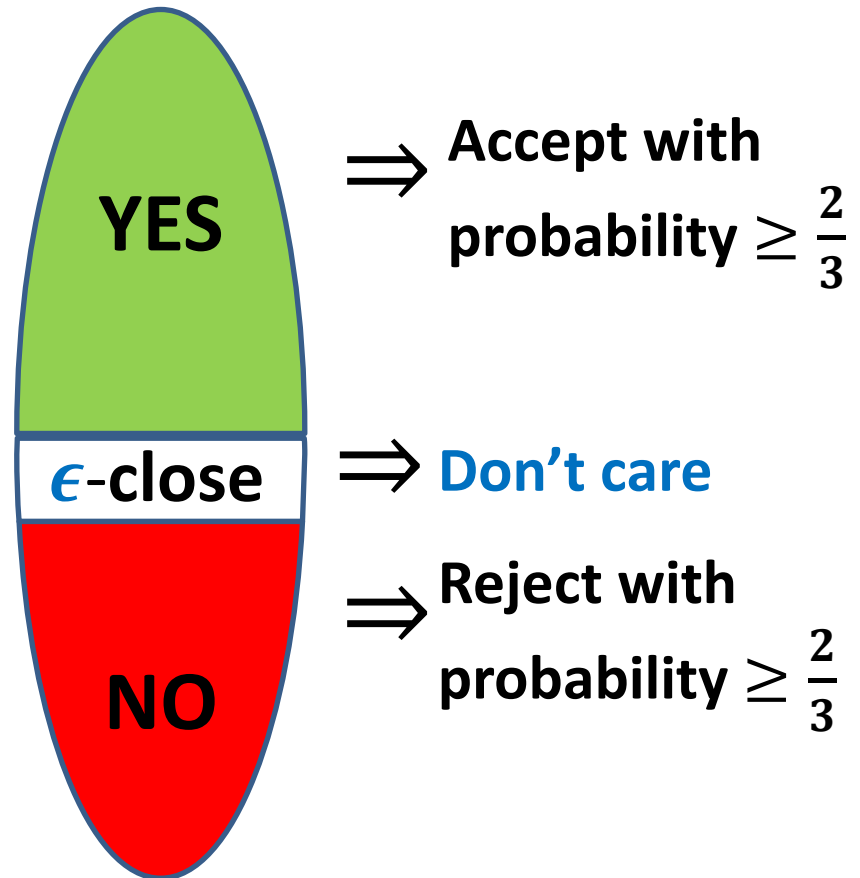
# Property Testing

[Goldreich, Goldwasser, Ron; Rubinfeld, Sudan]

