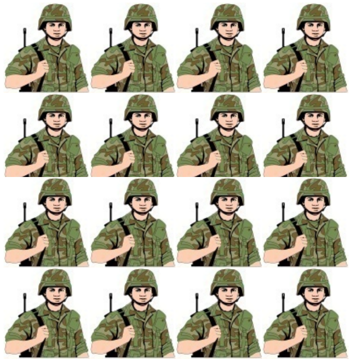
Bounds for Group Testing Algorithms to Determine the Number of Defectives

Nader H. Bshouty
Technion

Group Testing



Robert Dorfman's paper in 1943 introduced the field of (Combinatorial) **Group Testing**. The motivation arose during the Second World War when the United States Public Health Service and the Selective service embarked upon a large scale project. The objective was to weed out all syphilitic men called up for induction. However, syphilis testing back then was expensive and testing every soldier individually would have been very cost heavy and inefficient.

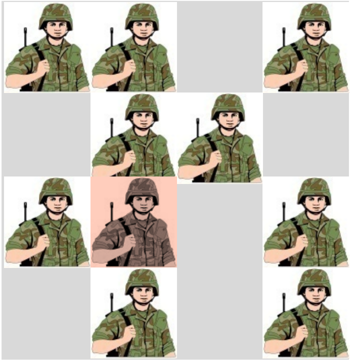
We can combine blood samples and test a combined sample together to check if at least one soldier has syphilis.

x_1 , x_2 , x_3 , x_4,
x_5 , x_6 , x_7 , x_8,
x_9 , x_{10} , x_{11} , x_{12}
x_{13} , x_{14}, x_{15}, x_{16}

Detecting the Defective items

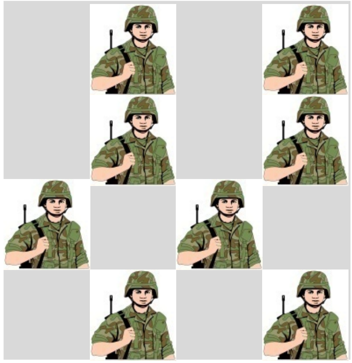
Determining the Number of Defective items

Estimate the Number of Defective items



Positive

1101
0110 1
1101
0101



Negative

0101
0101 0
1010
0101

Positive

0010
1010 1
0001
1010

Many other applications

Soldiers $\Rightarrow$ Items

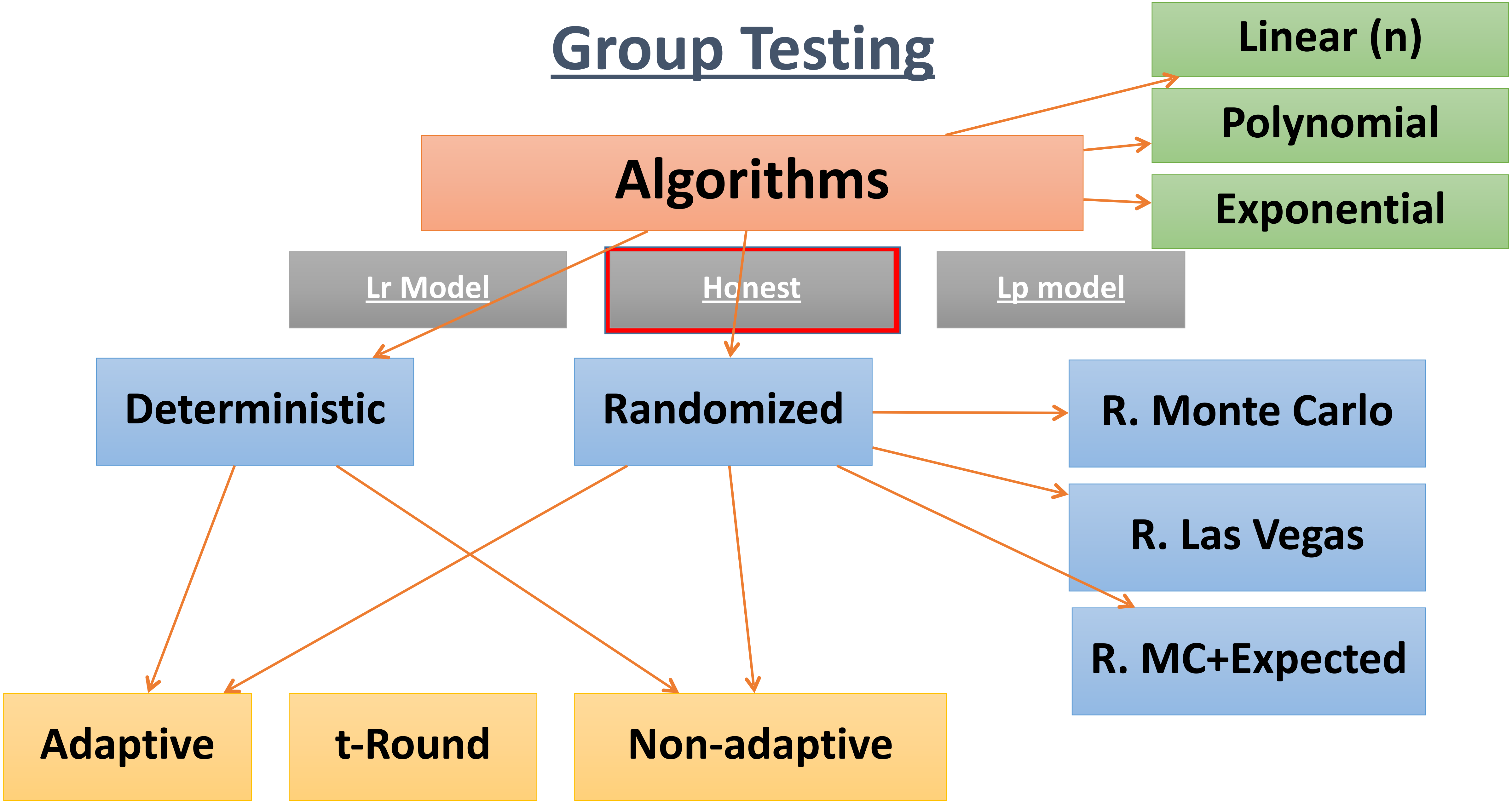
Sick soldiers $\Rightarrow$ defective items

