Distribution of Missing Sums in Sumsets

Oleg Lazarev, Princeton University
Advisor: Steven Miller, Williams College
SMALL Research Program 2011, Williams College

2011 Young Mathematicians Conference
The Ohio State University, August 21, 2011

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$

Interval: $[a, b] = \{x \in \mathbb{N} : a \leq x \leq b\}$

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$
Interval: $[a, b] = \{x \in \mathbb{N} : a \leq x \leq b\}$

Example: if $A = \{1, 2, 5\}$, then

$$A + A = \{2, 3, 4, 6, 7, 10\}$$

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$
Interval: $[a, b] = \{x \in \mathbb{N} : a \leq x \leq b\}$

Example: if $A = \{1, 2, 5\}$, then

$$A + A = \{2, 3, 4, 6, 7, 10\}$$

Why study sumsets?

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$
Interval: $[a, b] = \{x \in \mathbb{N} : a \leq x \leq b\}$

Example: if $A = \{1, 2, 5\}$, then

$$A + A = \{2, 3, 4, 6, 7, 10\}$$

Why study sumsets?

- Goldbach's conjecture: $\{4, 6, 8, \cdots\} \subseteq P + P$.

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$
Interval: $[a, b] = \{x \in \mathbb{N} : a \leq x \leq b\}$

Example: if $A = \{1, 2, 5\}$, then

$$A + A = \{2, 3, 4, 6, 7, 10\}$$

Why study sumsets?

- Goldbach's conjecture: $\{4, 6, 8, \cdots\} \subseteq P + P$.
- Fermat's last theorem: let A_n be the nth powers and then ask if $(A_n + A_n) \cap A_n = \emptyset$ for all $n > 2$.

Background

Let $A \subseteq \mathbb{N} \cup \{0\}$.

Definition

Sumset: $A + A = \{x + y : x, y \in A\}$
Interval: $[a, b] = \{x \in \mathbb{N} : a \leq x \leq b\}$

Example: if $A = \{1, 2, 5\}$, then

$$A + A = \{2, 3, 4, 6, 7, 10\}$$

Why study sumsets?

- Goldbach's conjecture: $\{4, 6, 8, \cdots\} \subseteq P + P$.
- Fermat's last theorem: let A_n be the nth powers and then ask if $(A_n + A_n) \cap A_n = \emptyset$ for all $n > 2$.

Key Question: What is the structure of $A + A$?

Structure of Random Sets

- Consider finite $A \subseteq [0, n-1]$ chosen randomly with uniform distribution from all subsets of $[0, n-1]$.

Structure of Random Sets

- Consider finite $A \subseteq [0, n-1]$ chosen randomly with uniform distribution from all subsets of $[0, n-1]$.
- Question: What is the structure of $A + A$ for such A? what the distribution of $|A + A|$ for such A?

Structure of Random Sets

- Consider finite $A \subseteq [0, n-1]$ chosen randomly with uniform distribution from all subsets of $[0, n-1]$.
- Question: What is the structure of $A + A$ for such A? what the distribution of $|A + A|$ for such A?

Theorem: Martin, O'Bryant

$$\mathrm{E}|A + A| = 2n - 1 - 10 + O((3/4)^{n/2})$$

Structure of Random Sets

- Consider finite $A \subseteq [0, n-1]$ chosen randomly with uniform distribution from all subsets of $[0, n-1]$.
- Question: What is the structure of $A + A$ for such A? what the distribution of $|A + A|$ for such A?

Theorem: Martin, O'Bryant

$E|A + A| = 2n - 1 - 10 + O((3/4)^{n/2})$

Theorem: Zhao

For each fixed k, $P(A \subseteq [0, n-1] : |A + A| = 2n - 1 - k)$ has a limit as $n \to \infty$

Structure of Random Sets

- Consider finite $A \subseteq [0, n-1]$ chosen randomly with uniform distribution from all subsets of $[0, n-1]$.
- Question: What is the structure of $A + A$ for such A? what the distribution of $|A + A|$ for such A?

Theorem: Martin, O'Bryant

$\mathrm{E}|A + A| = 2n - 1 - 10 + O((3/4)^{n/2})$

Theorem: Zhao

For each fixed k, $\mathrm{P}(A \subseteq [0, n-1] : |A + A| = 2n - 1 - k)$ has a limit as $n \to \infty$

Note: Both theorem can be more naturally stated in terms of missing sums (independent of n).

Structure of Random Sets, Continued

- Why is the expectation so high? $E|A+A| \sim 2n - 11$.

Structure of Random Sets, Continued

- Why is the expectation so high? $E|A+A| \sim 2n - 11$.
- Main characteristic of typical $A+A$: middle is full.