## Randomized Algorithm



## Property Tester



$\epsilon$ -close :  $\leq \epsilon$  fraction has to be changed to become YES

# Which stocks were growing?



**Microsoft**



**IBM**



Data from <http://finance.google.com>

# Property testing: testing monotonicity?



**Microsoft**



**IBM**

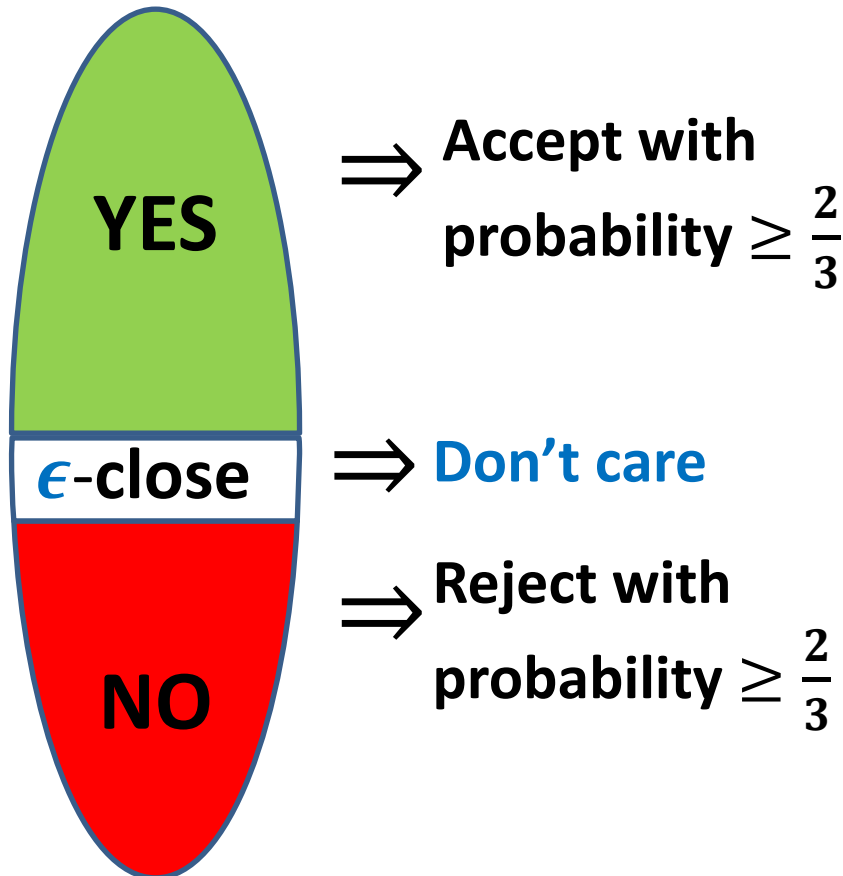


Data from <http://finance.google.com>

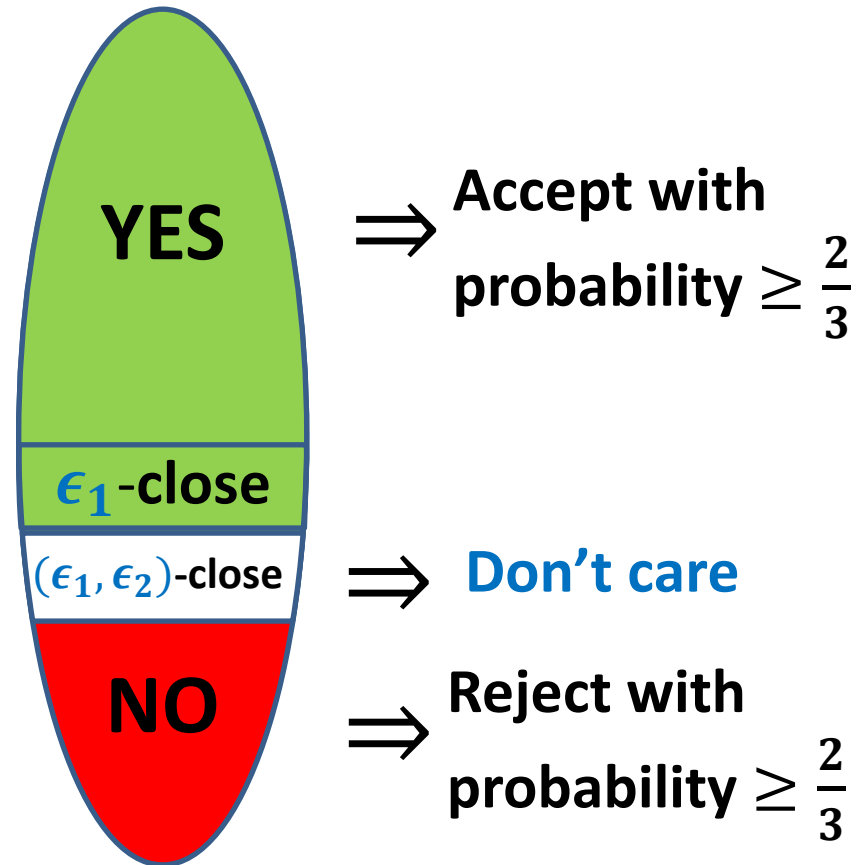
# Tolerant Property Testing

[Parnas, Ron, Rubinfeld]

