

Subquadratic Non-adaptive Threshold Group Testing

Gianluca De Marco¹ Tomasz Jurdziński² Michał Różański²
Grzegorz Stachowiak²

¹Dipartimento di Informatica,
University of Salerno, Italy

²Institute of Computer Science
University of Wrocław, Poland

FCT 2017
Bordeaux, France

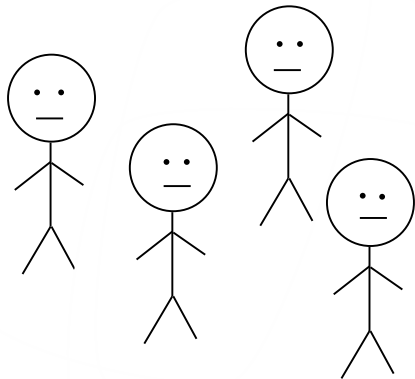
Motivation – Group Testing

- Introduced by Dorfman in '43, during World War II, as a part of a large project started by United States Public Health Service, to weed out syphilitic men called up for induction.

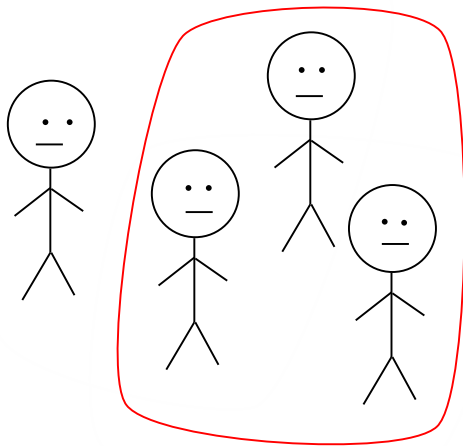
Motivation – Group Testing

- Introduced by Dorfman in '43, during World War II, as a part of a large project started by United States Public Health Service, to weed out syphilitic men called up for induction.
- Also used in: machine learning, cryptography, multiple access channel communication, DNA sequencing, medical examination, streaming algorithms, and more.

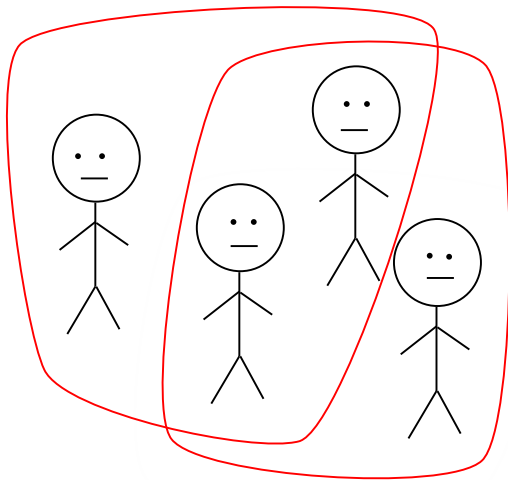
Motivation – Group Testing



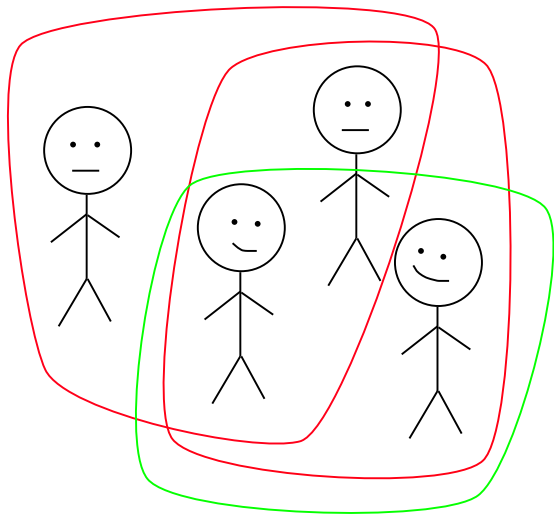
Motivation – Group Testing



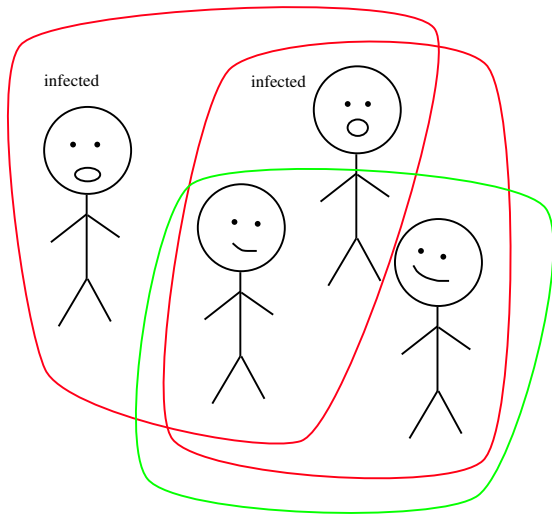
Motivation – Group Testing



Motivation – Group Testing



Motivation – Group Testing



Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.
- *Feedback function* $f_P : 2^{[n]} \rightarrow \{0, 1\}$
 $f_P(Q) = 1$ if $|Q \cap P| \geq 1$, and 0 otherwise.

