Active Testing

Liu Yang

Joint work with Nina Balcan, Eric Blais and Avrim Blum

Property Testing

- Instance space $X = R^n$ (Distri D over X)
- Tested function $f : X \to \{0,1\}$
- A property P of Boolean fn is a subset of all Boolean fns $h : X \to \{-1,1\}$ (e.g ltf)
- $dist_D(f, P) := \min_{g \in P} P_{x \sim D}[f(x) \neq g(x)]$
- Standard Type of query: membership query

Property Testing

If $f \in P$ should accept w/ prob ≥ 2/3

If dist(f,P)>ε should reject w/ prob ≥ 2/3

- E.g. Union of d Intervals

  0----++++----+++++++++-----++---+++--------1

  - $UINT_4$? Yes! $UINT_3$? Depend on ε
  - Model selection: testing can tell us how big d need to be close to target (double and guess, d = 2, 4, 8, 16, ...)

Property Testing and Learning : Motivation

- What is Property Testing for ?
  - Quickly tell if the right fn class to use
  - Estimate complexity of fn without actually learning

- **Want to do it with fewer queries than learning**

Standard Model uses Membership Query

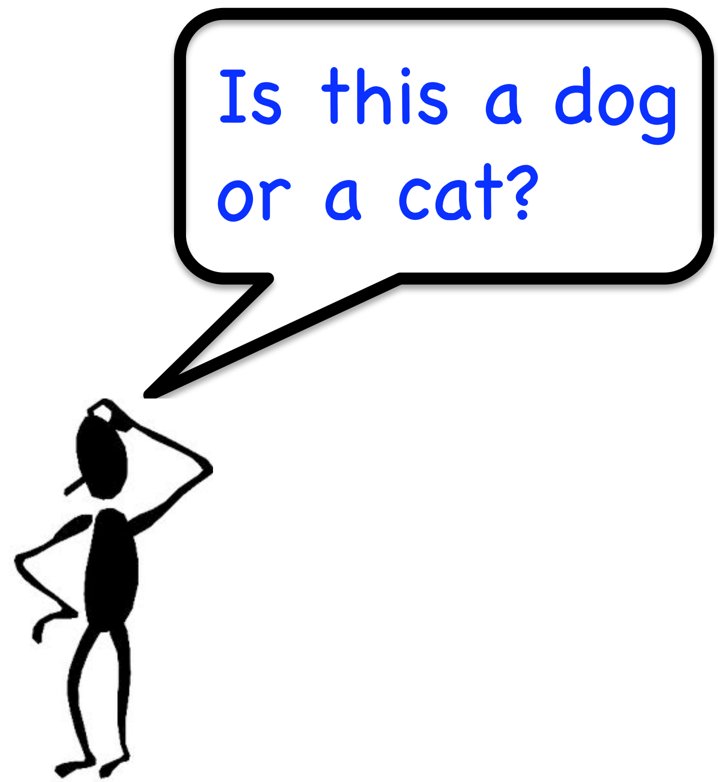
- Results of Testing basic Boolean fns using MQ:
- Constant QC for UINTd, dictator, ltf, …

Class of functions	Number of Queries	Reference
singletons and monomials	$O(1/\epsilon)$	[PRS02]
s-term monotone DNF	$\tilde{O}(s^2/\epsilon)$	[PRS02]
k-Juntas	$\tilde{O}(k^2/\epsilon)$, $\Omega(k)$	[FKR+04],[CG06]
decision lists	$\tilde{O}(1/\epsilon^2)$	[DLM+07]
size-s decision trees, size-s branching programs	$\tilde{O}(s^4/\epsilon^2)$	[DLM+07]
s-term DNF, size-s Boolean formulae	$\Omega(\log s/\log\log s)$	[DLM+07]
s-sparse polynomials over $GF(2)$	$\tilde{O}(s^4/\epsilon^2)$, $\tilde{\Omega}(\sqrt{s})$	[DLM+07]
size-s Boolean circuits	$\tilde{O}(s^6/\epsilon^2)$	[DLM+07]
functions with Fourier degree $\leq d$	$\tilde{O}(2^{6d}/\epsilon^2)$, $\tilde{\Omega}(\sqrt{d})$	[DLM+07]
linear threshold functions	poly$(1/\epsilon)$	[MORS07]