## Property Tester



## Tolerant Property Tester



$\epsilon$ -close :  $\leq \epsilon$  fraction has to be changed to become YES

# Tolerant monotonicity testing?



**Microsoft**



**IBM**

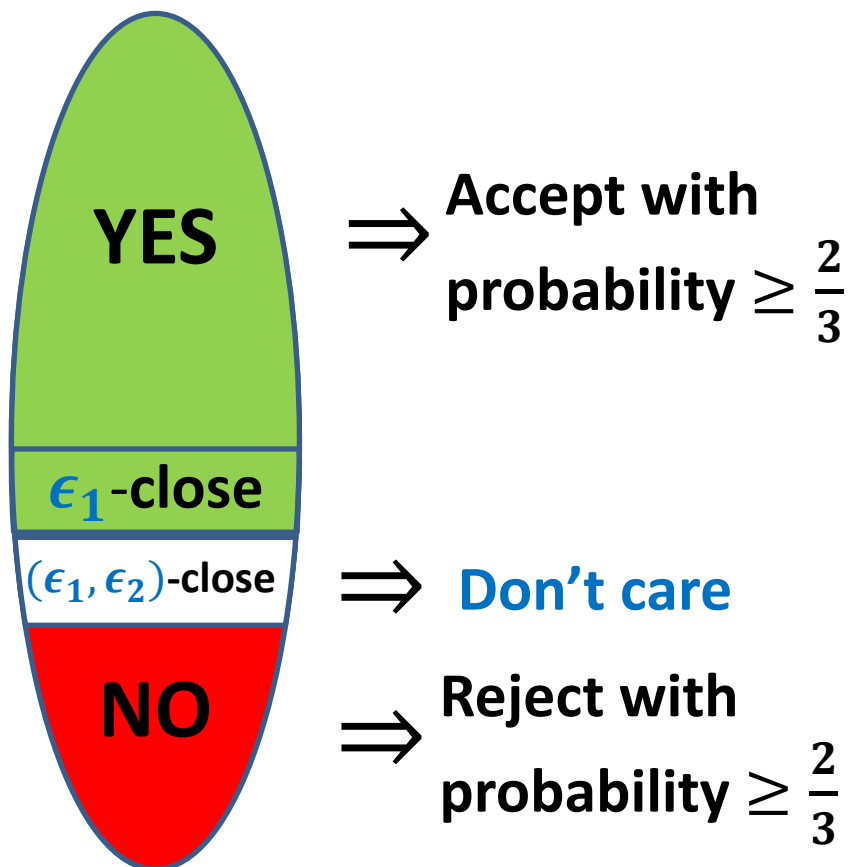


Data from <http://finance.google.com>

# Tolerant “ $L_1$ Property Testing”

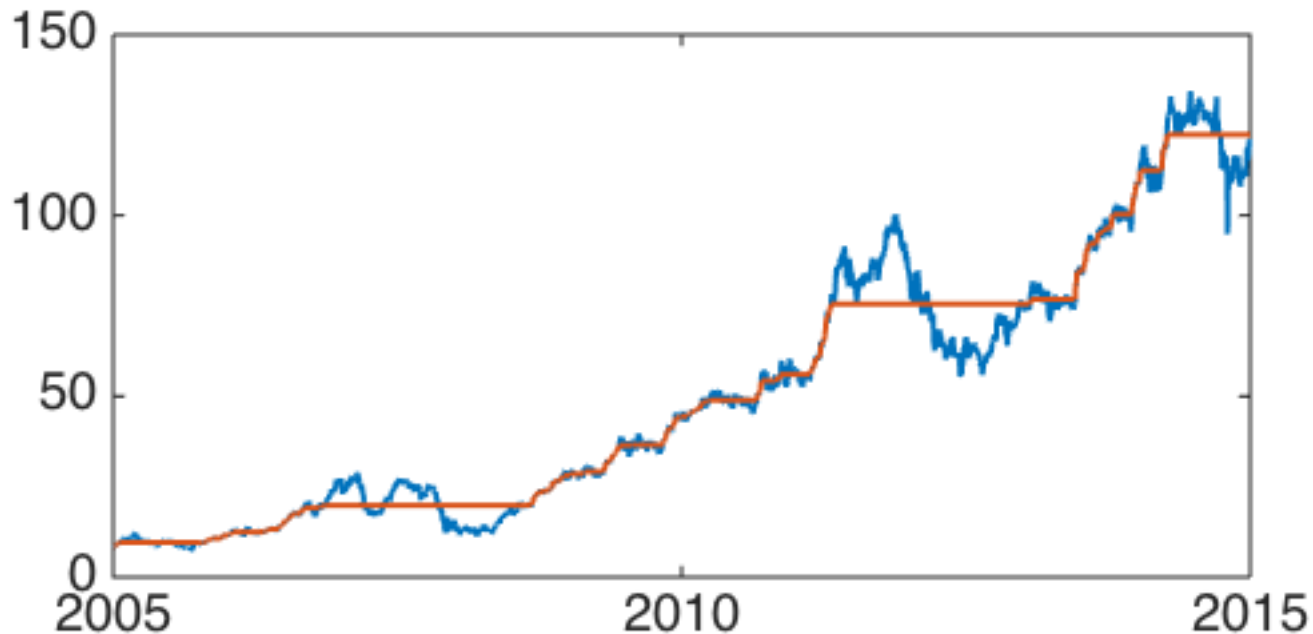
- $f: \{1, \dots, n\} \rightarrow [0,1]$
- $\mathcal{P}$  = class of monotone functions
- $dist_1(f, \mathcal{P}) = \frac{\min_{g \in \mathcal{P}} |f - g|_1}{n}$
- $\epsilon$ -close:  $dist_1(f, \mathcal{P}) \leq \epsilon$
- More general: distance approximation
- Even more general: isotonic regression

