On Learning Monotone probability distributions over the Boolean cube.

R. Rubinfeld, MIT

A. Vasilyan, MIT

Learning distributions

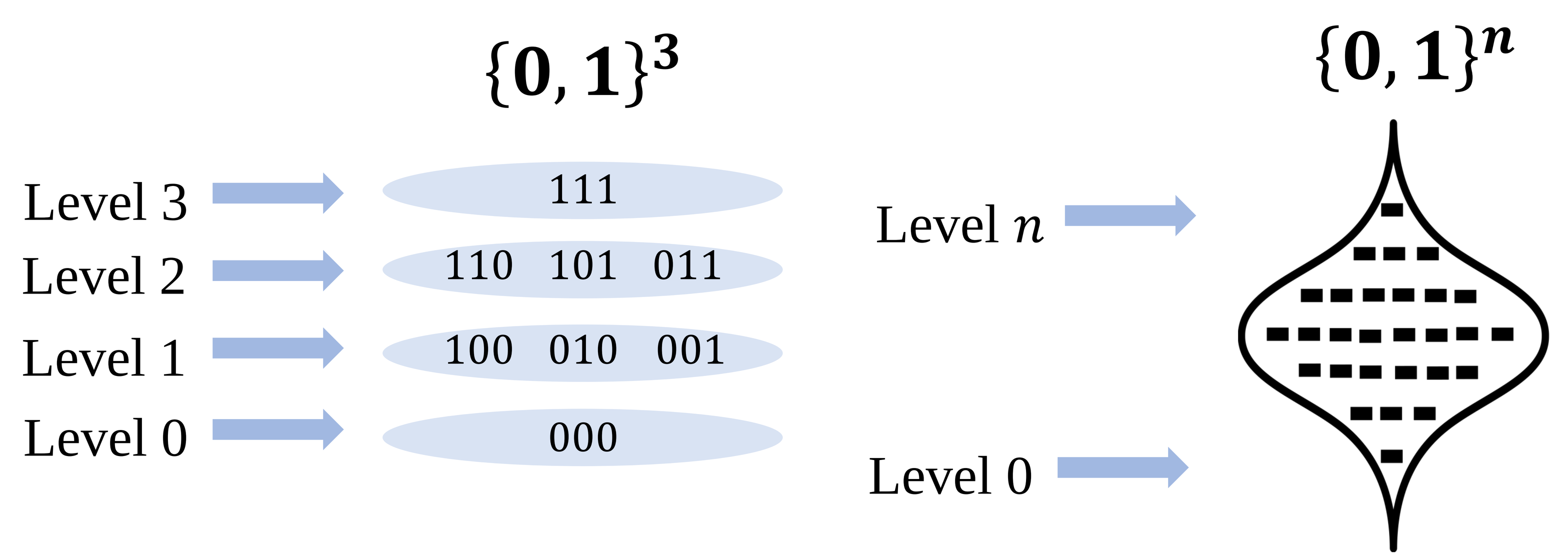
ρ is probability distribution over N elements

Samples from ρ → Algorithm → Estimate $\hat{\rho}$

Want $|\rho - \hat{\rho}|_1 \leq \varepsilon$ (w.p. $\geq 2/3$)

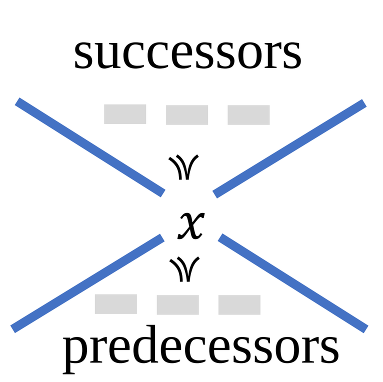
General ρ → need $\Omega(N)$ samples.
Motivates assuming ρ in special class.