Structure of Random Sets, Continued

- Why is the expectation so high? $E|A+A| \sim 2n - 11$.
- Main characteristic of typical $A + A$: middle is full.
- Many ways to write middle elements as sums

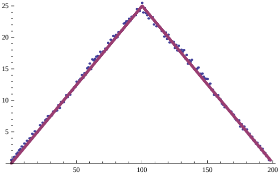


Figure: Comparison of predicted and observed number of representations of possible elements of the sumset, Miller

Structure of Random Sets, Continued

- Why is the expectation so high? $E|A+A| \sim 2n - 11$.
- Main characteristic of typical $A+A$: middle is full.
- Many ways to write middle elements as sums

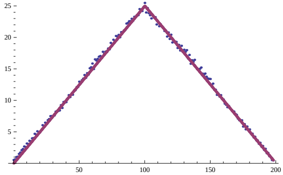


Figure: Comparison of predicted and observed number of representations of possible elements of the sumset, Miller

- Key fact: if $k < n$, then $\mathrm{P}(k \notin A+A) \sim \left(\frac{3}{4}\right)^{k/2}$

New Results

New Results

Theorem: Bounds on the distribution (Lazarev, 2011)

$$0.70^k \ll P(A + A \text{ has } k \text{ missing sums}) \ll 0.80^k$$

New Results

Theorem: Bounds on the distribution (Lazarev, 2011)

$$0.70^k \ll P(A + A \text{ has } k \text{ missing sums}) \ll 0.80^k$$

From numerical data, conjecture that

$$P(A + A \text{ has } k \text{ missing sums}) \sim 0.78^k$$

New Results

Theorem: Bounds on the distribution (Lazarev, 2011)

$$0.70^k \ll P(A + A \text{ has } k \text{ missing sums}) \ll 0.80^k$$

From numerical data, conjecture that

$$P(A + A \text{ has } k \text{ missing sums}) \sim 0.78^k$$

and Log(P($A + A$ has k missing sums)) is eventually linear:

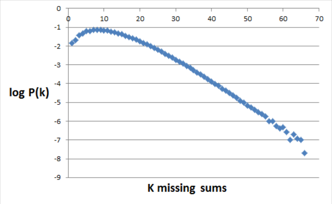


Figure: Log P(k missing sums) for $5 \cdot 10^8$ trial A of size 120

More Results

Theorem: Variance (Lazarev, 2011)

Exact series for the variance (too long to fit on page...); also Var $\sim$ 35.98.

More Results

Theorem: Variance (Lazarev, 2011)

Exact series for the variance (too long to fit on page...); also Var $\sim$ 35.98.

Theorem: Asymptotic formula for distribution of consecutive missing sums (Lazarev, 2011)

$$\left(\frac{1}{2}\right)^{(k+i)/2} \ll P(A+A \text{ is missing } [k, k+i]) \ll \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$$

More Results

Theorem: Variance (Lazarev, 2011)

Exact series for the variance (too long to fit on page...); also Var $\sim$ 35.98.

Theorem: Asymptotic formula for distribution of consecutive missing sums (Lazarev, 2011)

$$\left(\frac{1}{2}\right)^{(k+i)/2} \ll P(A+A \text{ is missing } [k, k+i]) \ll \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$$

Key Obstacle: Dependent random variables

More Results

Theorem: Variance (Lazarev, 2011)

Exact series for the variance (too long to fit on page...); also Var $\sim$ 35.98.

Theorem: Asymptotic formula for distribution of consecutive missing sums (Lazarev, 2011)

$$\left(\frac{1}{2}\right)^{(k+i)/2} \ll P(A+A \text{ is missing } [k, k+i]) \ll \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$$

Key Obstacle: Dependent random variables
Key Approach: Graph theory!

More Results

Theorem: Variance (Lazarev, 2011)

Exact series for the variance (too long to fit on page...); also Var $\sim$ 35.98.

Theorem: Asymptotic formula for distribution of consecutive missing sums (Lazarev, 2011)

$$\left(\frac{1}{2}\right)^{(k+i)/2} \ll P(A+A \text{ is missing } [k, k+i]) \ll \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$$

Key Obstacle: Dependent random variables
Key Approach: Graph theory! (and Fibonacci Numbers?)

Bounds on the Distribution

Bound on Distribution: Upper Bound

Weaker Upper bound: $P(A + A \text{ has } k \text{ missing sums}) < 0.93^k$

Proof sketch:

Bound on Distribution: Upper Bound

Weaker Upper bound: $P(A + A \text{ has } k \text{ missing sums}) < 0.93^k$

Proof sketch:

- Recall $P(k \notin A + A) = \left(\frac{3}{4}\right)^{k/2}$

Bound on Distribution: Upper Bound

Weaker Upper bound: $P(A + A \text{ has } k \text{ missing sums}) < 0.93^k$

Proof sketch:

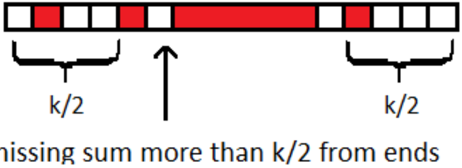
- Recall $P(k \notin A + A) = \left(\frac{3}{4}\right)^{k/2}$
- If k elements are missing, then missing one at least $k/2$ from the edges

Bound on Distribution: Upper Bound

Weaker Upper bound: $P(A + A \text{ has } k \text{ missing sums}) < 0.93^k$

Proof sketch:

- Recall $P(k \notin A + A) = \left(\frac{3}{4}\right)^{k/2}$
- If k elements are missing, then missing one at least $k/2$ from the edges

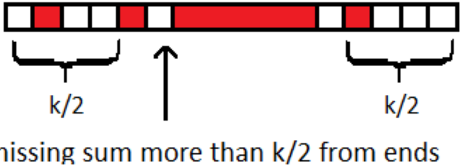


Bound on Distribution: Upper Bound

Weaker Upper bound: $P(A + A \text{ has } k \text{ missing sums}) < 0.93^k$

Proof sketch:

- Recall $P(k \notin A + A) = \left(\frac{3}{4}\right)^{k/2}$
- If k elements are missing, then missing one at least $k/2$ from the edges



- $P(A + A \text{ has } k \text{ missing sums}) < P(k/2 \notin A + A) < \left(\frac{3}{4}\right)^{k/4} \sim 0.93^k$

Bound on Distribution: Upper Bound

Weaker Upper bound: $P(A + A \text{ has } k \text{ missing sums}) < 0.93^k$

Proof sketch:

- Recall $P(k \notin A + A) = \left(\frac{3}{4}\right)^{k/2}$
- If k elements are missing, then missing one at least $k/2$ from the edges

- $P(A + A \text{ has } k \text{ missing sums}) < P(k/2 \notin A + A) < \left(\frac{3}{4}\right)^{k/4} \sim 0.93^k$

Note: Since the expectation is around 10 and expansion decay, all the higher moments are bounded as $n \to \infty$.

Bound on Distribution: Lower Bound

Lower bound: $P(A + A \text{ has } k \text{ missing sums}) > 0.01 \cdot 0.70^k$
Proof sketch: Construction.

Bound on Distribution: Lower Bound

Lower bound: $P(A + A \text{ has } k \text{ missing sums}) > 0.01 \cdot 0.70^k$

Proof sketch: Construction.

- Let the first $k/2$ be missing from A

Bound on Distribution: Lower Bound

Lower bound: $P(A + A \text{ has } k \text{ missing sums}) > 0.01 \cdot 0.70^k$

Proof sketch: Construction.

- Let the first $k/2$ be missing from A
- For the rest of elements, pick any set that fills in

Bound on Distribution: Lower Bound

Lower bound: $P(A + A \text{ has } k \text{ missing sums}) > 0.01 \cdot 0.70^k$
Proof sketch: Construction.

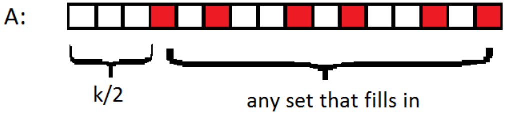
- Let the first $k/2$ be missing from A
- For the rest of elements, pick any set that fills in
- Martin/O'Bryant: $P(\text{ fills in}) > 0.01$ independent of n

Bound on Distribution: Lower Bound

Lower bound: $P(A + A \text{ has } k \text{ missing sums}) > 0.01 \cdot 0.70^k$

Proof sketch: Construction.

- Let the first $k/2$ be missing from A
- For the rest of elements, pick any set that fills in
- Martin/O'Bryant: $P(\text{ fills in}) > 0.01$ independent of n

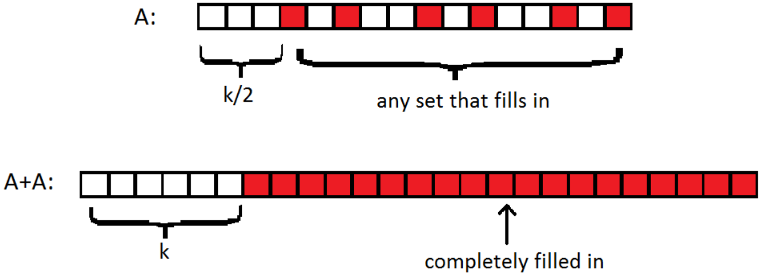


Bound on Distribution: Lower Bound

Lower bound: $P(A + A$ has k missing sums$) > 0.01 \cdot 0.70^k$

Proof sketch: Construction.

- Let the first $k/2$ be missing from A
- For the rest of elements, pick any set that fills in
- Martin/O'Bryant: $P($ fills in$) > 0.01$ independent of n

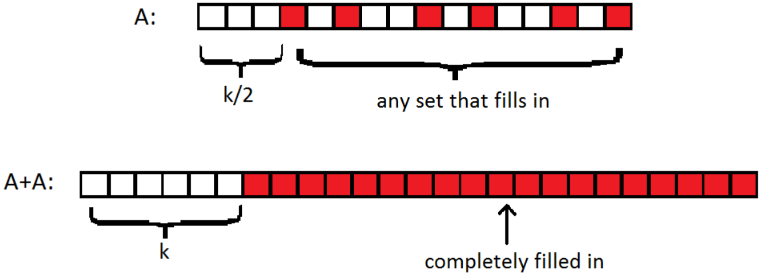


Bound on Distribution: Lower Bound

Lower bound: $P(A + A \text{ has } k \text{ missing sums}) > 0.01 \cdot 0.70^k$

Proof sketch: Construction.