## Tolerant “ $L_1$ Property Tester”



# $L_1$ -Isotonic Regression

- Pool Adjacent Violators Algorithm
- Running time  $O(n \log n)$  [Folklore]
- Available in Matlab/R packages





# New $L_p$ -Testing Model for Real-Valued Data

- **Generalizes** standard Hamming testing
- For  $p > 0$  still has a **probabilistic interpretation**:
$$d_p(f, g) = (\mathbf{E}[|f - g|^p])^{1/p}$$
- Compatible with existing **PAC-style learning models** that have  $L_p$ -error (preprocessing for model selection)
- For Boolean functions,  $d_0(f, g) = d_p(f, g)^p$ .
- Various distances used widely in distribution testing

# Our Contributions

1. Relationships between  $L_p$ -testing models
2. Algorithms
  - $L_p$ -testers for  $p \geq 1$ 
    - monotonicity, Lipschitzness, convexity
  - Tolerant  $L_p$ -tester for  $p \geq 1$ 
    - monotonicity in 1D (sublinear algorithm for isotonic regression)
    - monotonicity in 2D
  - ❖ Our  $L_p$ -testers **beat lower bounds** for Hamming testers
  - ❖ **Simple algorithms** backed up by involved analysis
  - ❖ Uniformly sampled (or **easy to sample**) data suffices
3. Nearly tight lower bounds in many cases

# Implications for Hamming Testing

Some techniques/results carry over to Hamming testing

- Improvement on **Levin's work investment strategy**
  - Connectivity of bounded-degree graphs [[Goldreich, Ron '02](#)]
  - Properties of images [[Raskhodnikova '03](#)]
  - Multiple-input problems [[Goldreich '13](#)]
- First example of **monotonicity testing** problem where **adaptivity helps**
- Improvements to Hamming testers for Boolean functions

# Definitions

- $f: D \rightarrow [0,1]$  ( $D =$  finite domain/poset)
- $\|f\|_p = (\sum_{x \in D} |f(x)|^p)^{1/p}$ , for  $p \geq 1$
- $\|f\|_0 =$  Hamming weight (# of non-zero values)
- Property  $P =$  class of functions (monotone, convex, Lipschitz, ...)
- $dist_p(f, P) = \frac{\min_{g \in P} \|f - g\|_p}{\|1\|_p}$

# Relationships: $L_p$ -Testing

$Q_p(\mathbf{P}, \epsilon)$  = query complexity of  $L_p$ -testing property  $\mathbf{P}$  at distance  $\epsilon$

- $Q_1(\mathbf{P}, \epsilon) \leq Q_0(\mathbf{P}, \epsilon)$
- $Q_1(\mathbf{P}, \epsilon) \leq Q_2(\mathbf{P}, \epsilon)$  (Cauchy-Schwarz)
- $Q_1(\mathbf{P}, \epsilon) \geq Q_2(\mathbf{P}, \sqrt{\epsilon})$

Boolean functions  $f: D \rightarrow \{0,1\}$

$$Q_0(\mathbf{P}, \epsilon) = Q_1(\mathbf{P}, \epsilon) = Q_2(\mathbf{P}, \sqrt{\epsilon})$$

# Relationships: Tolerant $L_p$ -Testing

$Q_p(\mathbf{P}, \epsilon_1, \epsilon_2)$  = query complexity of tolerant  $L_p$ -testing property  $\mathbf{P}$  with distance parameters  $\epsilon_1, \epsilon_2$

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

$$Q_1(\mathbf{P}, \epsilon_1^p, \epsilon_2) \leq Q_p(\mathbf{P}, \epsilon_1, \epsilon_2) \leq Q_1(\mathbf{P}, \epsilon_1, \epsilon_2^p)$$