Motivation

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.
- *Feedback function* $f_P : 2^{[n]} \rightarrow \{0, 1\}$
 $f_P(Q) = 1$ if $|Q \cap P| \geq 1$, and 0 otherwise.
- A sequence $\mathcal{Q} = (Q_1, \dots, Q_m)$ is called a *pooling strategy* of size m .

Motivation

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.
- *Feedback function* $f_P : 2^{[n]} \rightarrow \{0, 1\}$
 $f_P(Q) = 1$ if $|Q \cap P| \geq 1$, and 0 otherwise.
- A sequence $\mathcal{Q} = (Q_1, \dots, Q_m)$ is called a *pooling strategy* of size m .
- A *measurement* is a vector $y = (f_P(Q_1), \dots, f_P(Q_m))$.

Motivation

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.
- *Feedback function* $f_P : 2^{[n]} \rightarrow \{0, 1\}$
 $f_P(Q) = 1$ if $|Q \cap P| \geq 1$, and 0 otherwise.
- A sequence $\mathcal{Q} = (Q_1, \dots, Q_m)$ is called a *pooling strategy* of size m .
- A *measurement* is a vector $y = (f_P(Q_1), \dots, f_P(Q_m))$.

Pooling strategy \mathcal{Q} is valid if for any set of positives P of size d , it is possible to discover the set P basing only on the measurement vector.

Motivation

Group Testing

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.
- *Feedback function* $f_P : 2^{[n]} \rightarrow \{0, 1\}$
 $f_P(Q) = 1$ if $|Q \cap P| \geq 1$, and 0 otherwise.
- A sequence $\mathcal{Q} = (Q_1, \dots, Q_m)$ is called a *pooling strategy* of size m .
- A *measurement* is a vector $y = (f_P(Q_1), \dots, f_P(Q_m))$.

Pooling strategy \mathcal{Q} is valid if for any set of positives P of size d , it is possible to discover the set P basing only on the measurement vector.

Goal: design a valid pooling strategy with small number of tests.

Motivation

Example

Let $\mathcal{Q} = (Q_1, \dots, Q_n)$, where $Q_i = \{i\}$.

Motivation

Example

Let $\mathcal{Q} = (Q_1, \dots, Q_n)$, where $Q_i = \{i\}$.

For a measurement $y = (1, 0, 0, 1, 0, 1, 0, \dots, 0)$ we can derive the set of positives $P = \{1, 3, 7\}$.

Motivation

Example

Let $\mathcal{Q} = (Q_1, \dots, Q_n)$, where $Q_i = \{i\}$.

For a measurement $y = (1, 0, 0, 1, 0, 1, 0, \dots, 0)$ we can derive the set of positives $P = \{1, 3, 7\}$.

- \mathcal{Q} discovers all positives explicitly.

Motivation

Example

Let $\mathcal{Q} = (Q_1, \dots, Q_n)$, where $Q_i = \{i\}$.

For a measurement $y = (1, 0, 0, 1, 0, 1, 0, \dots, 0)$ we can derive the set of positives $P = \{1, 3, 7\}$.

- \mathcal{Q} discovers all positives explicitly.
- How to decode measurement vectors for more complex queries?

Motivation

Example

Let $\mathcal{Q} = (Q_1, \dots, Q_n)$, where $Q_i = \{i\}$.

For a measurement $y = (1, 0, 0, 1, 0, 1, 0, \dots, 0)$ we can derive the set of positives $P = \{1, 3, 7\}$.

- \mathcal{Q} discovers all positives explicitly.
- How to decode measurement vectors for more complex queries?
- The length of \mathcal{Q} is n . Can we do better knowing that there are $d \ll n$ positives?

Group Testing

Results

- Introduction – [Dorfman '43]
- Explicit construction $O(d^2 \log^2 n)$, existential result $O(d^2 \log n)$, connections to coding theory – [Kautz, Singleton '64]
- Lower bound $\Omega(d^2 \log_d n)$ – [Dyachkov, Rykov '82; Furedi '96]
- Explicit construction with $O(d^2 \log n)$ tests – [Porat, Rotschild '08]

Group Testing

Results

- Introduction – [Dorfman '43]
- Explicit construction $O(d^2 \log^2 n)$, existential result $O(d^2 \log n)$, connections to coding theory – [Kautz, Singleton '64]
- Lower bound $\Omega(d^2 \log_d n)$ – [Dyachkov, Rykov '82; Furedi '96]
- Explicit construction with $O(d^2 \log n)$ tests – [Porat, Rotschild '08]