n #items

d #defective items

$f = x_3 \vee x_{10}$

Group Testing



n #items

d #defective items

Detecting the Defective items

Algorithm	Lower Bound	Upper Bound
Schlaghoff and Triesch 2005 and Cheng et al. 2014, 2015 and many others		
Adaptive Deterministic or Randomized (poly)		
Folklore Result		
Non-Adaptive Randomized MC (poly)		
A.G. D'yachkov, V.V. Rykov 1982 Porat and Rothschild 2011		
Non-Adaptive Deterministic (poly)		
A. De Bonis, L. Gasieniec, U. Vaccaro 2005		
Two-round Deterministic (exp)		
M. Cheraghchi 2013		
Two-round Deterministic (poly)		

$$n \text{ tests } \rightarrow d\log n - d^2\log n \text{ tests}$$

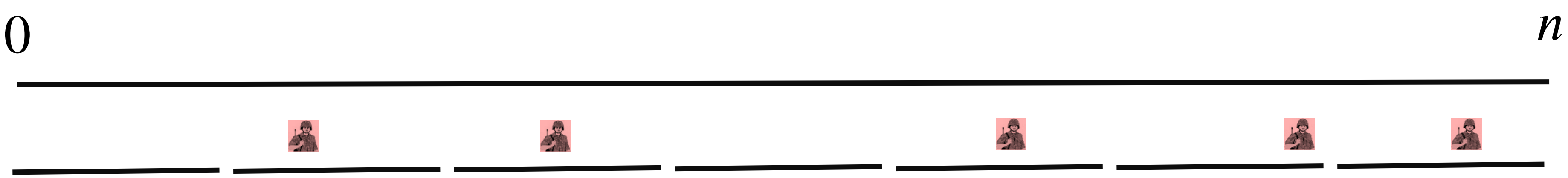
Determining the Number of Defective items

Algorithm	Lower Bound	Upper Bound
Cheng. 2011		
Adaptive MC Rand.		
Bshouty, Haddad-Zaknoon, Boulos, Moalem, Nada, Noufi, Zaknoon 2020		
Adaptive MC Rand.		
Bshouty, Bshouty-Hurani, Haddad, Hashem, Khoury, Sharafy 2018		
Adaptive Deterministic or LV Rand.		
Damaschke and Sheikh Muhammad 2009		
Non-Adaptive Deterministic		
Bshouty 2019		
Non-Adaptive MC and LV		