- Let the first $k/2$ be missing from A
- For the rest of elements, pick any set that fills in
- Martin/O'Bryant: $P(\text{ fills in}) > 0.01$ independent of n



$P(A + A \text{ has } k \text{ missing sums}) > 0.01 \left(\frac{1}{2}\right)^{k/2} \sim 0.01 \cdot 0.70^k$

Variance

Variance

Can study $\text{Var}|A+A|$ instead.
Recall that

$$\text{Var}|A+A| = \text{E}(|A+A|^2) - (\text{E}|A+A|)^2$$

Variance

Can study $\mathrm{Var}|A+A|$ instead.
Recall that

$$\mathrm{Var}|A+A| = \mathrm{E}(|A+A|^2) - (\mathrm{E}|A+A|)^2$$

Therefore just need $\mathrm{E}(|A+A|^2)$:

$$\begin{aligned} \mathrm{E}(|A+A|^2) &= \frac{1}{2^n} \sum_{A \subseteq [0,n-1]} |A+A|^2 \\ &= \frac{1}{2^n} \sum_{A \subseteq [0,n-1]} \sum_{i,j \in A+A} 1 \\ &= \frac{1}{2^n} \sum_{0 \leq i,j \leq 2n-2} \sum_{A : i,j \in A+A} 1 \\ &= \sum_{0 \leq i,j \leq 2n-2} \mathrm{P}(A : i \text{ and } j \in A+A) \end{aligned}$$

Problem: Dependent Random Variables

It is sufficient to study $\mathrm{P}(A : i \text{ and } j \notin A + A)$:

$$\begin{aligned}\mathrm{P}(A : i \text{ and } j \in A + A) &= 1 - \mathrm{P}(A : i \notin A + A) - \mathrm{P}(A : j \notin A + A) \\ &\quad + \mathrm{P}(A : i \text{ and } j \notin A + A).\end{aligned}$$

Problem: Dependent Random Variables

It is sufficient to study $\mathrm{P}(A : i \text{ and } j \notin A+A)$:

$$\begin{aligned}\mathrm{P}(A : i \text{ and } j \in A+A) &= 1 - \mathrm{P}(A : i \notin A+A) - \mathrm{P}(A : j \notin A+A)\\ &\quad + \mathrm{P}(A : i \text{ and } j \notin A+A).\end{aligned}$$

Example: $\mathrm{P}(A : 3 \text{ and } 7 \notin A+A)$

Problem: Dependent Random Variables

It is sufficient to study $\mathrm{P}(A : i \text{ and } j \notin A+A)$:

$$\begin{aligned}\mathrm{P}(A : i \text{ and } j \in A+A) &= 1 - \mathrm{P}(A : i \notin A+A) - \mathrm{P}(A : j \notin A+A) \\ &\quad + \mathrm{P}(A : i \text{ and } j \notin A+A).\end{aligned}$$

Example: $\mathrm{P}(A : 3 \text{ and } 7 \notin A+A)$

- Conditions:

$$i = 3: \quad 0 \text{ or } 3 \notin A \text{ and } 1 \text{ or } 2 \notin A \qquad j = 7: \quad 0 \text{ or } 7 \notin A \text{ and } 1 \text{ or } 6 \notin A \text{ and } 2 \text{ or } 5 \notin A \text{ and } 3 \text{ or } 4 \notin A.$$

Problem: Dependent Random Variables

It is sufficient to study $\mathrm{P}(A : i \text{ and } j \notin A+A)$:

$$\begin{aligned}\mathrm{P}(A : i \text{ and } j \in A+A) &= 1 - \mathrm{P}(A : i \notin A+A) - \mathrm{P}(A : j \notin A+A) \\ &\quad + \mathrm{P}(A : i \text{ and } j \notin A+A).\end{aligned}$$

Example: $\mathrm{P}(A : 3 \text{ and } 7 \notin A+A)$

- Conditions:

$i = 3$: 0 or $3 \notin A$ and 1 or $2 \notin A$

$j = 7$: 0 or $7 \notin A$ and 1 or $6 \notin A$ and 2 or $5 \notin A$ and 3 or $4 \notin A$.

- Since there are common integers in both lists, the events $3 \notin A+A$ and $7 \notin A+A$ are dependent.

Solution: Use Graphs!

- Transform the conditions into a graph!

Solution: Use Graphs!

- Transform the conditions into a graph!
- For each integers in $[0, 7]$, add a vertex with that integer.

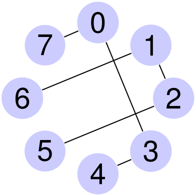
Solution: Use Graphs!

- Transform the conditions into a graph!
- For each integers in $[0, 7]$, add a vertex with that integer.
- Then connect two vertices if add up to 3 or 7.

Solution: Use Graphs!

- Transform the conditions into a graph!
- For each integers in $[0, 7]$, add a vertex with that integer.
- Then connect two vertices if add up to 3 or 7.

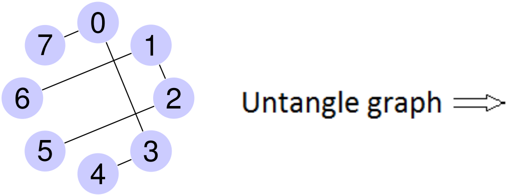
Example $i = 3, j = 7$:



Solution: Use Graphs!

- Transform the conditions into a graph!
- For each integers in $[0, 7]$, add a vertex with that integer.
- Then connect two vertices if add up to 3 or 7.

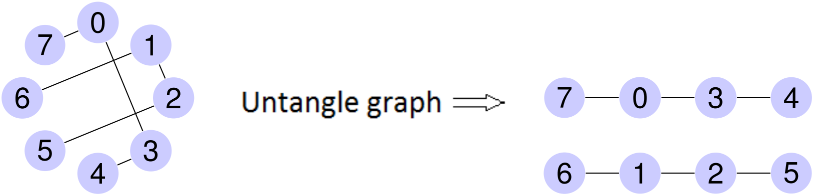
Example $i = 3, j = 7$:



Solution: Use Graphs!