For Boolean functions  $f: D \rightarrow \{0,1\}$

$$Q_0(\mathbf{P}, \epsilon_1, \epsilon_2) = Q_1(\mathbf{P}, \epsilon_1, \epsilon_2) = Q_p(\mathbf{P}, \epsilon_1^{1/p}, \epsilon_2^{1/p})$$

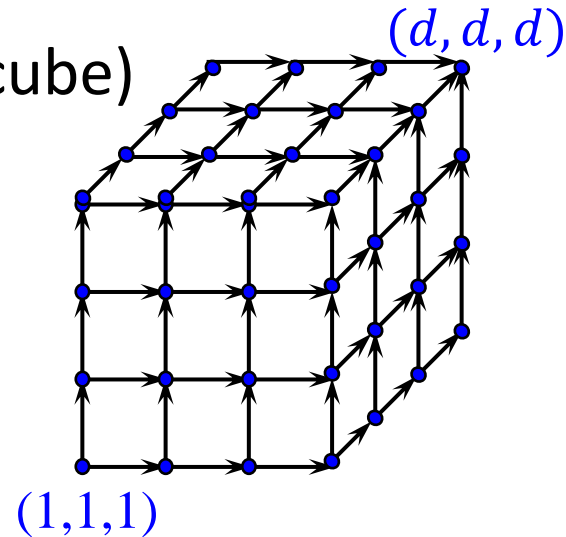
# Testing Monotonicity

- Line ( $D = [n]$ )

	$L_0$	$L_1$
Upper bound	$O(\log n/\epsilon)$ [Ergun, Kannan, Kumar, Rubinfeld, Viswanathan'00]	$O(1/\epsilon)$
Lower bound	$\Omega(\log n/\epsilon)$ [Fischer'04]	$\Omega(1/\epsilon)$

# Monotonicity

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



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

One of the most studied properties in property testing [Ergün  
Kannan Kumar Rubinfeld Viswanathan, Goldreich Goldwasser Lehman Ron, Dodis Goldreich Lehman  
Raskhodnikova Ron Samorodnitsky, Batu Rubinfeld White, Fischer Lehman Newman Raskhodnikova  
Rubinfeld Samorodnitsky, Fischer, Halevy Kushilevitz, Bhattacharyya Grigorescu Jung Raskhodnikova  
Woodruff, ..., Chakrabarty Seshadhri, Blais, Raskhodnikova Yaroslavtsev, Chakrabarty Dixit Jha Seshadhri, ...]



# Monotonicity: Key Lemma

- $M$  = class of monotone functions
- Boolean slicing operator  $f_{\mathbf{y}}: D \rightarrow \{0,1\}$

$$f_{\mathbf{y}}(x) = 1, \text{ if } f(x) \geq \mathbf{y},$$

$$f_{\mathbf{y}}(x) = 0, \text{ otherwise.}$$

- **Theorem:**

$$\text{dist}_1(f, M) = \int_0^1 \text{dist}_0(f_{\mathbf{y}}, M) d\mathbf{y}$$

# Proof sketch: slice and conquer

1) Closest monotone function with **minimal  $L_1$ -norm** is **unique** (can be denoted as an operator  $M_f^1$ ).

2)  $\|f - g\|_1 = \int_0^1 \|f_{\mathbf{y}} - g_{\mathbf{y}}\|_1 d\mathbf{y}$

3)  $M_f^1$  and  $f_{\mathbf{y}}$  commute:  $(M_f^1)_{\mathbf{y}} = M^1_{(f_{\mathbf{y}})}$

$$\begin{aligned} \text{dist}_1(f, M) &= \frac{\|f - M_f^1\|_1}{|D|} \stackrel{2)}{=} \frac{\int_0^1 \|f_{\mathbf{y}} - (M_f^1)_{\mathbf{y}}\|_1 d\mathbf{y}}{|D|} \stackrel{3)}{=} \\ &= \frac{\int_0^1 \|f_{\mathbf{y}} - M^1_{(f_{\mathbf{y}})}\|_1 d\mathbf{y}}{|D|} = \int_0^1 \text{dist}_0(f_{\mathbf{y}}, M) d\mathbf{y} \end{aligned}$$

