

"INPUT-DRIVEN COMPUTING"

INPUT-DRIVEN COMPUTING



- Probe input
- Streaming
- Sketching
- Priced Computation
- Uncertainty
- On-line learning

$$a_1, \dots, a_n \quad \longrightarrow \quad \frac{1}{n} \sum_{i=1}^n a_i$$

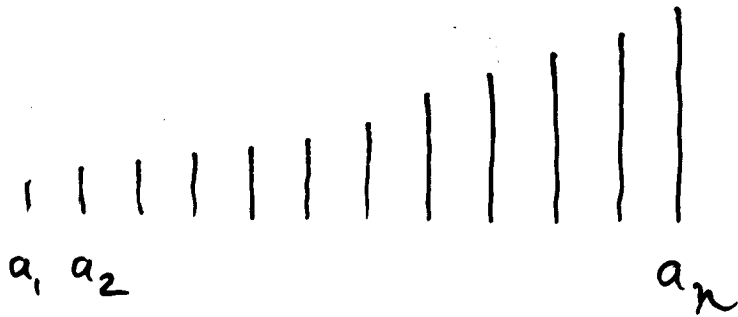
SUBLINEAR ALGO.

1. Pick random i_1, \dots, i_s

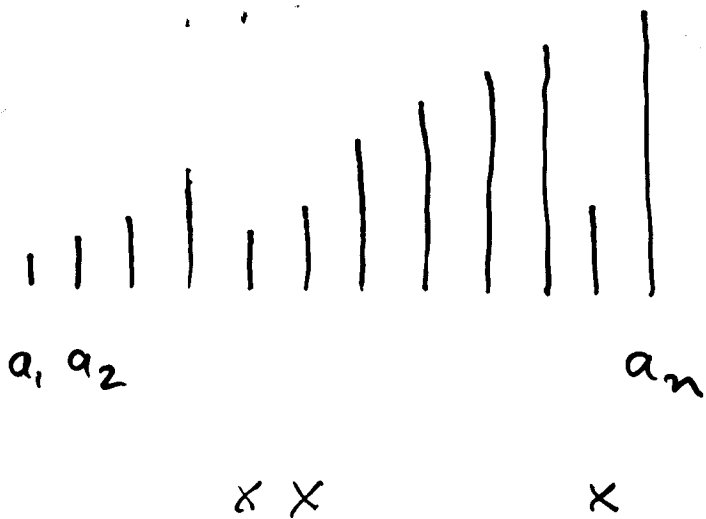
2. Output $\frac{1}{s} \sum_{k=1}^s a_{i_k}$

$|a_i| = O(1)$; choice $s = O(\epsilon^{-2})$

Relative error $< \epsilon$ w/ prob > 0.99



• SORTED



• $\frac{1}{4}$ -Far From SORTED

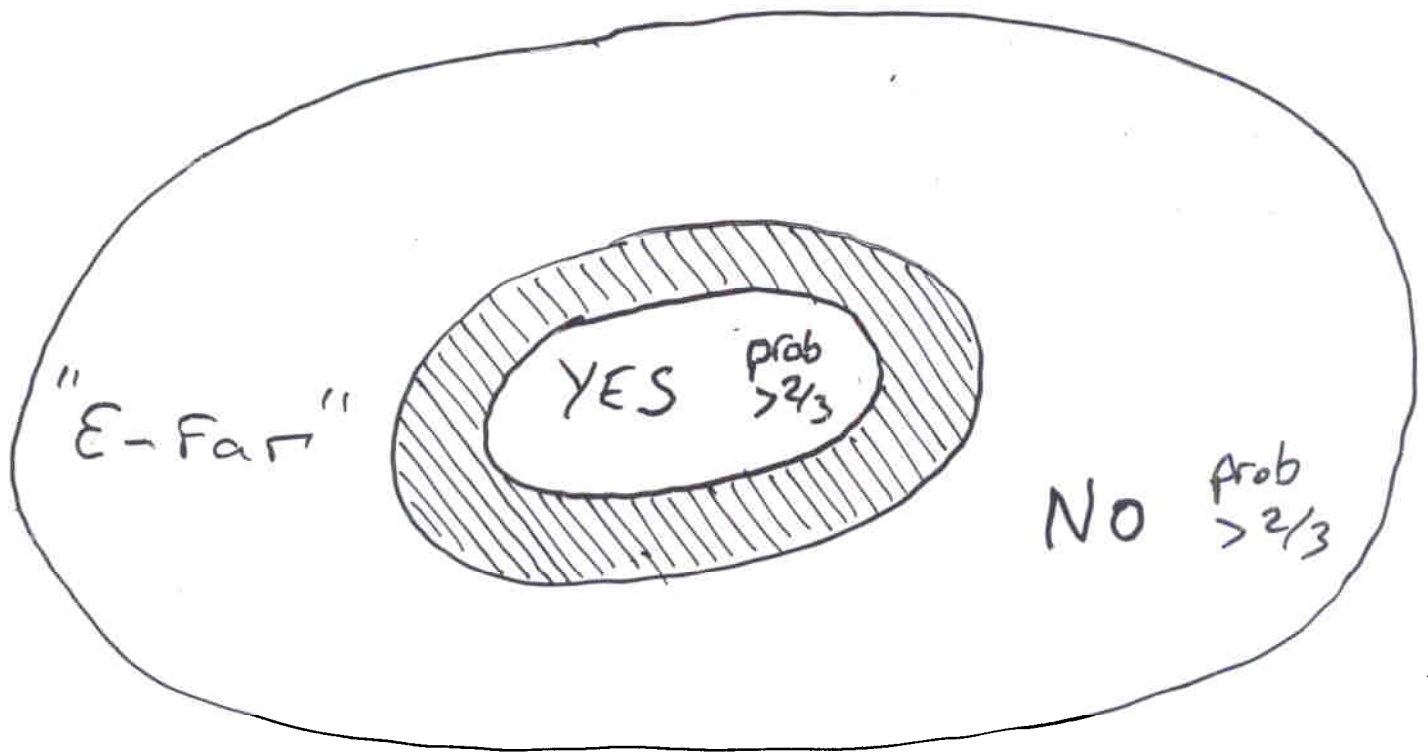
ALGORITHM: Given table, if sorted **YES**
 if ϵ -far from sorted **NO**