Determining the Number of Defective items

Algorithm	Lower Bound	Upper Bound
Cheng. 2011		
Adaptive MC Rand.		
Bshouty, Haddad-Zaknoon, Boulos, Moalem, Nada, Noufi, Zaknoon 2020		
Adaptive MC Rand.		
Bshouty, Bshouty-Hurani, Haddad, Hashem, Khoury, Sharafy 2018		
Adaptive Deterministic or LV Rand.		
Damaschke and Sheikh Muhammad 2009		
Non-Adaptive Deterministic		
Bshouty 2019		
Non-Adaptive MC and LV		

Monte Carlo Adaptive – Upper Bound



$d \le Y \le 2d$ with failure probability at most $\dfrac{\delta}{2}$ - can be done in $d + \log d \log 1/\delta$

$r = \dfrac{Y^2}{\delta}$ sets $\quad \Pr[Fail] = 1 - \left(1 - \dfrac{1}{r}\right)\left(1 - \dfrac{2}{r}\right)\cdots\left(1 - \dfrac{d-1}{r}\right)$

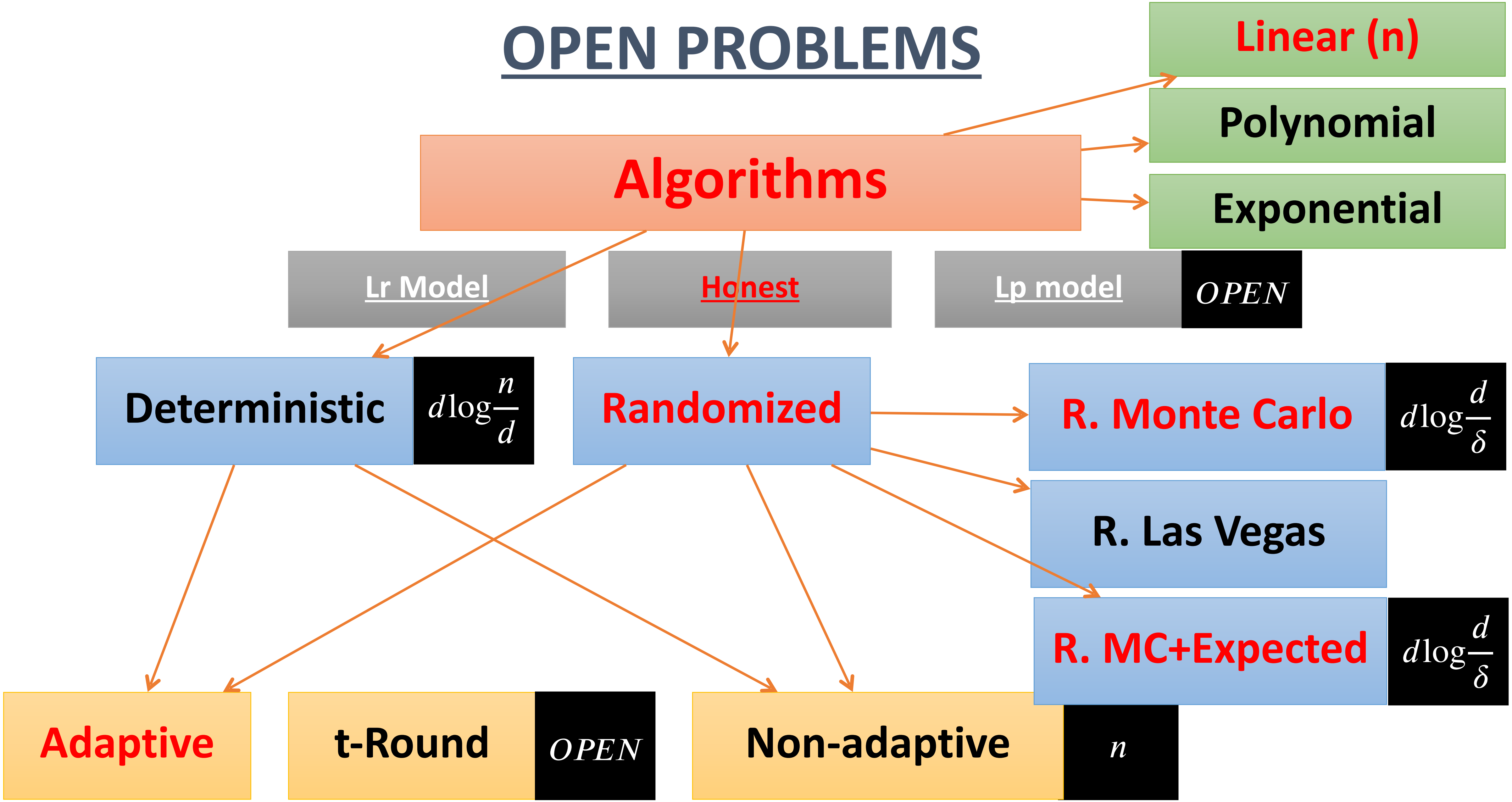
$$\le \frac{1}{r} + \frac{2}{r} + \cdots + \frac{d-1}{r} \le \frac{d^2}{2r} \le \frac{\delta}{2}$$