Monotone distributions over Boolean cube

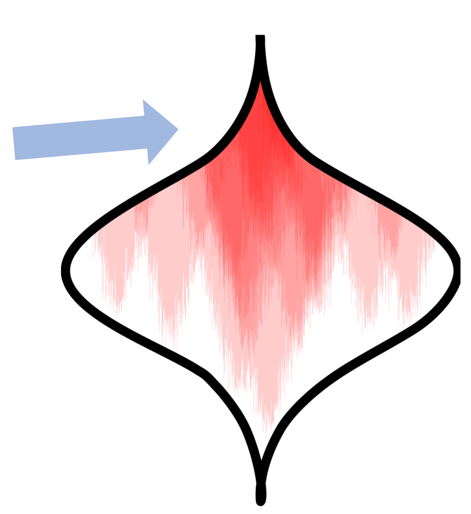


Partial order on $\{0,1\}^n$: if $x \succcurlyeq y$ if $x_i \geq y_i$ for all i.

ρ is **monotone** over $\{0,1\}^n$ if whenever $x \succcurlyeq y$, then $\rho(x) \geq \rho(y)$



Darker color means higher probability.



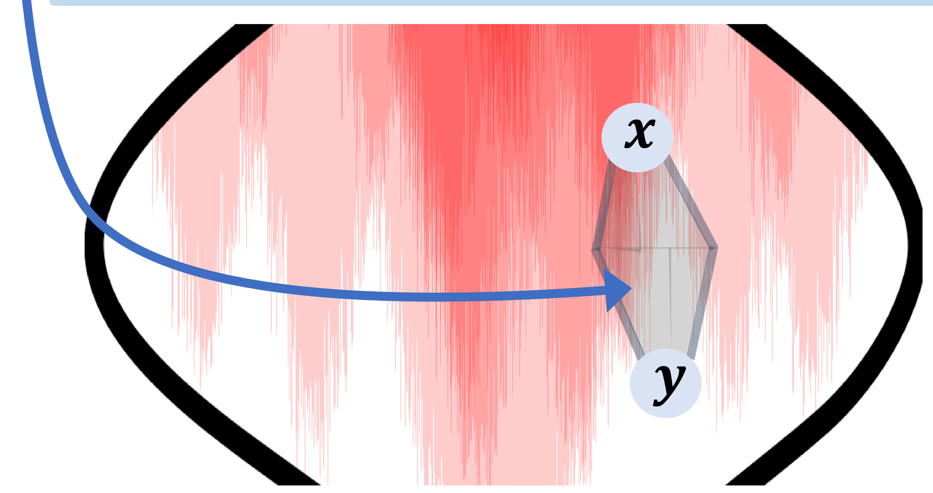
Main result

If ρ is monotone, can learn with $\frac{2^n}{2^{\Theta(n^{1/5})}}$ samples.

Together with the L_1 distance tester in [VV11], can be applied to test whether a distribution is monotone with $O\left(\frac{2^n}{n}\right)$ samples.

Algorithm outline

1. For every $i \in \{0, \cdots n\}$ carefully pick a parameter $L(i)$.
2. Define $\eta(x,y) := \frac{1}{2^{||x||-||y||}} \Pr_{z\sim\rho}[x \succcurlyeq z \succcurlyeq y]$, graphically the average density here. And $\hat{\eta}(x,y) := \frac{1}{2^{||x||-||y||}} \Pr_{z\sim\{\text{samples}\}}[x \succcurlyeq z \succcurlyeq y]$ is the empirical estimate of $\eta(x,y)$.
3. To estimate $\hat{\rho}(x)$, compute
$$\max_{y\ s.t.\ y \preccurlyeq x\ and\ ||y|| = ||x|| - L(|x|)} \hat{\eta}(x,y)$$