# $L_1$ -Testers from Boolean Testers

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

Proof:

$$f(x) > f(y)$$



- 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  $f: D \rightarrow [0,1]$  is monotone,  $Q$  will not contain a violation.
- If  $d_1(f, M) \geq \varepsilon$  then  $\exists t^*: d_0(f_{(t^*)}, M) \geq \varepsilon$
- W.p.  $\geq 2/3$ , set  $Q$  contains a violation  $(x, y)$  for  $f_{(t^*)}$

$$f_{(t^*)}(x) = 1, f_{(t^*)}(y) = 0$$

$\Downarrow$

$$f(x) > f(y)$$

- For Boolean functions  $O(1/\varepsilon)$  sample is enough

# Our Results: Testing Monotonicity

- Hypergrid ( $D = [n]^d$ )

	$L_0$	$L_1$
Upper bound	$O\left(\frac{d \log n}{\epsilon}\right)$ [Dodis et al. '99,..., Chakrabarti, Seshadhri '13]	$O\left(\frac{d}{\epsilon} \log \frac{d}{\epsilon}\right)$
Lower bound	$\Omega\left(\frac{d \log n}{\epsilon}\right)$ [Dodis et al.'99..., Chakrabarti, Seshadhri '13]	$\Omega\left(\frac{1}{\epsilon} \log \frac{1}{\epsilon}\right)$ Non-adaptive 1-sided error

- $2^{O(d)}/\epsilon$  **adaptive** tester for Boolean functions

# Testing Monotonicity of $[n]^d \rightarrow \{0,1\}$

- $e^i = (0 \dots 1 \dots 0) = i$ -th unit vector.
- For  $i \in [d]$ ,  $\alpha \in [n]^d$  where  $\alpha_i = 0$  an axis-parallel line along dimension  $i$ :  $\{\alpha + x_i e^i \mid x_i \in [n]\}$
- $L_{n,d}$  = set of all  $dn^{d-1}$  axis-parallel lines
- Dimension reduction for  $f: [n]^d \rightarrow \{0,1\}$  [Dodis et al.'99]:

$$E_{\ell \sim L_{n,d}} \left[ \text{dist} \left( f|_{\ell}, M \right) \right] \geq \frac{\text{dist}(f, M)}{2d}$$

- If  $\text{dist}(f|_{\ell}, M) \geq \delta \Rightarrow O\left(\frac{1}{\delta}\right)$ -sample detects a violation

# Testing Monotonicity on $[n]^d$

- Dimension reduction for  $f: [n]^d \rightarrow \{0,1\}$  [Dodis et al.'99]:

$$E_{\ell \sim L_{n,d}} \left[ \text{dist} \left( f|_{\ell}, M \right) \right] \geq \frac{\text{dist}(f, M)}{2d}$$

- If  $\text{dist}(f|_{\ell}, M) \geq \delta \Rightarrow O\left(\frac{1}{\delta}\right)$ -sample can detect a violation

- “Inverse Markov”: For r. v.  $X \in [0,1]$  with  $E[X] = \mu$  and  $c < 1$

$$\Pr[X \leq c\mu] \leq \frac{1 - \mu}{1 - c\mu} \Rightarrow \Pr\left[X \leq \frac{\mu}{2}\right] \leq 1 - \frac{\mu}{2 - \mu} \leq 1 - \frac{\mu}{2}$$

- $\Pr\left[\text{dist}(f|_{\ell}, M) \geq \frac{\text{dist}(f, M)}{4d}\right] \geq \frac{\text{dist}(f, M)}{4d} \Rightarrow O\left(\frac{d^2}{\epsilon^2}\right)$ -test

- [Dodis et al.]  $O\left(\frac{d}{\epsilon} \log^2 \frac{d}{\epsilon}\right)$  via “Levin’s economical work investment strategy” (used in other papers for testing connectedness of a graph, properties of images, etc.)

# Testing Monotonicity on $[n]^d$

- “Discretized Inverse Markov”

For r. v.  $X \in [0,1]$  with  $E[X] = \mu \leq \frac{1}{2}$  and  $t = 3 \log 1/\mu$

$$\exists j \in [t]: \Pr[X \geq 2^{-j}] \geq \frac{2^j \mu}{4}$$

- For each  $i \in [t]$  pick  $O\left(\frac{1}{\mu 2^i}\right)$  samples of size  $O(2^i) \Rightarrow$  complexity  $O\left(\frac{1}{\mu} \log \frac{1}{\mu}\right)$
- For the right value  $j$  the test rejects with constant probability
- $\mu = E_{\ell \sim L_{n,d}}[\text{dist}(\mathbf{f}|_{\ell}, M)] \geq \frac{\text{dist}(\mathbf{f}, M)}{2d} \Rightarrow O\left(\frac{d}{\epsilon} \log \frac{d}{\epsilon}\right)$ -test