Schlaghoff and Triesch 2005
and Cheng et al. 2014, 2015

$d \log \dfrac{n}{d} + O(d)$

$$d \log \frac{r}{d} = d \log \frac{Y^2}{d\delta} = d \log \frac{4d}{\delta} = d \log \frac{d}{\delta} + O(d) \quad \blacksquare$$

OPEN PROBLEMS



Estimating the Number of Defective items

$$d/\Delta \le x \le \Delta d \qquad \Delta = 1 + \Omega(1)$$

Algorithm	Lower Bound	Upper Bound
Damaschke and Sheikh Muhammad 2010		
Non-Adaptive MC Rand.		

Bshouty 2019		
Non-Adaptive MC Rand.		

Bshouty 2019		
Non-Adaptive Deterministic or LV Rand.		

Estimating the Number of Defective items

Given some upper bound $d \le D$

$$d/\Delta \le x \le \Delta d$$

Algorithm	Lower Bound	Upper Bound
Damaschke and Sheikh Muhammad 2010		
Non-Adaptive MC Rand.		

Bshouty 2019		
Non-Adaptive MC Rand.		

<u>Deterministic</u> <u>Estimating the Number of Defective items</u>

Bshouty Haddad-Zaknoon 2020

<u>Given some upper bound $d \le D$</u>

$d/\Delta \le x \le \Delta d$

Bounds	Adaptive/ Non-Adapt.	Result	Explicit/ Non-Expl.	Ref.
Lower B.	Non-Adapt.	$\frac{D}{\Delta^2} \log \frac{n}{D}$	-	[1]
Lower B.	Adaptive	$\frac{D}{\Delta^2} \log \frac{n}{D}$	-	Ours
Upper B.	Adaptive	$\frac{D}{\Delta^2} \left(\log \frac{n}{D} + \log \Delta\right)$	Explicit	Ours
Upper B.	Non-Adapt.	$\frac{\log D}{\log \Delta} D \log n$	Non-Expl.	[1]
Upper B.	Non-Adapt.	$\frac{D}{\Delta^2} \left(\log \frac{n}{D} + \log \Delta\right)$	Non-Expl.	Ours
Upper B.	Non-Adapt.	$\frac{D^{1+o(1)}}{\Delta^2} \log n$	Explicit[1]	Ours
Upper B.	Non-Adapt.	$\frac{D}{\Delta^2} \cdot \text{Quazipoly}(\log n)$	Explicit	Ours

[1] Abhishek Agarwal, Larkin Flodin, and Arya Mazumdar. Estimation of sparsity via simple measurements. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 456–460. IEEE, 2017.

Estimate the Number of Defective items

$$(1-\epsilon)d \leq Y \leq (1+\epsilon)d$$

Algorithm	Lower Bound	Upper Bound
Cheng and Xu (2014)		
Adaptive Rand. MC with		

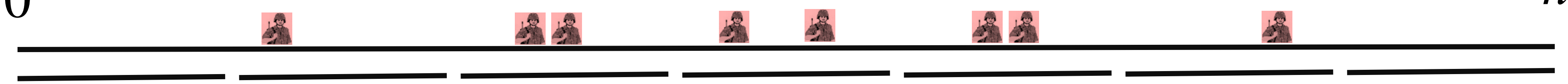
Ron and Tsur (2014)		
Adaptive Rand. MC with		

Falahatgar et al. (2016)		
Adaptive Randomized MC + Expected number queries		

Algorithm	Lower Bound	Upper Bound
Cheng and Xu (2014)		
Adaptive Randomized MC with		
Ron and Tsur (2014)		
Adaptive Randomized MC with		
Falahatgar et al. (2016)		
Adaptive Randomized MC + Expected number of queries		
Bshouty, Bshouty-Hurani, Haddad, Hashem, Khoury, Sharafy 2018		
Adaptive Deterministic and Las Vegas		
Adaptive Randomized MC + Expected number of queries		
Adaptive Randomized MC		