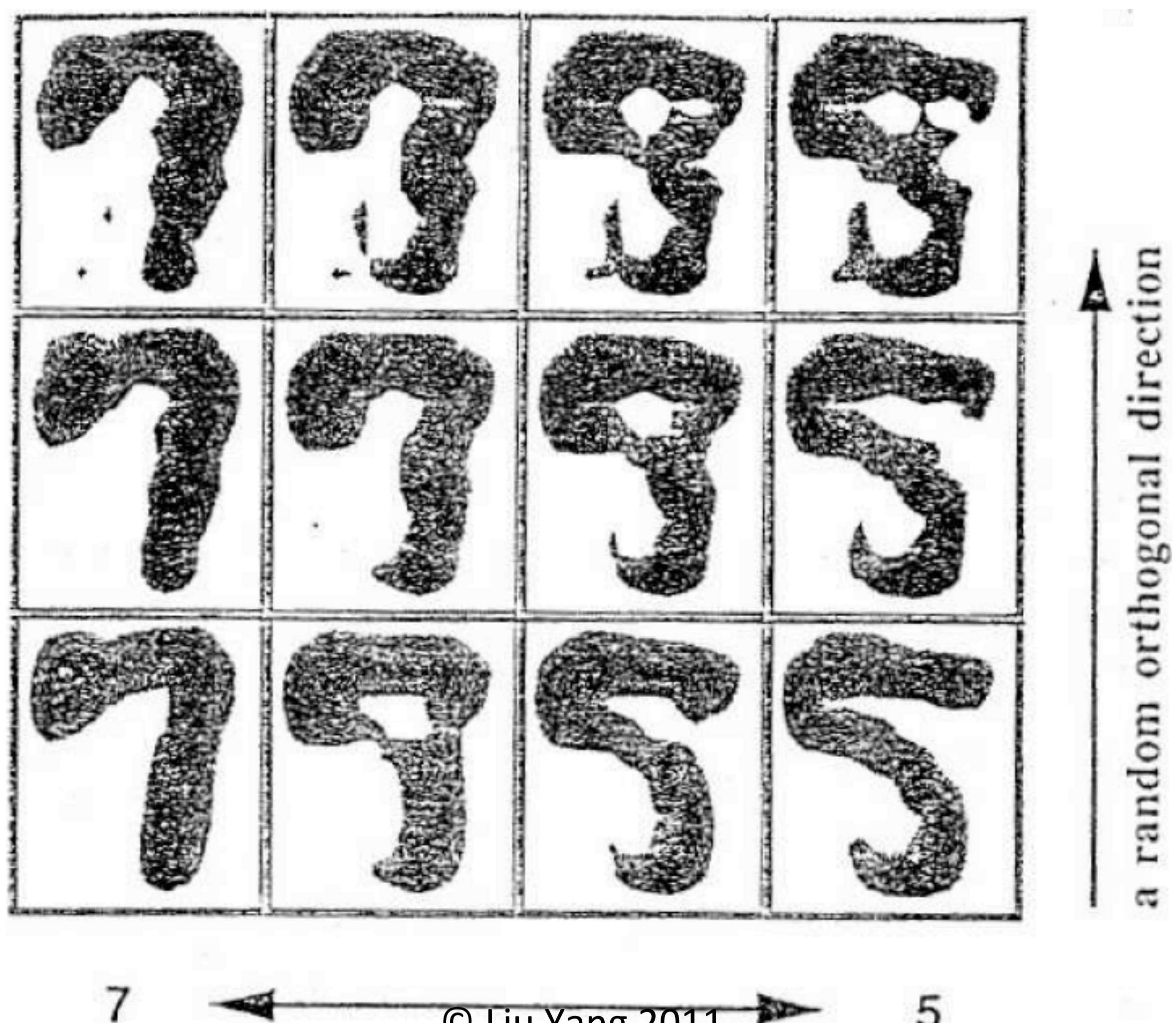
However …

Membership Query is Unrealistic for Machine Learning Problems

Recognizing cat/dog ? MQ gives …



Membership Query is Unrealistic for Machine Learning Problems



Passive : Waste Too Many Queries

- ML people move on

- Passive Model (sample from D)
  query samples exist in **NATURE**; but quite wasteful (many examples uninformative)

- Can we **SAVE** #queries ?