# Distance Approximation and Tolerant Testing

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

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

- Sublinear algorithm for isotonic regression
- Improves  $\tilde{O}\left(\frac{1}{\delta^2}\right)$  adaptive distance approximation of [Fattal,Ron'10] for Boolean functions
- Time complexity of tolerant  $L_1$ -testing for monotonicity is

$$O\left(\frac{\epsilon_2}{(\epsilon_2 - \epsilon_1)^2}\right)$$

- Better dependence than what follows from distance approximation for  $\epsilon_2 \ll 1$



# Distance Approximation $f: [n] \rightarrow [0,1]$

**Theorem:** with constant probability over the choice of a random sample  $\mathbf{S}$  of size  $O\left(\frac{1}{\delta^2}\right)$ :

$$|dist_1(f|_{\mathbf{S}}, M) - dist_1(f, M)| < \delta$$

- Implies an  $O\left(\frac{1}{(\epsilon_2 - \epsilon_1)^2}\right)$  tolerant tester by setting

$$\delta = \frac{(\epsilon_2 - \epsilon_1)}{3}$$

- $dist_1(\mathbf{f}, M) = \int_0^1 dist_0(\mathbf{f}_{\mathbf{y}}, M) d\mathbf{y}$
- Suffices:  $\forall \mathbf{y}: |dist_0(\mathbf{f}_{\mathbf{y}}|_{\mathbf{S}}, M) - dist_0(\mathbf{f}_{\mathbf{y}}, M)| < \delta$
- Improves previous  $\tilde{O}(1/\delta^2)$  algorithm [Fattal, Ron'10]

# Distance Approximation

For  $f: [n] \rightarrow \{0,1\}$  violation graph  $G_f([n], E)$ :  
edge  $(x_1, x_2)$  if  $x_1 \leq x_2, f(x_1) = 1, f(x_2) = 0$

**MM**(G) = maximum matching

**VC**(G) = minimum vertex cover

- $dist_0(f, M) = \frac{|MM(G_f)|}{|D|} = \frac{|VC(G_f)|}{|D|}$  [Fischer et al.'02]
- $dist_0(f|_S, M) = \frac{|MM(G_{f|_S})|}{|S|} = \frac{|VC(G_{f|_S})|}{|S|}$

$$\text{dist}_0(\mathbf{f}|\mathbf{S}, M) - \text{dist}_0(\mathbf{f}, M) < 0 \left( \frac{1}{\sqrt{|\mathbf{S}|}} \right)$$

Define:  $Y(\mathbf{S}) = \frac{|\text{VC}_{f \cap \mathbf{S}}|}{|\mathbf{S}|}$

- $\text{dist}_0(\mathbf{f}|\mathbf{S}, M) = \frac{|\text{VC}_{f|\mathbf{S}}|}{|\mathbf{S}|} \leq \frac{|\text{VC}_{f \cap \mathbf{S}}|}{|\mathbf{S}|} = Y(\mathbf{S})$