Algorithm	Lower Bound	Upper Bound
Cheng and Xu (2014)		
Adaptive Randomized MC with		
Ron and Tsur (2014)		
Adaptive Randomized MC with		
Falahatgar et al. (2016)		
Adaptive Randomized MC + Expected number of queries		
Bshouty, Bshouty-Hurani, Haddad, Hashem, Khoury, Sharafy 2018		
Adaptive Deterministic and Las Vegas		
Adaptive Randomized MC + Expected number of queries		
Adaptive Randomized MC		

Deterministic Adaptive – Upper Bound



$\frac{1}{1-\epsilon}$

$(1-\epsilon)n$ sets

Schlaghoff and Triesch 2005
and Cheng et al. 2014, 2015

$d \log \frac{n}{d} + O(d)$

$d \log \frac{(1-\epsilon)n}{d}$

$\frac{Y}{1-\epsilon} \geq d$

Y

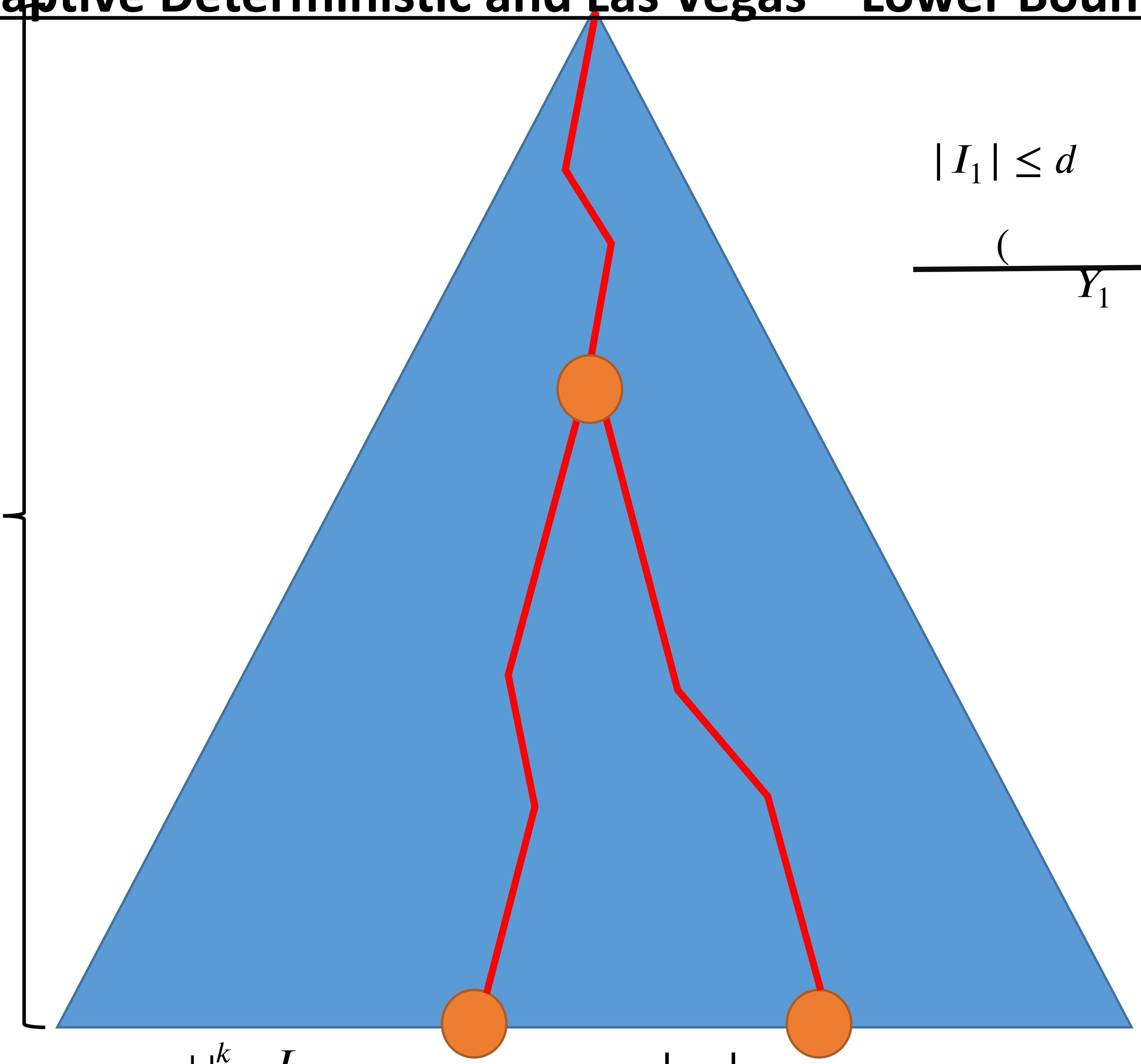
$Y \leq d$

$Y \geq (1-\epsilon)d$

$(1-\epsilon)d \leq Y \leq d$

Algorithm	Lower Bound	Upper Bound
Cheng and Xu (2014)		
Adaptive Randomized MC with		
Ron and Tsur (2014)		
Adaptive Randomized MC with		
Falahatgar et al. (2016)		