- Transform the conditions into a graph!
- For each integers in $[0, 7]$, add a vertex with that integer.
- Then connect two vertices if add up to 3 or 7.

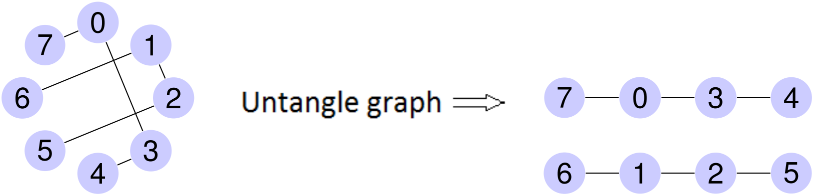
Example $i = 3, j = 7$:



Solution: Use Graphs!

- Transform the conditions into a graph!
- For each integers in $[0, 7]$, add a vertex with that integer.
- Then connect two vertices if add up to 3 or 7.

Example $i = 3, j = 7$:



- One-to-one correspondence between conditions/edges (and integers/vertices).

Interpretation of Graphs

Transformed into:

Interpretation of Graphs

Transformed into:

- Need to pick integers so that each condition is satisfied.

Interpretation of Graphs

Transformed into:

- Need to pick integers so that each condition is satisfied.
- Therefore, need to pick vertices so that each edge has a vertex chosen.

Interpretation of Graphs

Transformed into:

- Need to pick integers so that each condition is satisfied.
- Therefore, need to pick vertices so that each edge has a vertex chosen.
- So need to pick a vertex cover!

Interpretation of Graphs

Transformed into:

- Need to pick integers so that each condition is satisfied.
- Therefore, need to pick vertices so that each edge has a vertex chosen.
- So need to pick a vertex cover!
- Each integer is equally likely so can ignore vertex labels!

Vertex Covers

Have:

Vertex Covers

Have:

Example:

$7, 0, 4$ and $6, 2$ form a vertex cover

Vertex Covers

Have:

Example:

$7, 0, 4$ and $6, 2$ form a vertex cover

$\Longleftrightarrow$

If $7, 0, 4, 6, 2 \notin A$, then $3, 7 \notin A + A$

Vertex Covers

Have:

Example:

$7, 0, 4$ and $6, 2$ form a vertex cover

$\iff$

If $7, 0, 4, 6, 2 \notin A$, then $3, 7 \notin A + A$

Lemma(Lazarev, 2011)

$$P(i, j \notin A + A) = P(\text{pick a vertex cover for graph})$$

Number of Vertex Covers

Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

Number of Vertex Covers

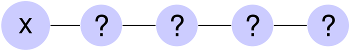
Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

- **Case 1**: If the first vertex is chosen:

Number of Vertex Covers

Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

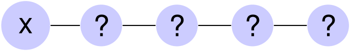
- **Case 1**: If the first vertex is chosen:



Number of Vertex Covers

Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

- **Case 1**: If the first vertex is chosen:

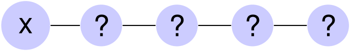


Need an vertex cover for the rest of the graph: $g(n-1)$.

Number of Vertex Covers

Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

- **Case 1**: If the first vertex is chosen:

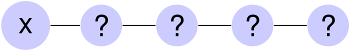


  Need an vertex cover for the rest of the graph: $g(n-1)$.
- **Case 2**: If the first vertex is not chosen:

Number of Vertex Covers

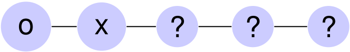
Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

- **Case 1**: If the first vertex is chosen:



  Need an vertex cover for the rest of the graph: $g(n-1)$.

- **Case 2**: If the first vertex is not chosen:

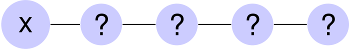


  Need an vertex cover for the rest of the graph: $g(n-2)$.

Number of Vertex Covers

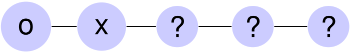
Condition graphs are always ‘segment’ graphs. So we just need $g(n)$, the number of vertex covers for a ‘segment’ graph with n vertices.

- **Case 1**: If the first vertex is chosen:



  Need an vertex cover for the rest of the graph: $g(n-1)$.

- **Case 2**: If the first vertex is not chosen:



  Need an vertex cover for the rest of the graph: $g(n-2)$.

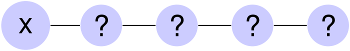
- Fibonacci recursive relationship!

$$g(n) = g(n-1) + g(n-2)$$

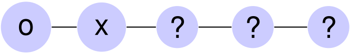
Number of Vertex Covers

Condition graphs are always 'segment' graphs. So we just need $g(n)$, the number of vertex covers for a 'segment' graph with n vertices.

- **Case 1**: If the first vertex is chosen:



  Need an vertex cover for the rest of the graph: $g(n-1)$.
- **Case 2**: If the first vertex is not chosen:



  Need an vertex cover for the rest of the graph: $g(n-2)$.
- Fibonacci recursive relationship!

$$g(n) = g(n-1) + g(n-2)$$
$$\Longrightarrow g(n) = F_{n+2}$$

General i, j

In particular

$$P(3 \text{ and } 7 \notin A + A) = \frac{1}{2^8} F_{4+2} F_{4+2} = \frac{1}{4}$$

since there were two graphs each of length 4.

General i, j

In particular

$$P(3 \text{ and } 7 \notin A + A) = \frac{1}{2^8} F_{4+2} F_{4+2} = \frac{1}{4}$$

since there were two graphs each of length 4.