Research directions

- Gap between lower and upper bound
- Efficient construction
- Decoding – how to obtain a set of positives from a measurement vector effectively?
- Practical aspects, applications & generalizations

Threshold Group Testing

Definition

- The *population* is identified with $[n]$. An *individual* is a number from the set.
- Distinguished, unknown set of elements $P \subseteq [n]$ of size d , called *positives*.
- *Testing pool* – $Q \subseteq [n]$.
- *Feedback function* $f_P : 2^{[n]} \rightarrow \{0, 1\}$
 $f_{P,t}(Q) = 1$ if $|Q \cap P| \geq t$, and 0 otherwise.
- A sequence $\mathcal{Q} = (Q_1, \dots, Q_m)$ is called a *pooling strategy* of size m .
- A *measurement* is a vector $y = (f_P(Q_1), \dots, f_P(Q_m))$.

Group Testing \equiv Threshold Group Testing with threshold 1.

Threshold Group Testing

Previous Results

- Introduction of threshold group testing (adaptive)– Damaschke ('05)
- Non-adaptive result with $O(d^{t+1} \log n)$ tests – Chen, Fu ('09)
- Improved result for *constant* threshold, $O(d^2 \log d \log(n/d))$ (existential), explicit construction with $O(d^3 \log d \log(n/d))$ tests – Cheraghchi ('13)

Threshold Group Testing

Previous Results

- Introduction of threshold group testing (adaptive)– Damaschke ('05)
- Non-adaptive result with $O(d^{t+1} \log n)$ tests – Chen, Fu ('09)
- Improved result for *constant* threshold, $O(d^2 \log d \log(n/d))$ (existential), explicit construction with $O(d^3 \log d \log(n/d))$ tests – Cheraghchi ('13)

Our result

- There exists a pooling design of size $O\left(\frac{d^2}{q(d,t)} \log \frac{n}{d}\right)$ for a threshold group testing, where $q(d, t) = \Omega\left(\sqrt{\frac{dt}{d-t}}\right)$.

Threshold Group Testing

Previous Results

- Introduction of threshold group testing (adaptive)– Damaschke ('05)
- Non-adaptive result with $O(d^{t+1} \log n)$ tests – Chen, Fu ('09)
- Improved result for *constant* threshold, $O(d^2 \log d \log(n/d))$ (existential), explicit construction with $O(d^3 \log d \log(n/d))$ tests – Cheraghchi ('13)

Our result

- There exists a pooling design of size $O\left(\frac{d^2}{q(d,t)} \log \frac{n}{d}\right)$ for a threshold group testing, where $q(d,t) = \Omega\left(\sqrt{\frac{dt}{d-t}}\right)$.
- Corollary: there exists a pooling design of size $O(d^{3/2} \log(n/d))$ for $t = d/2$.

Threshold Group Testing

Previous Results

- Introduction of threshold group testing (adaptive)– Damaschke ('05)
- Non-adaptive result with $O(d^{t+1} \log n)$ tests – Chen, Fu ('09)
- Improved result for *constant* threshold, $O(d^2 \log d \log(n/d))$ (existential), explicit construction with $O(d^3 \log d \log(n/d))$ tests – Cheraghchi ('13)

Our result

- There exists a pooling design of size $O\left(\frac{d^2}{q(d,t)} \log \frac{n}{d}\right)$ for a threshold group testing, where $q(d,t) = \Omega\left(\sqrt{\frac{dt}{d-t}}\right)$.
- Corollary: there exists a pooling design of size $O(d^{3/2} \log(n/d))$ for $t = d/2$.
- Lower bound: $\Omega(\min(d/t)^2, n/t)$.

Threshold Group Testing

Previous Results

- Introduction of threshold group testing (adaptive)– Damaschke ('05)
- Non-adaptive result with $O(d^{t+1} \log n)$ tests – Chen, Fu ('09)
- Improved result for *constant* threshold, $O(d^2 \log d \log(n/d))$ (existential), explicit construction with $O(d^3 \log d \log(n/d))$ tests – Cheraghchi ('13)

Our result

- There exists a pooling design of size $O\left(\frac{d^2}{q(d,t)} \log \frac{n}{d}\right)$ for a threshold group testing, where $q(d,t) = \Omega\left(\sqrt{\frac{dt}{d-t}}\right)$.
- Corollary: there exists a pooling design of size $O(d^{3/2} \log(n/d))$ for $t = d/2$.
- Lower bound: $\Omega(\min(d/t)^2, n/t)$.
- Multi-threshold group testing: $O(d^2/t \log n)$ for $\Theta(\sqrt{t})$ thresholds.

Further directions

- Tight bounds
- Explicit construction
- Decoding algorithms

Thank you!