Active Learning

↓

Active Testing

The **NEW!** Model of
Property Testing

Property Tester

- **Definition.** An s-sample, q-query ε-tester for P over the distribution D is a randomized algorithm A that draws **s** samples from D, sequentially queries for the value of f on **q** of those samples, and then
  1. Accepts w.p. at least 2/3 when $f \in P$
  2. Rejects w.p. at least 2/3 when $dist_D(f,P) > \varepsilon$

cheap

expensive

Active Tester

- **Definition.** A randomized algorithm is a q-query active ε-tester for $P \subseteq R^n \to \{0, 1\}$ of over D if it is a poly(n)-sample, q-query ε-tester for P over D.

Active Property Testing

- Testing as preprocessing step of learning
- Need an example? where Active testing
  - get same QC saving as MQ
  - better in QC than Passive
  - need fewer queries than Learning

- **Union of d Intervals, active testing help!**

0----++++----+++++++++-----++---+++--------1

- Testing tells how big d need to be close to target
- #Label: Active Testing need O(1), Passive Testing need $\Theta(\sqrt{d})$, Active Learning need $\Theta(d)$

Outline

- Our Results of Various Classes
- Testing Disjoint Unions of Testable Properties
- General Testing Dimension

Our Results

NEW !!

	Active Testing	Passive Testing	Active Learning
Union of d Intervals	$O(1)$	$\Theta(\sqrt{d})$	$\Theta(d)$
Union of d Thresh	$O(1)$	$\Theta(\sqrt{d})$	const
Dictator	$\Theta(\log n)$	$\Theta(\log n)$	$\Theta(\log n)$
Linear Threshold Fn	$O(\sqrt{n})$	$\tilde{\Theta}(\sqrt{n})$	$\Theta(n)$
Cluster Assumption	$O(1)$	$\Omega(\sqrt{N})$	$\Theta(N)$

✔ Passive-like on testing Dictator

MQ-like

Passive-like

Testing Unions of Intervals

0----++++----+++++++++-----++---+++--------- 1

- **Theorem.** Testing UINTd in the active testing model can be done using $O(1/\varepsilon^3)$ queries. If uniform distribution, we need only $O(\sqrt{d}/\varepsilon^5)$ unlabeled egs.

- Proof Idea:
  - Noise Sensitivity:=Pr[two close pts label diff]
  - all UINTd have low NS whereas all fns far from this class have noticeably larger NS
  - a tester that est noise sensitivity of input fn.

Testing Unions of Intervals (cont.)

- **Definition:** Fix $\delta > 0$. The local δ-noise sensitivity of fn f: [0, 1]->{0, 1} at $x \in [0; 1]$ is $\mathrm{NS}_\delta(f, x) = \Pr_{y \sim_\delta x}[f(x) \neq f(y)]$. The noise sensitivity of f is $\mathrm{NS}_\delta(f) = \Pr_{x, y \sim_\delta x}[f(x) \neq f(y)]$

easy

- **Proposition:** Fix $\delta > 0$. Let f: [0, 1] -> {0,1} be a union of d intervals. $\mathrm{NS}_\delta(f) \leq d\delta$.