- In general, graphs will have many components with different lengths.

General i, j

In particular

$$\mathrm{P}(3 \text{ and } 7 \notin A + A) = \frac{1}{2^8} F_{4+2} F_{4+2} = \frac{1}{4}$$

since there were two graphs each of length 4.

- In general, graphs will have many components with different lengths.
- For odd $i < j < n$:

$$P(A : i \text{ and } j \notin A + A)$$
$$= \frac{1}{2^{j+1}} F_{2\left\lceil \frac{i+1}{j-i} \right\rceil + 2}^{\frac{1}{2}\left((j-i)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1)\right)} \times F_{2\left\lceil \frac{i+1}{j-i} \right\rceil + 4}^{\frac{1}{2}\left(j+1-(j-i)\left\lceil \frac{i+1}{j-i} \right\rceil\right)}$$

General i, j

In particular

$$\mathrm{P}(3 \text{ and } 7 \notin A+A) = \frac{1}{2^8} F_{4+2} F_{4+2} = \frac{1}{4}$$

since there were two graphs each of length 4.

- In general, graphs will have many components with different lengths.
- For odd $i < j < n$:

$$P(A : i \text{ and } j \notin A+A) = \frac{1}{2^{j+1}} F_{2\left\lceil \frac{i+1}{j-i} \right\rceil + 2}^{\frac{1}{2}\left((j-i)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1)\right)} \times F_{2\left\lceil \frac{i+1}{j-i} \right\rceil + 4}^{\frac{1}{2}\left(j+1-(j-i)\left\lceil \frac{i+1}{j-i} \right\rceil\right)}$$

- Need to find the formulas for when i, j are even....

General i, j

In particular

$$\mathrm{P}(3 \text{ and } 7 \notin A+A) = \frac{1}{2^8} F_{4+2} F_{4+2} = \frac{1}{4}$$

since there were two graphs each of length 4.

- In general, graphs will have many components with different lengths.
- For odd $i < j < n$:

$$P(A : i \text{ and } j \notin A+A) = \frac{1}{2^{j+1}} F_{2\left\lceil \frac{i+1}{j-i} \right\rceil + 2}^{\frac{1}{2}\left((j-i)\left\lceil \frac{i+1}{j-i} \right\rceil - (i+1)\right)} \times F_{2\left\lceil \frac{i+1}{j-i} \right\rceil + 4}^{\frac{1}{2}\left(j+1-(j-i)\left\lceil \frac{i+1}{j-i} \right\rceil\right)}$$

- Need to find the formulas for when i, j are even....
- And to get variance, need to sum up over all $i < j < 2n$....

Variance Formula

$$Var|A+A| = -40 + 4\sum_{i<j<n} P(i,j \notin A+A)$$

$$= -40 + O(c^n)$$

$$+4\sum_{i,j\ odd} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+1-(j-i)\left\lceil\frac{i+1}{j-i}\right\rceil\right)}$$

$$+4\sum_{i\ even,j\ odd} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{i/2+1}{j-i}\right\rceil+1} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)+2\left\lceil\frac{i/2+1}{j-i}\right\rceil-1\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+1-(j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil-2\left\lceil\frac{i/2+1}{j-i}\right\rceil\right)}$$

$$+4\sum_{i\ odd,j\ even} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{j/2+1}{j-i}\right\rceil} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)+2\left\lceil\frac{j/2+1}{j-i}\right\rceil-2\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+2-(j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil+2\left\lceil\frac{j/2+1}{j-i}\right\rceil\right)}$$

$$+4\sum_{i,j\ even} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{i/2+1}{j-i}\right\rceil+1} F_{2\left\lceil\frac{j+2}{j-i}\right\rceil} \times$$

$$F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i-2)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)+2\left\lceil\frac{i/2+1}{j-i}\right\rceil+2\left\lceil\frac{j/2+1}{j-i}\right\rceil-3\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+2-(j-i-2)\left\lceil\frac{i+1}{j-i}+2\left\lceil\frac{i/2+1}{j-i}\right\rceil+2\left\lceil\frac{j/2+1}{j-i}\right\rceil\right\rceil\right)}$$

Variance Formula

$$Var|A+A| = -40 + 4\sum_{i<j<n} P(i,j \notin A+A)$$

$$= -40 + O(c^n)$$

$$+4\sum_{i,j\ odd} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+1-(j-i)\left\lceil\frac{i+1}{j-i}\right\rceil\right)}$$

$$+4\sum_{i\ even,j\ odd} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{i/2+1}{j-i}\right\rceil+1} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)+2\left\lceil\frac{i/2+1}{j-i}\right\rceil-1\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+1-(j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil-2\left\lceil\frac{i/2+1}{j-i}\right\rceil\right)}$$

$$+4\sum_{i\ odd,j\ even} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{j/2+1}{j-i}\right\rceil} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)+2\left\lceil\frac{j/2+1}{j-i}\right\rceil-2\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+2-(j-i-1)\left\lceil\frac{i+1}{j-i}\right\rceil+2\left\lceil\frac{j/2+1}{j-i}\right\rceil\right)}$$

$$+4\sum_{i,j\ even} \frac{1}{2^{j+1}} F_{2\left\lceil\frac{i/2+1}{j-i}\right\rceil+1} F_{2\left\lceil\frac{j+2}{j-i}\right\rceil} \times$$

$$F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+2}^{\frac{1}{2}\left((j-i-2)\left\lceil\frac{i+1}{j-i}\right\rceil-(i+1)+2\left\lceil\frac{i/2+1}{j-i}\right\rceil+2\left\lceil\frac{j/2+1}{j-i}\right\rceil-3\right)} F_{2\left\lceil\frac{i+1}{j-i}\right\rceil+4}^{\frac{1}{2}\left(j+2-(j-i-2)\left\lceil\frac{i+1}{j-i}+2\left\lceil\frac{i/2+1}{j-i}\right\rceil+2\left\lceil\frac{j/2+1}{j-i}\right\rceil\right\rceil\right\rceil\right)}$$