Adaptive Randomized MC + Expected number of queries		
Bshouty, Bshouty-Hurani, Haddad, Hashem, Khoury, Sharafy 2018		
Adaptive Deterministic and Las Vegas		
Adaptive Randomized MC + Expected number of queries		
Adaptive Randomized MC		

Adaptive Deterministic and Las Vegas – Lower Bound



$$\log\frac{\binom{n}{d}}{k}$$
$$= d\log\frac{(1-\epsilon)n}{d}$$

$$Expect = d\log\frac{(1-\epsilon)n}{d}$$

Yao MinMax Theorem

$\cup_{i=1}^{k} J_i$

$J_1, J_2, J_3, \cdots, J_k$ $\quad \left| J_i \right| = d$

$(1-\epsilon)d \le Y \le (1+\epsilon)d$

$|I_1| \le d$ $\quad \left| I_2 \right| \ge \frac{1+\epsilon}{1-\epsilon}(d+1)$

( ) ( )

Y_1 $\quad Y_2$

$Y_1 \ne Y_2$

$$\left| \cup_{i=1}^{k} J_i \right| < \frac{1+\epsilon}{1-\epsilon}(d+1)$$

$$k \le \binom{\frac{1+\epsilon}{1-\epsilon}(d+1)}{d}$$

$$\frac{\binom{n}{d}}{k}$$

OPEN PROBLEMS

Algorithm	Lower Bound	Upper Bound
Randomized MC + Expected number of queries		
Randomized MC		
Randomized MC		

