Distribution of Missing Differences in Diffsets

Scott Harvey-Arnold sharveyarnold@gmail.com
Steven Miller sjm1@williams.edu
Fei Peng fpeng1@andrew.cmu.edu

CANT 2020, June 5th

The problem

- Define [n] as $\{0, 1, \ldots, n-1\}$.

- Given a set S ⊆ [n], we can define its sumset and diffset
  - $S + S := \{x + y : x, y \in S\}, \ S - S := \{x - y : x, y \in S\}$.

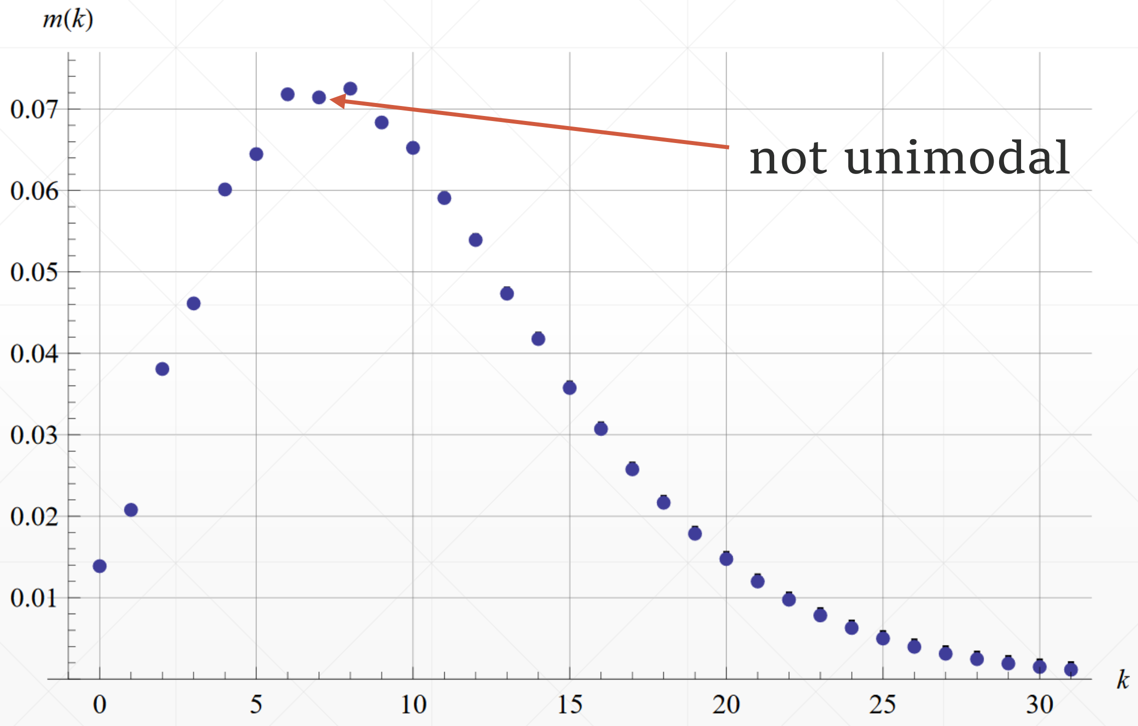
- Q: What is the typical size of S+S and S-S?
- (Observe: both sizes are at most 2n-1.)

Related work

- Q: What is the typical size of S+S and S-S? ($S \subseteq [n]$ uniformly random)
- Martin and O'Bryant (2006): When $n \to \infty$, the expected number of missing sums goes to 10 ($\lim_{n\to\infty} \mathbb{E}[2n-1-|S+S|] = 10$), and missing differences to 6.
- Zhao (2009): The "limiting probabilities" of missing k sums (differences) exist and sum to 1.