So clearly

$$Var|A+A| \sim 35.98$$

Consecutive Missing Sums

Improved Upper Bound

- Use $P(i, j \notin A + A)$ to improve the upper bound for $P(A + A$ has k missing sums $)$

Improved Upper Bound

- Use $P(i, j \notin A + A)$ to improve the upper bound for $P(A + A$ has k missing sums)
- From previous section:

$$P(i, j \notin A + A) \sim \left(\frac{\phi^2}{2 \cdot 5^{1/4}}\right)^j \left(\frac{5^{1/4}}{\phi}\right)^i$$

where ϕ is the golden ratio

Improved Upper Bound

- Use $P(i, j \notin A + A)$ to improve the upper bound for $P(A + A \text{ has } k \text{ missing sums })$
- From previous section:

$$P(i, j \notin A + A) \sim \left(\frac{\phi^2}{2 \cdot 5^{1/4}}\right)^j \left(\frac{5^{1/4}}{\phi}\right)^i$$

where ϕ is the golden ratio
- If $A + A$ is missing at k elements, then missing at least two elements greater than k:

$$P(A{+}A \text{ has } k \text{ missing sums }) < \left(\frac{\phi^2}{2 \cdot 5^{1/4}}\right)^k \left(\frac{5^{1/4}}{\phi}\right)^k \sim 0.80^k$$

Improved Upper Bound

- Use $P(i, j \notin A + A)$ to improve the upper bound for $P(A + A \text{ has } k \text{ missing sums })$
- From previous section:

$$P(i, j \notin A + A) \sim \left(\frac{\phi^2}{2 \cdot 5^{1/4}}\right)^j \left(\frac{5^{1/4}}{\phi}\right)^i$$

where ϕ is the golden ratio

- If $A + A$ is missing at k elements, then missing at least two elements greater than k:

$$P(A{+}A \text{ has } k \text{ missing sums }) < \left(\frac{\phi^2}{2 \cdot 5^{1/4}}\right)^k \left(\frac{5^{1/4}}{\phi}\right)^k \sim 0.80^k$$

- General idea: The better we know $P(a_1, \cdots, a_j \notin A + A)$, the better we know the $P(A + A \text{ has } k \text{ missing sums })$

Consecutive Missing Sums in A+A

- Will study the particular case of $P(a_1, \cdots, a_j \notin A + A)$ of consecutive missing sums: $P(k, k+1, \cdots, k+i \notin A + A)$

Consecutive Missing Sums in A+A

- Will study the particular case of $P(a_1, \cdots, a_j \notin A+A)$ of consecutive missing sums: $P(k, k+1, \cdots, k+i \notin A+A)$
- Example: $P(16, 17, 18, 19, 20 \notin A+A)$

Consecutive Missing Sums in A+A

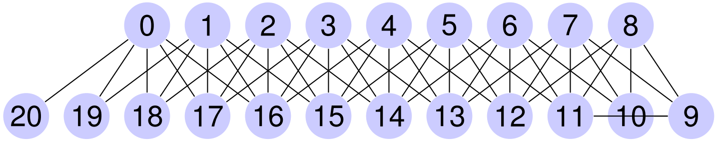
- Will study the particular case of $P(a_1, \cdots, a_j \notin A + A)$ of consecutive missing sums: $P(k, k+1, \cdots, k+i \notin A + A)$
- Example: $P(16, 17, 18, 19, 20 \notin A + A)$

Start with original graph and remove some conditions (edges):

Consecutive Missing Sums in A+A

- Will study the particular case of $P(a_1, \cdots, a_j \notin A + A)$ of consecutive missing sums: $P(k, k+1, \cdots, k+i \notin A + A)$
- Example: $P(16, 17, 18, 19, 20 \notin A + A)$

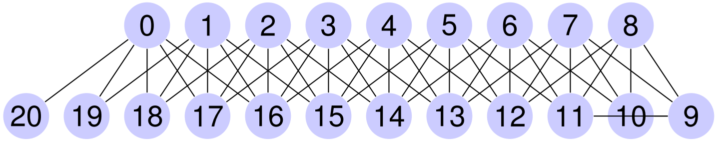
Start with original graph and remove some conditions (edges):



Consecutive Missing Sums in A+A

- Will study the particular case of $P(a_1, \cdots, a_j \notin A + A)$ of consecutive missing sums: $P(k, k+1, \cdots, k+i \notin A + A)$
- Example: $P(16, 17, 18, 19, 20 \notin A + A)$

Start with original graph and remove some conditions (edges):