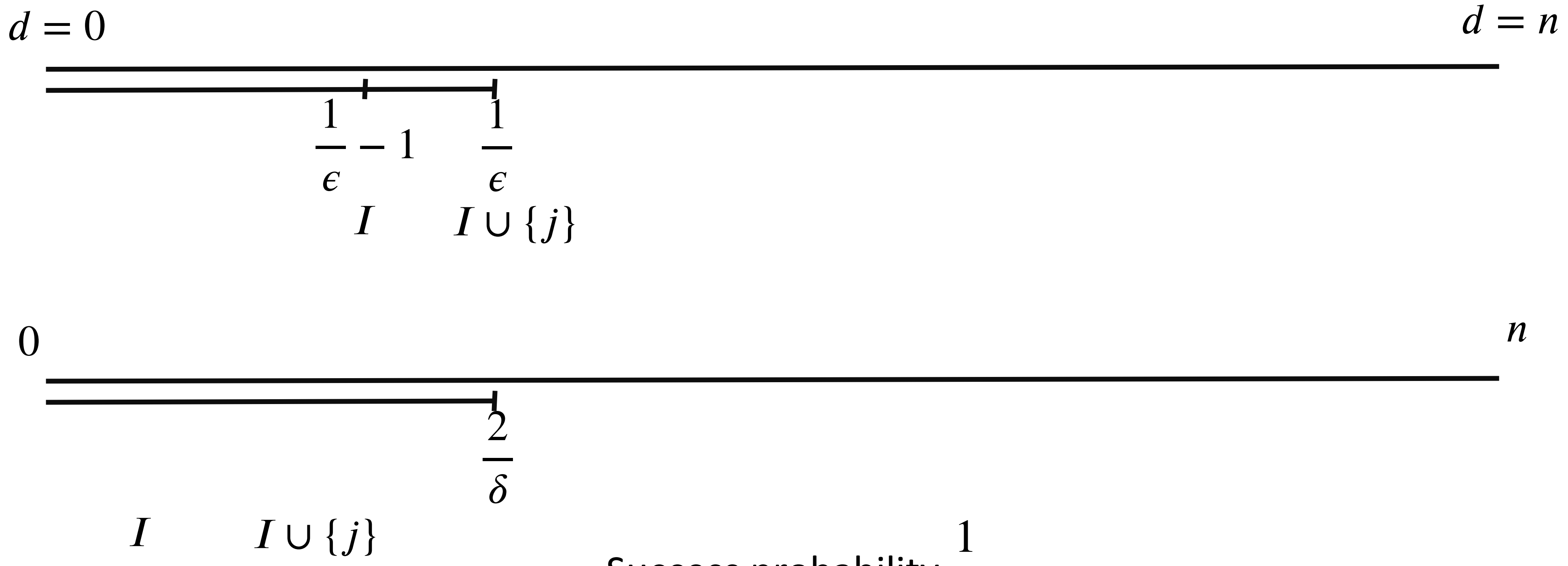
Thank You

Thank You

Algorithm	Lower Bound	Upper Bound
	Cheng and Xu (2014)	
Randomized MC with		
	Ron and Tsur (2014)	
Randomized MC with		
	Falahatgar et al. (2016)	
Randomized MC + Expected number of queries		
	Ours	
Deterministic and Las Vegas		
Randomized MC + Expected number of queries		
Randomized MC		

Monte Carlo adaptive algorithm

$$(1-\epsilon)d \le D \le (1+\epsilon)d$$

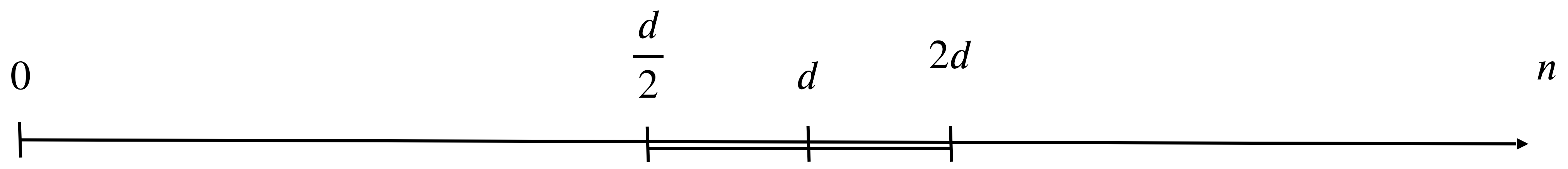


Success probability $\frac{1}{2}$

$$M = \frac{1}{2}\binom{\frac{2}{\delta}}{\frac{1}{\epsilon}}$$

$$\log M = \Omega\left(\frac{1}{\epsilon}\log\frac{1}{\delta}\right)$$

Algorithm	Lower Bound	Upper Bound
Cheng and Xu (2014)		
Randomized MC with		
Ron and Tsur (2014)		
Randomized MC with		
Falahatgar et al. (2016)		
Randomized MC + Expected number of queries		
Ours		
Deterministic and Las Vegas		
Randomized MC + Expected number of queries		
Randomized MC		



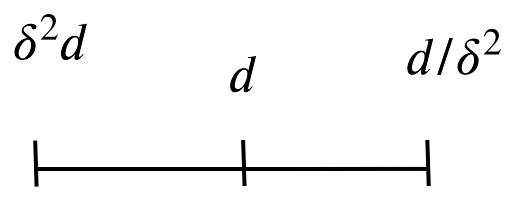
Doubling technique $a_i = 2^i$ $\log d$

Falahatgar et al. (2016) $a_i = 2^{2^i}$ $d^{\frac{1}{2}} \leq D_1, D_2 < d^2$ $b_1 = \left(D_1 D_2\right)^{1/2}$

$\log\log d$ $\log\log d$ $2\log\log d$

Ours $a_i = 2^{2^{i^2}}$ $d^{2^{-2\sqrt{\log\log d}}} \leq D_1, D_2 < d^{2^{2\sqrt{\log\log d}}}$ $b_1 = \left(D_1 D_2\right)^{1/2}$

$\sqrt{\log\log d}$ $\log\log d + \sqrt{\log\log d}$ $\log\log d$



$\log\log d$ $+O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$ $(1-\delta)\log\log d + O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$