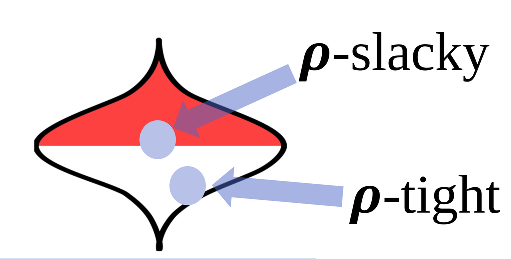
Definition: Tight and slacky elements

If f is monotone then $f(x) \geq \max_{y \prec x} f(y)$.

x is f**-tight** if have = above
Otherwise x is f**-slacky**

Examples:

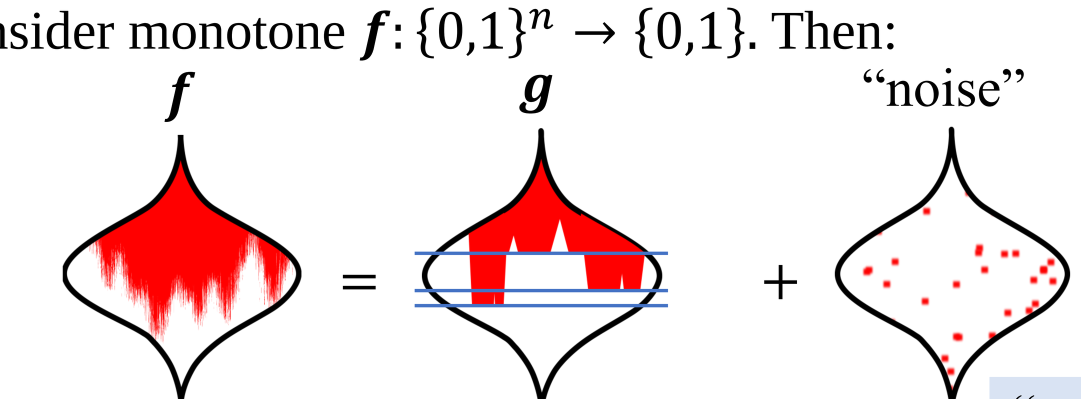
- ρ uniform in upper half:



- If f is {0,1}-valued : f**-slacky** elements ↔ DNF minterms

The case of {0,1}-valued functions: most levels are completely tight

Lemma [Blais, Håstad, Servedio, Tan '14] (restated):
Consider monotone $f: \{0,1\}^n \to \{0,1\}$. Then:

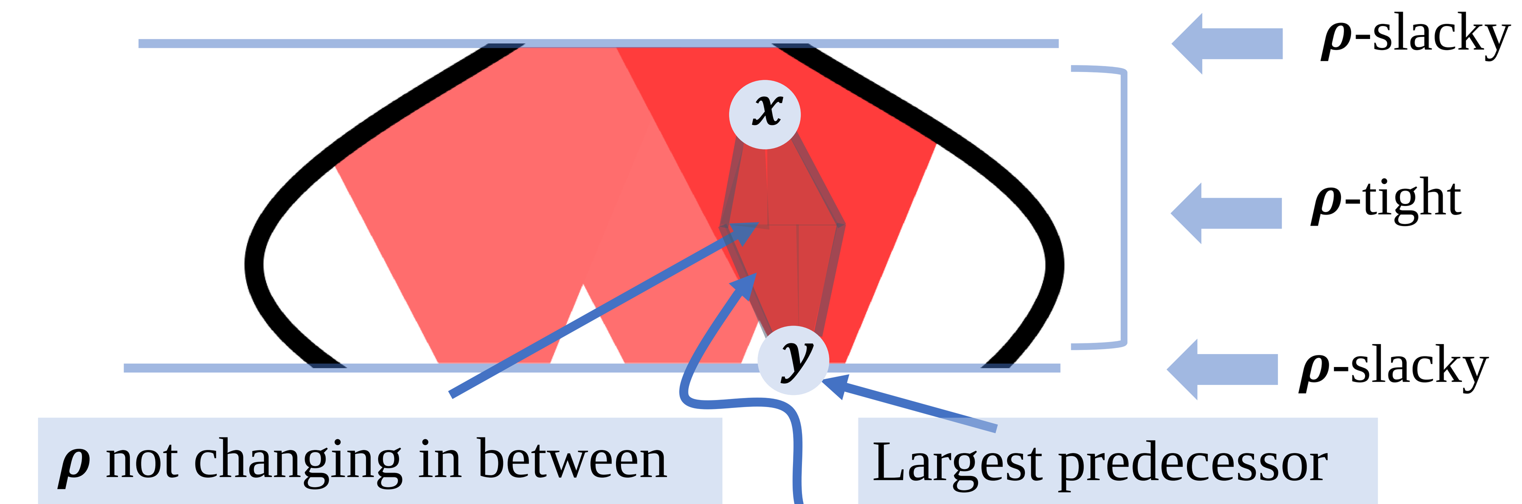


x is g-slacky ⇓ x in one of $O_\varepsilon(1)$ special levels

"noise" ≥ 0

$\frac{|\text{"noise"}|_1}{2^n} \leq \varepsilon$

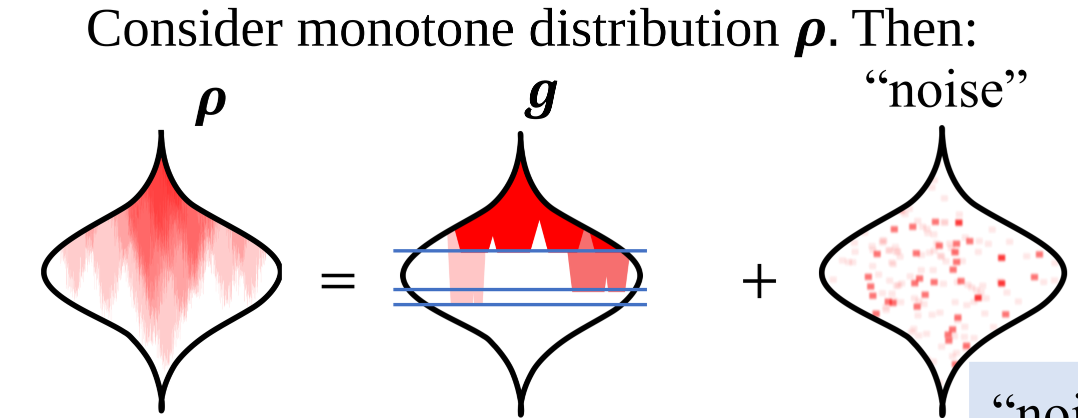
Few slacky levels is good for our algorithm



We then have $\rho(x) = \rho(y) = \eta(x,y)$
(recall $\eta(x,y)$ defined as average of ρ here).

Extending to monotone distributions

"Lemma":
Consider monotone distribution ρ. Then:



x is g-slacky ⇓ x in one of $O_\varepsilon(1)$ special levels

"noise" ≥ 0

$|\text{"noise"}|_1 \leq \varepsilon$

Unfortunately, one cannot guarantee literally this.
We get around this issue by carefully assigning weights to levels and bounding total weight of special slacky levels.

References

[AGPRY19] Maryam Aliakbarpour, Themis Gouleakis, John Peebles, Ronitt Rubinfeld and Anak Yodpinyanee. Towards Testing Monotonicity of Distributions Over General Posets COLT 2019/

[BHST14] Eric Blais, Johan Hastad, Rocco A Servedio, and Li-Yang Tan. On DNF approximators for monotone Boolean functions. ICALP 2014.

[BFRV11] Arnab Bhattacharyya, Eldar Fischer, Ronitt Rubinfeld and Paul Valiant. Testing monotonicity of distributions over general partial orders. ICS 2011

[VV11] Paul Valiant. Testing symmetric properties of distributions. SIAM Journal on Computing, 40(6):1927–1968, 2011.