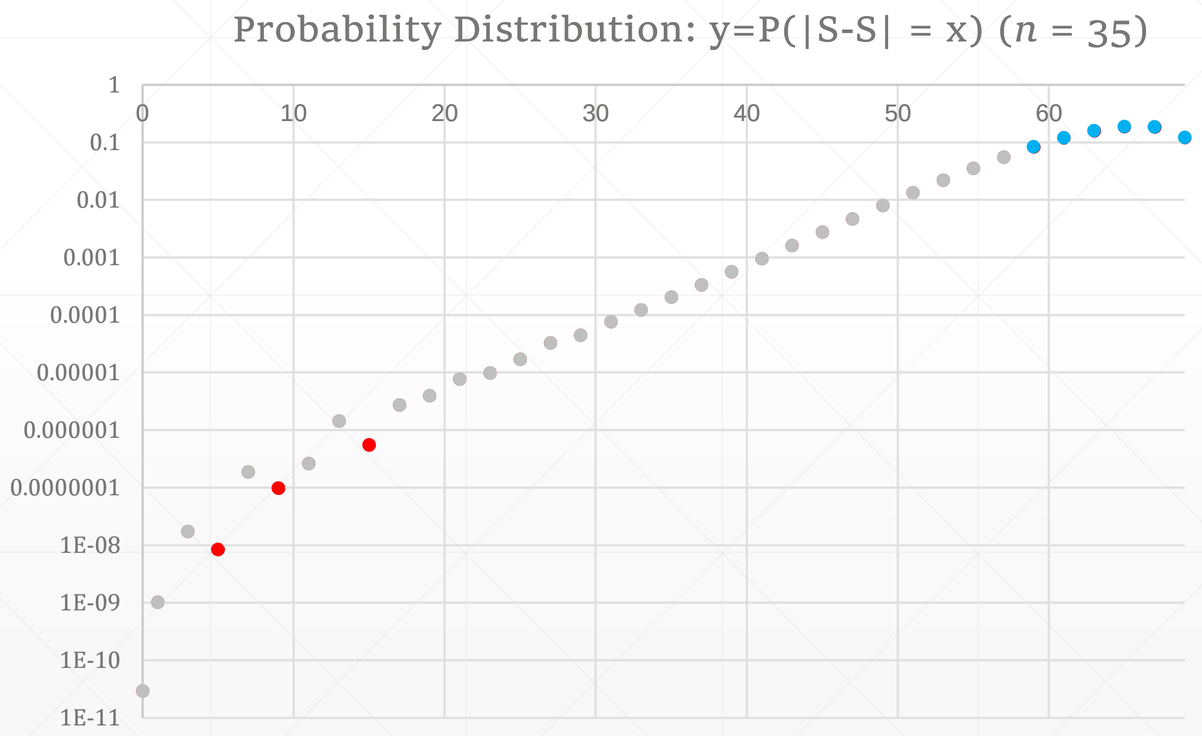
Ok.. how are they distributed?

- Lazarev, Miller and O’Bryant (2012): missing sums distribute like



What we can say about differences

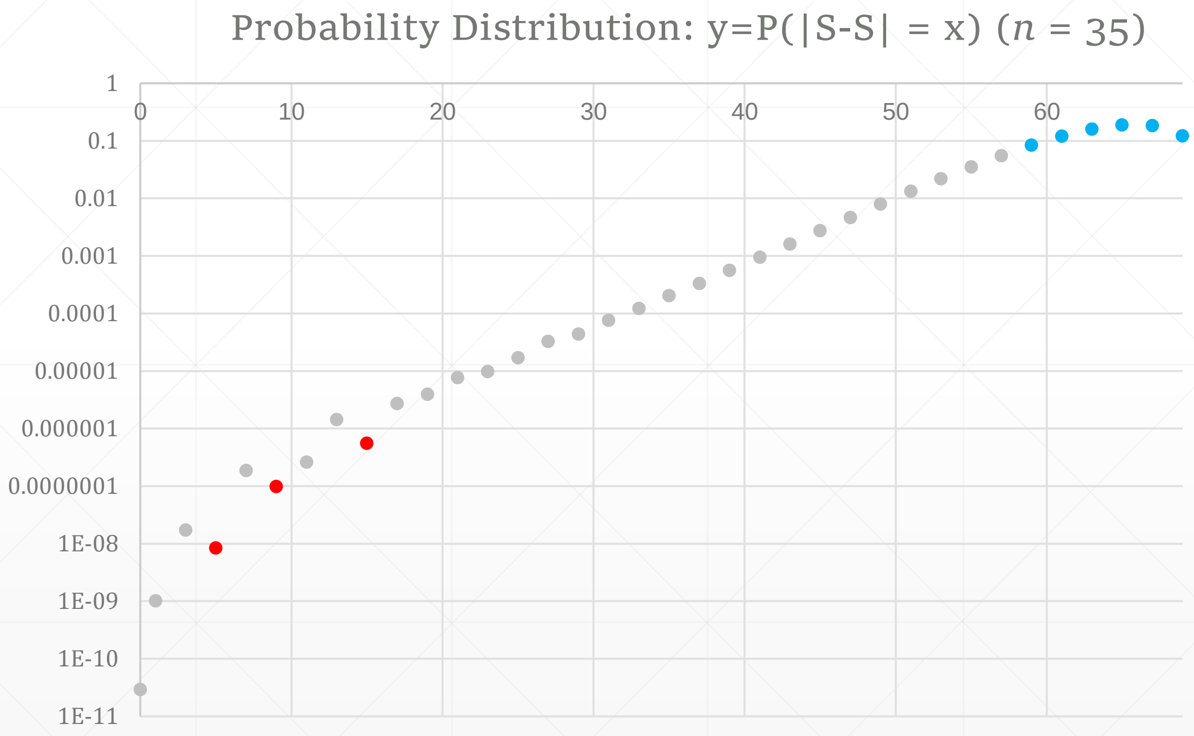
- Look at the distribution of |S-S| for $n = 35$