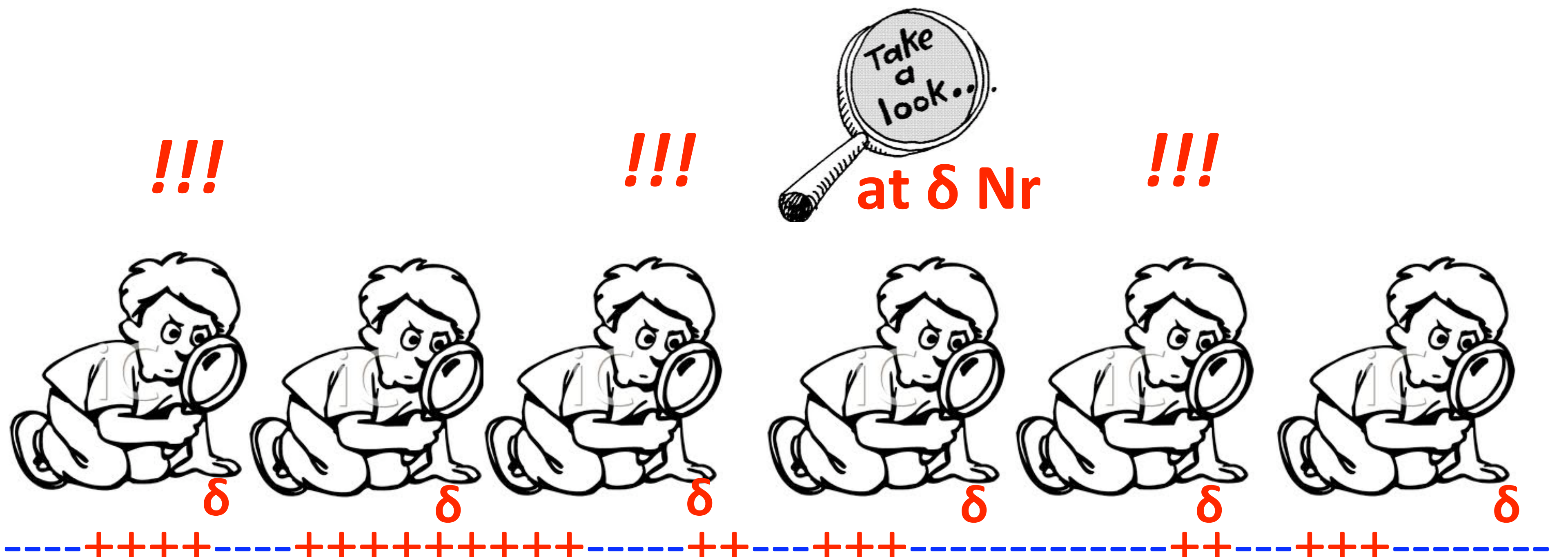
hard

- **Lemma:** Fix $\delta = \varepsilon^2/(32d)$. Let f : [0, 1] -> {0, 1} be a fn with noise sensitivity bounded by $\mathrm{NS}_\delta(f) \leq d\delta(1 + \varepsilon/4)$. Then f is ε-close to a union of d intervals.

Union of Intervals Tester(f, d, ϵ)

Parameters: $\delta = \frac{\epsilon^2}{32d}, r = O(\epsilon^{-3})$.

1. For rounds $i = 1, \ldots, r$,
   - 1.1 Draw $x \in [0, 1]$ uniformly at random.
   - 1.2 Draw samples until we obtain $y \in (x - \delta, x + \delta)$.
   - 1.3 Set $Z_i = \mathbf{1}[f(x) \neq f(y)]$.
2. **Accept** iff $\frac{1}{r} \sum Z_i \leq d\delta(1 + \frac{\epsilon}{8})$.



Testing Unions of Intervals (cont.)

- **Theorem.** Testing UINTd in the active testing model can be done using $O(1/\varepsilon^3)$ queries. If uniform distribution, we need only $O(\sqrt{d}/\varepsilon^5)$ unlabeled egs.

Testing Linear Threshold Fns

- **Theorem.** We can efficiently test LTFs under the Gaussian distribution with $\tilde{O}(\sqrt{n})$ labeled examples in both active and passive testing models. We have lower bounds of $\tilde{\Omega}(n^{1/3})$ for active testing and $\tilde{\Omega}(\sqrt{n})$ on #labels needed for passive testing.

- Learn ltf need Omega(n) under Gaussian

So testing is better than learning in this case.

Testing Linear Threshold Fns (cont.)

- **Definition:** Hermite polynomials : $h_0(x) = 1$, $h_1(x) = x$; $h_2(x) = 1/\sqrt{2}(x^2 - 1)$,…,
  - complete orthogonal basis under <f,g> = $E_x[f(x)g(x)]$, where E_x over std Gaussian distrib
- For any S in N^n, define $H_S = \prod_{i=1}^{n} h_{S_i}(x_i)$
  - Hermite coefficient of f: Rn -> R corresponding to S is $\hat{f}(S) = \langle f, H_S \rangle = \mathbb{E}_x[f(x)H_S(x)]$
  - Hermite decomposition of f: $f(x) = \sum_{S \in \mathbb{N}^n} \hat{f}(S) H_S(x)$

  The degree of the coefficient $\hat{f}(S)$ is $|S| := \sum_{i=1}^{n} S_i$

Testing Linear Threshold Fns (cont.)

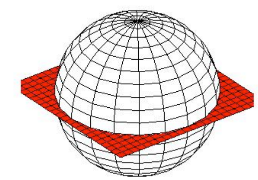
- **Lemma:** There is an explicit continuous fn W: R->R w/bounded derivative $\|W'\|_\infty \leq 1$ and peak value $W(0) = \frac{2}{\pi}$ s.t. every ltf f : Rn -> {1, -1} satisfies $\sum_{i=1}^{n} \hat{f}(e_i)^2 = W(\mathbb{E}_x f)$ Also, every fn g : R^n-> {-1, 1} that satisfies $\left|\sum_{i=1}^{n} \hat{g}(e_i)^2 - W(\mathbb{E}_x g)\right|$ ≤4 ε ; is ε -close to be ltf.