[Ergun-Kannan-Kumar-Rubinfeld-Viswanathan]

$O(\epsilon^{-1})$ queries & $O(\epsilon^{-1} \log n)$ time

PROPERTY TESTING

[Rubinfeld - Sudan]



- Need metric in input space
- Generalized to Functions

[Goldreich - Goldwasser - Ron]

PCP THEOREM

[Arora-Lund-Motwani-Safra-Sudan-Stegeedy]

My cool proof
of the Riemann
Hypothesis
⋮

1. Look up $O(1)$ random spots
2. Think
3. IF proof correct ACCEPT
else REJECT w/ prob 0.99

IS STRING A CODEWORD OR FAR FROM ONE?

PROPERTY TESTING CAN BE :

- TRIVIAL (Nonbipartiteness)

- IMPOSSIBLE

Some graph properties in NP
require looking at $\Omega(\text{input})$

[Goldreich-Trevisan]

- SURPRISING

PCP, clique, coloring, etc.

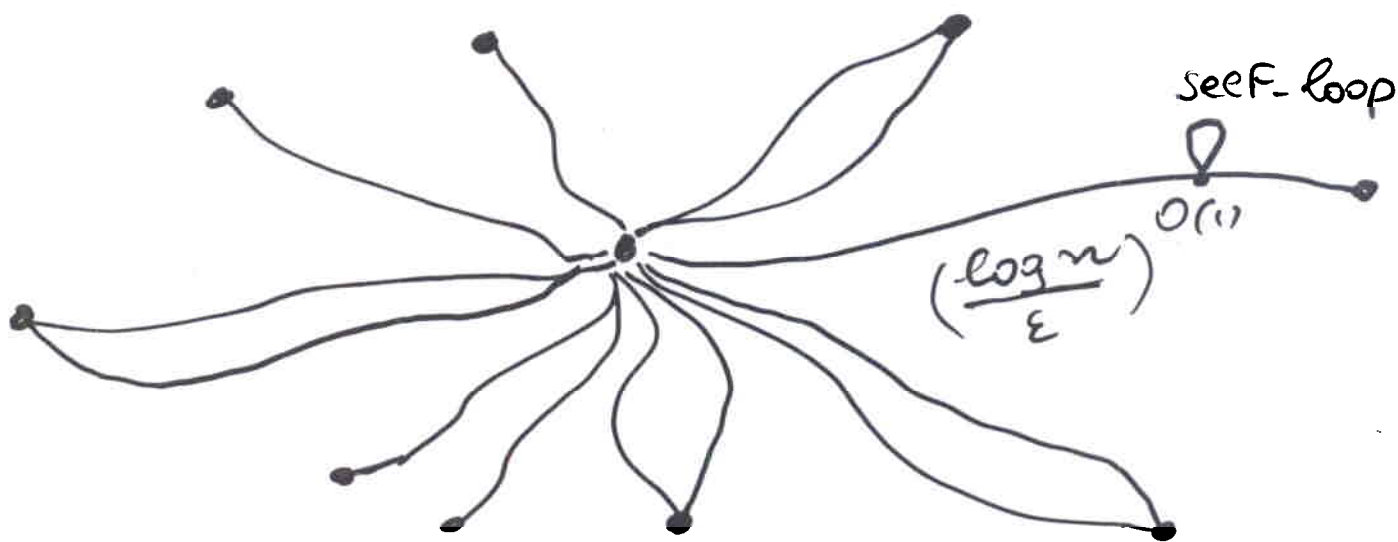
Any general techniques
& principles ?