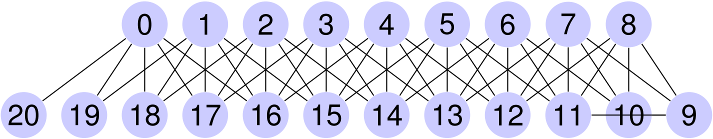


$\Longrightarrow$ Transforms to:

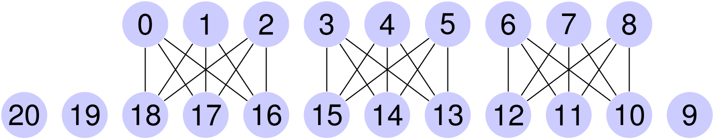
Consecutive Missing Sums in A+A

- Will study the particular case of $P(a_1, \cdots, a_j \notin A + A)$ of consecutive missing sums: $P(k, k+1, \cdots, k+i \notin A + A)$
- Example: $P(16, 17, 18, 19, 20 \notin A + A)$

Start with original graph and remove some conditions (edges):

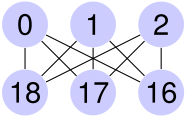


$\Longrightarrow$ Transforms to:



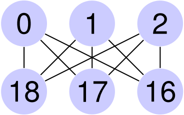
Consecutive Missing Sums in A+A

- So get around $20/6$ complete bipartite graphs like:



Consecutive Missing Sums in A+A

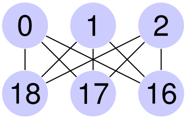
- So get around $20/6$ complete bipartite graphs like:



- For a vertex cover, need one side to have all vertices chosen; occurs with probability $\leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$

Consecutive Missing Sums in A+A

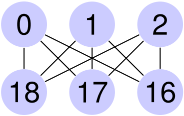
- So get around $20/6$ complete bipartite graphs like:



- For a vertex cover, need one side to have all vertices chosen; occurs with probability $\leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$
- By independence, $P(16, 17, 18, 19, 20) \leq \left(\frac{1}{4}\right)^{20/6}$

Consecutive Missing Sums in A+A

- So get around $20/6$ complete bipartite graphs like:



- For a vertex cover, need one side to have all vertices chosen; occurs with probability $\leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$
- By independence, $P(16, 17, 18, 19, 20) \leq \left(\frac{1}{4}\right)^{20/6}$
- In general $P(k, k+1, k+2, k+3, k+4) \leq \left(\frac{1}{4}\right)^{(k+4)/6} \sim 0.79^{k+4}$ which is a slight improvement!

Consecutive Missing Sums

- But most general is:
  $P(k, k+1, \cdots, k+i \notin A+A) \leq \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$

Consecutive Missing Sums

- But most general is:
  $P(k, k+1, \cdots, k+i \notin A+A) \leq \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$
- However the trivial lower bound is:
  $\left(\frac{1}{2}\right)^{(k+i)/2} \leq P(k, k+1, \cdots, k+i \notin A+A)$

Consecutive Missing Sums

- But most general is:
  $P(k, k+1, \cdots, k+i \notin A+A) \leq \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$
- However the trivial lower bound is:
  $\left(\frac{1}{2}\right)^{(k+i)/2} \leq P(k, k+1, \cdots, k+i \notin A+A)$
- Why interesting? Bounds almost match!

Consecutive Missing Sums

- But most general is:
  $P(k, k+1, \cdots, k+i \notin A+A) \leq \left(\frac{1}{2}\right)^{(k+i)/2} (1+\epsilon_i)^k$
- However the trivial lower bound is:
  $\left(\frac{1}{2}\right)^{(k+i)/2} \leq P(k, k+1, \cdots, k+i \notin A+A)$
- Why interesting? Bounds almost match!
- Essentially the only way to miss a block of i consecutive sums is to miss all elements before the block as well.

Summary

We found

Summary

We found

- Exponential bounds for distribution of number of missing sums

Summary

We found

- Exponential bounds for distribution of number of missing sums
- Series for the variance in terms of Fibonacci numbers

Summary

We found

- Exponential bounds for distribution of number of missing sums
- Series for the variance in terms of Fibonacci numbers
- Asymptotic formula for consecutive missing sums

Summary

We found

- Exponential bounds for distribution of number of missing sums
- Series for the variance in terms of Fibonacci numbers
- Asymptotic formula for consecutive missing sums

by developing graph theoretic framework for studying dependent random variables.

Future Research

And currently investigating/other research ideas:

Future Research

And currently investigating/other research ideas:

- Is distribution of missing sums approximately exponential?

Future Research

And currently investigating/other research ideas:

- Is distribution of missing sums approximately exponential?
- Is $P(k + a_1, \cdots, k + a_m) \sim \lambda^k$ for any fixed $a_1, \cdots, a_m$?

Future Research

And currently investigating/other research ideas:

- Is distribution of missing sums approximately exponential?
- Is $P(k+a_1,\cdots,k+a_m) \sim \lambda^k$ for any fixed $a_1,\cdots,a_m$?
- Higher moments: third moment involves $P(i,j,k)$ which has more complicated graphs

Future Research

And currently investigating/other research ideas:

- Is distribution of missing sums approximately exponential?
- Is $P(k + a_1, \cdots, k + a_m) \sim \lambda^k$ for any fixed $a_1, \cdots, a_m$?
- Higher moments: third moment involves $P(i, j, k)$ which has more complicated graphs
- Distribution of missing differences in $A - A = \{x - y : x, y \in A\}$

Thank you!