- **Lemma:** For any fn f : R^n -> R, we have $\sum_{i=1}^{n} \hat{f}(e_i)^2 = \mathbb{E}_{x,y}[f(x)f(y)\langle x, y\rangle]$, <x, y> = $\sum_{i=1}^{n} x_i y_i$ is the standard vector dot product.

Testing Linear Threshold Fns (cont.)

The **U-statistic** (of order 2) with symmetric kernel function g : $R^n \times R^n \rightarrow R$ is

$$U_g^m(x^1,\ldots,x^m) := \binom{m}{2}^{-1} \sum_{1\le i<j\le m} g(x^i,x^j)$$



LTF TESTER(f, ϵ)

Parameters: $\tau = \sqrt{4n\log(4n/\epsilon^3)}$, $m = 800\tau/\epsilon^3 + 32/\epsilon^6$.

1. Draw $x^1, x^2, \ldots, x^m$ independently at random from $\mathbb{R}^n$.
2. Query $f(x^1), f(x^2), \ldots, f(x^m)$.
3. Set $\tilde{\mu} = \frac{1}{m}\sum_{i=1}^m f(x^i)$.
4. Set $\tilde{\nu} = \binom{m}{2}^{-1}\sum_{i\neq j} f(x^i)f(x^j)\langle x^i, x^j\rangle \cdot \mathbf{1}[|\langle x^i, x^j\rangle| \le \tau]$.
5. **Accept** iff $|\tilde{\nu} - W(\tilde{\mu})| \le 2\epsilon^3$.

Outline

- Our Results of Various Classes
- **Testing Disjoint Unions of Testable Properties**
- General Testing Dimension

Testing Disjoint Unions of Testable Properties

- Combine a collection of properties P_i via their disjoint union.
- **Theorem.** Given properties P1,…,PN, if each Pi is testable over Di w/ $q(\varepsilon)$ queries & $U(\varepsilon)$ unlabeled samples, then their disjoint union P is testable over the combined distribution D with $O(q(\varepsilon/2)(\log^3 1/\varepsilon))$ queries and $O(U(\varepsilon/2)(N/\varepsilon \log^3 1/\varepsilon))$ unlabeled samples.

Outline

- Our Results of Various Classes
- Testing Disjoint Unions of Testable Properties
- **General Testing Dimension**

Why a General Theory ?

- These are only a few among many possible testing problems

- We don't want to solve each problem one-at-a-time

- It will be good have some general theory: distinguish easily-testable vs hard-to-test problems

General Testing Dim

- Testing dim characterize (up to constant factors) the intrinsic #label requests needed to test the given property w.r.t. the given distribution
- All our lower bounds are proved via testing dim

Minimax Argument

- $\min_{Alg}\max_f$ P(Alg mistaken) = $\max_{\pi_0}\min_{Alg}$ P(Alg mistaken)
- wolg, $\pi_0 = \alpha\, \pi + (1-\alpha)\, \pi'$, $\pi \in \Pi_0, \pi' \in \Pi_\varepsilon$
- Let π_S, π'_S be induced distributions on labels of S. $$d_S(\pi,\pi') = (1/2)\sum_{y\in\{0,1\}^{|S|}} |\pi_S(y) - \pi'_S(y)|$$
- For a given π_0,

$\min_{Alg}$P(Alg makes mistake|S) $\leq$ 1-$dist_S(\pi, \pi')$

Passive Testing Dim

- Define $d_{passive}$ largest q in N, s.t.

$$\sup_{\pi \in \Pi_0} \sup_{\pi' \in \Pi_\epsilon} \Pr_{S \sim D^q}(\mathrm{d}_S(\pi, \pi') > 1/4) \leq 1/4$$

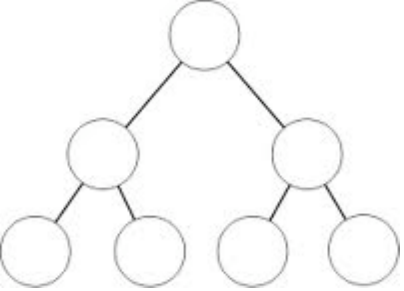
- **Theorem:** Sample Complexity of passive testing is $\Theta(d_{passive})$.