BIRTHDAY PARADOX

IS GRAPH BIPARTITE ?

[Goldreich-Ron]

Take $\sqrt{n} \left(\frac{\log n}{\epsilon} \right)^{O(1)}$ random walks ; check cycles



if odd cycle REJECT else ACCEPT

IF RAPIDLY MIXING B'DAY PARADOX

ELSE \exists SMALL CUTS

SAMPLE & SOLVE

Adjacency Matrix G, H are ϵ -near iff

$$\|G - H\|_1 \leq \epsilon n^2$$

S & S PARADIGM :

1. $W \leftarrow$ random sample of vertices
2. ACCEPT iff subgraph induced by W has some property

[Goldreich-Trevisan]

S & S is UNIVERSAL FOR

$f(\epsilon)$ QUERY COMPLEXITY

[Alon]

IS GRAPH H -FREE ?

- query time $f(|H|, \epsilon)$
- f is poly iff H bipartite

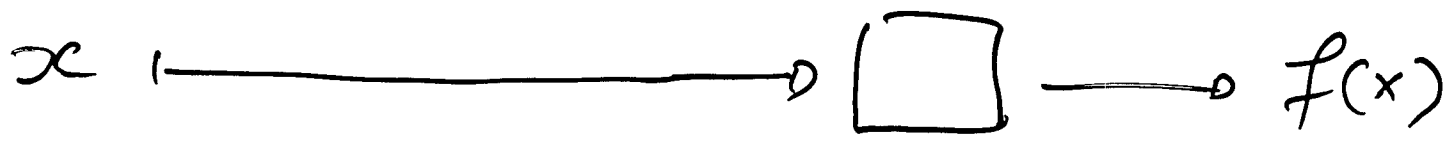
[Alon-Fischer-Krivelevich-Szegedy]

- $\exists V \forall E$ in $f(\epsilon)$ queries
- $\forall V \exists E$ impossible

"Property indistinguishability"

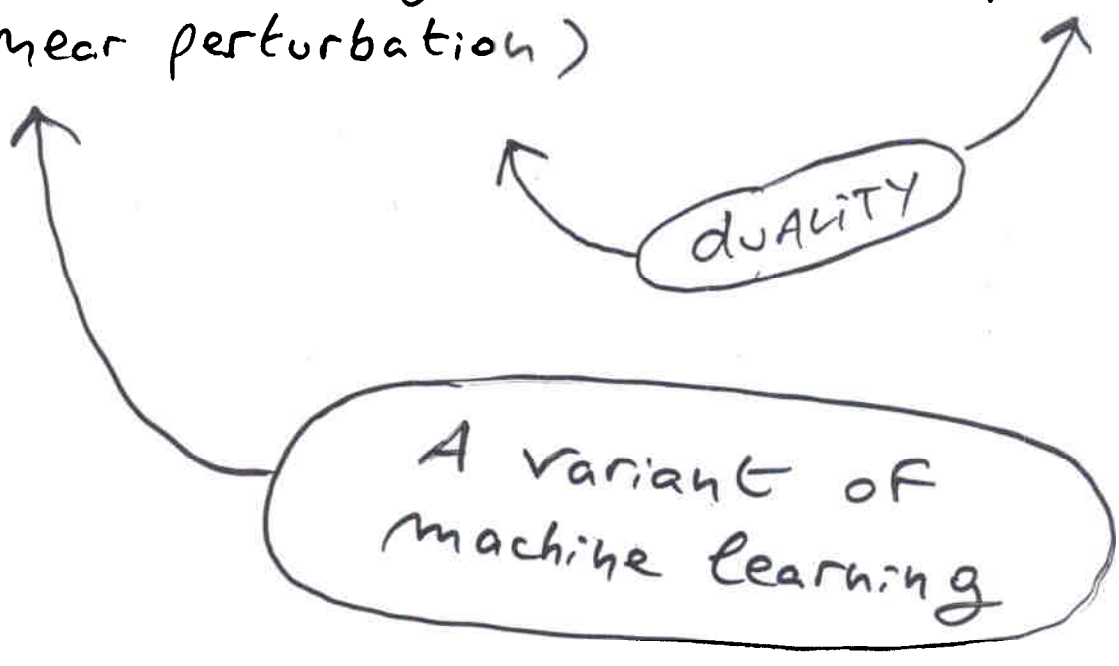
[Czumaj-Sohler]

- Abstract Combinatorial program



- Property checking
(ϵ -near perturbation)

Approximation



[Frieze-Kannan]

- Matrix Approximation

[Czumaj - Ergun - Fortnow - Magen - Newman -
Rubinfeld - Sohler]

- Euclidean MST

[Indyk] [Mishra - Oblinger - Pitt]

- Clustering

WORK IN PROGRESS

w/ Nir Ailon

Ding Liu

- Priced Point Location
- High-dim nearest neighbor
w/ $\epsilon_1, \dots, \epsilon_m$
- On line learning Genetic optimization