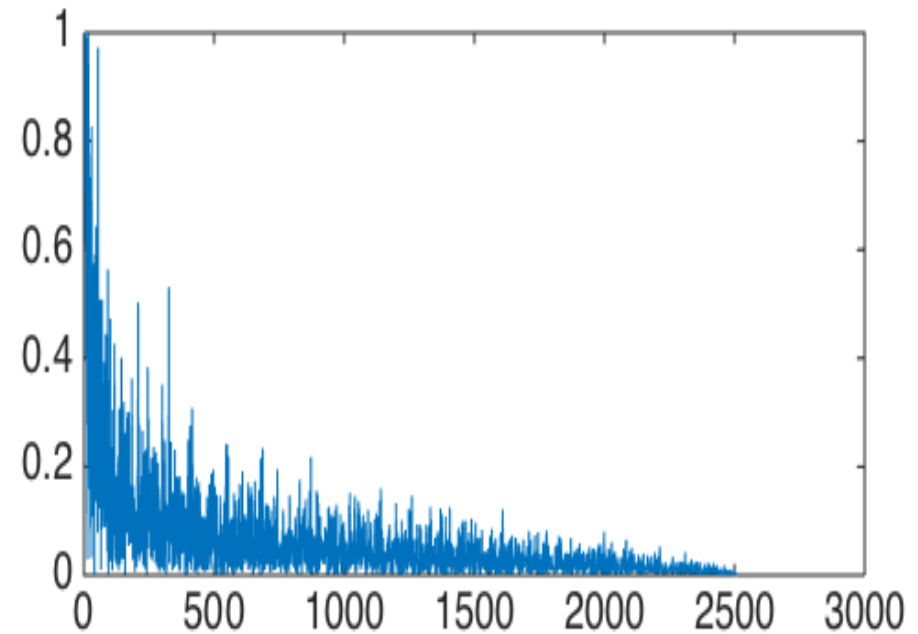
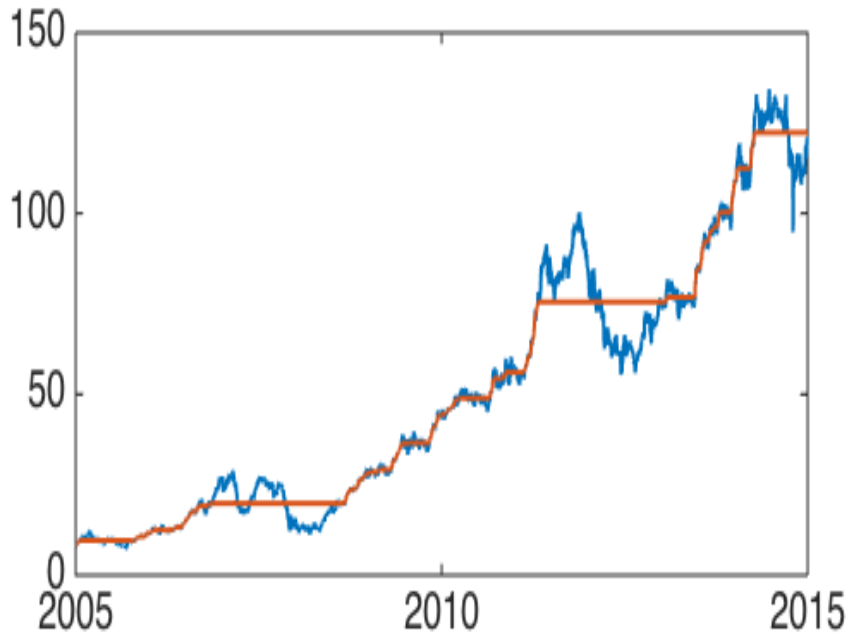
$Y(\mathbf{S})$  has hypergeometric distribution:

- $E[Y(\mathbf{S})] = \frac{|\text{VC}_f|}{|D|} = \text{dist}_0(\mathbf{f}, M)$

- $\text{Var}[Y(\mathbf{S})] \leq \frac{|\mathbf{S}| |\text{VC}_f|}{|D| |\mathbf{S}|^2} = \frac{\text{dist}_0(\mathbf{f}, M)}{|\mathbf{S}|} \leq \frac{1}{|\mathbf{S}|}$

# Experiments

- Data: Apple stock price data (2005-2015) from Google Finance
- Left:  $L_1$ -isotonic regression
- Right: multiplicative error vs. sample size



# $L_1$ -Testers for Other Properties

Via combinatorial characterization of  $L_1$ -distance to the property

- Lipschitz property  $f: [n]^d \rightarrow [0,1]$ :

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

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

- Convexity  $f: [n]^d \rightarrow [0,1]$ :

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

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

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

# $L_p$ -Testing for Convex Optimization

- **Theory:** Convergence rates of gradient descent methods depends on:
  - Convexity / strong convexity constant
  - Lipschitz constant of the derivative
- **Practice:**
  - Q: How to pick learning rate in ML packages?
  - A: Set 0.01 and hope it converges fast
- Even non-tolerant  $L_p$ -testers can be used to sanity check convexity/Lipschitzness



# A lot of open problems!

- $L_p$ -Testing Fourier sparsity [Arnold, Backurs, Blais, Kapralov, Onak, Y.]
- Eric Price: Hey, I can do this better!

**STAND BACK**



**I'M GOING TO TRY  
SCIENCE**



# Open Problems

- Our complexity for  $L_p$ -testing convexity grows exponentially with  $d$

Is there an  $L_p$ -testing algorithm for convexity with subexponential dependence on the dimension?

- Only have tolerant monotonicity for  $d = 1, 2$ .

Tolerant testers for higher dimensions?