Coarse Active Testing Dim

- Define d_{coarse} as the largest q in N, s.t.

$$\sup_{\pi \in \Pi_0} \sup_{\pi' \in \Pi_\epsilon} \Pr_{S \sim D^q} (\mathrm{d}_S(\pi, \pi') > 1/4) \leq 1/n^q$$

- **Theorem:** If d_{coarse} = O(1) the active testing of P can be done with O(1) queries, and if d_{coarse} = $\omega(1)$ then it cannot.



Active Testing Dim

- Fair(π,π′,U): distri. of labeled (y; l): w.p.½ choose $y \sim \pi_U$, l= 1; w.p.½ choose $y \sim \pi'_U$, l= 0.
- err*(H; P): err of optimal fn in H w.r.t data drawn from distri. P over labeled egs.
- Given u=poly(n) unlabeled egs, $d_{active}(u)$: largest q in N s.t.

$$\sup_{\pi\in\Pi_0}\ \sup_{\pi'\in\Pi_\epsilon}\ \Pr_{U\sim D^u}\left(\mathrm{err}^*(\mathrm{DT}_q, \mathrm{Fair}(\pi,\pi',U)) < 1/4\right) \le 1/4$$

- **Theorem:** Active testing w/ failure prob 1/8 using u unlabeled egs needs $\Omega(d_{active}(u))$ label queries; can be done w/ O(u) unlabeled egs and $O(d_{active}(u))$ label queries

Testing Dim: A Powerful Notion

- We use testing dimension to prove LBs for testing union of intervals and ltfs.

- **Lemma A.** Let $\pi \in \Pi_0$, $\pi' \in \Pi_\varepsilon$. Fix $U \subseteq X$ to be a set of allowable queries. Suppose any $S \subseteq U$, with $|S| = q$, there is a set $E_S \subseteq R^q$ (possibly empty) satisfying $\pi_S(E_S) \leq 2^{-q}/5$ s.t.

$$\pi_S(y) < \frac{6}{5}\pi'_S(y) \text{ for every } y \in \mathbb{R}^q \setminus E_S$$

Then any q-query tester has "large" prob of making mistake.

Application: Dictator fns

- **Theorem:** Active testing of dictatorships under the uniform distribution requires (log n) queries. This holds even for distinguishing dictators from random functions.

- Any class that contains dictator functions requires (log n) queries to test in the active model, including decision trees, functions of low Fourier degree, juntas, DNFs, etc.

Application: Dictator fns (cont.)

- π and π' uniform over dictator fns &over all boolean fns
- S: a set of q vectors in $\{0,1\}^n$: a qxn boolean matrix.
- $c_1(S),\ldots,c_n(S)$: the cols of this matrix

$$\pi_S(y) = \frac{|\{i \in [n] : c_i(S) = y\}|}{n} \quad \text{and} \quad \pi'_S(y) = 2^{-q}$$

- Set S of q vectors chosen unif indep at random from $\{0, 1\}^n$. For any y in $\{0, 1\}^n$, E[# cols of S equal to y] = $n2^{-q}$. Cols are drawn indep. at random, by Chernoff bounds:

$$\Pr\left[\pi_S(y) > \tfrac{6}{5}2^{-q}\right] \le e^{-(\frac{1}{5})^2 n2^{-q}/3} < e^{-\frac{1}{75}n2^{-q}}$$

Now apply Lemma A.

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

Application: LTFs

- **Theorem.** For LTFs under the standard n-dim Gaussian distrib, $d_{passive} = \Omega((n/\log n)^{1/2})$ and active QC = $\Omega((n/\log n)^{1/3})$.

- π: distrib over LTFs obtained by choosing $w \sim N(0, I_{nxn})$ and outputting $f(x) = \mathrm{sgn}(w \bullet x)$.
- π': uniform distrib over all functions.

- Obtain $d_{passive}$:bound tvd(distrib of $Xw/\sqrt{n}$, $N(0, I_{nxn})$).
- Obtain active QC: similar to dictator LB but rely on strong concentration bounds on spectrum of random matrices

Open Problem

- Matching lb/ub for active testing LTF: $\sqrt{n}$
- Tolerant Testing $\varepsilon/2$ vs. ε (UINTd, LTF)

Thanks !