Let’s define the limiting probabilities

- $\ell(k) \coloneqq \lim_{n\to\infty} \mathbb{P}(2n-1-|S-S|=k)$

- Snapshot at n=35:
- $\mathbb{P}[2n-1-|S-S|=0] \approx 0.12132$
- $\mathbb{P}[2n-1-|S-S|=2] \approx 0.18424$
- $\mathbb{P}[2n-1-|S-S|=4] \approx 0.18755$
- $\mathbb{P}[2n-1-|S-S|=6] \approx 0.15825$
- $\mathbb{P}[2n-1-|S-S|=8] \approx 0.11945$
- $\mathbb{P}[2n-1-|S-S|=10] \approx 0.08362$

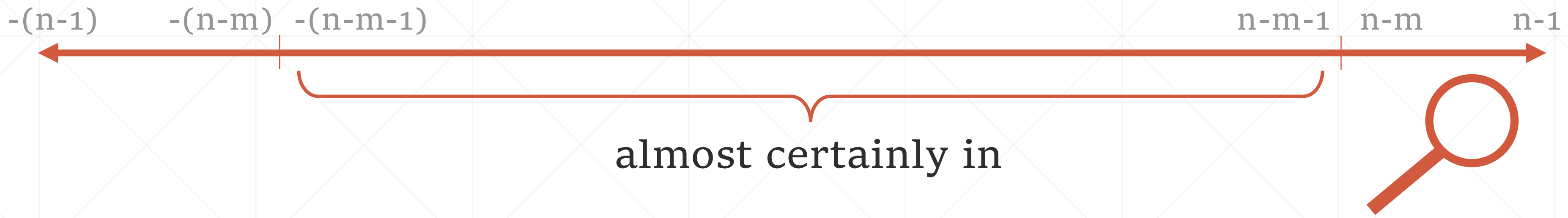


Towards a rigorous bound

- The possible differences are –n+1, -n+2, …, 0, …, n-2, n-1.
- Martin and O'Bryant: $\mathbb{P}(k \notin S - S) \leq \begin{cases} 0.75^{\frac{n}{3}} & \left(1 \leq k \leq \frac{n}{2}\right) \\ 0.75^{n-k} & \left(\frac{n}{2} \leq k \leq n-1\right) \end{cases}$
- Numbers close to zero are more likely to be in the diffset
- Union bound → $\mathbb{P}(\{-(n-m-1), \dots, n-m-1\} \nsubseteq S - S) < 4 \cdot 0.75^{m+1} + o(1)_{n \to \infty}$
  - So *most of the times*, the middle part is entirely in

We are close

- $\mathbb{P}(\{-(n-m-1), \dots, n-m-1\} \not\subseteq S-S) < 4 \cdot 0.75^{m+1} + o(1)_{n\to\infty}$



- Investigate the distribution of $|\{n-m, \dots, n-1\} \cap (S-S)|$
- We only care about the first and the last m numbers in [n]
- Use finite computing™ to make the error arbitrarily small

We are far…

- Simulating 2^{2m} choices, to reduce the error to $4 \cdot 0.75^{m+1}$.
- Wanted to show $\ell(2) < \ell(4) > \ell(6)$
  - $\mathbb{P}[2n - 1 - |S - S| = 2] \approx 0.18424$
  - $\mathbb{P}[2n - 1 - |S - S| = 4] \approx 0.18755$
  - $\mathbb{P}[2n - 1 - |S - S| = 6] \approx 0.15825$
- Need the error to be about 0.0016. That needs some $m \geq 27$, which implies at least 2^{54} (1.8×10^{16}) sets to loop through.
- 25.2 years ☹

That’s conditional

- $j(k) \coloneqq \lim_{n\to\infty} \mathbb{P}(2n-1-|S-S|=k \mid 0, n-1 \in S)$
- We can write $\ell(k)$ in terms of $j(k)$ (and vice versa):
- $\ell(0) = \frac{1}{4} j(0),$ $\ell(2) = \frac{1}{4} j(2) + \frac{2}{8} j(0),$
- $\ell(4) = \frac{1}{4} j(4) + \frac{2}{8} j(2) + \frac{3}{16} j(0),$ $\ell(6) = \frac{1}{4} j(6) + \frac{2}{8} j(4) + \frac{3}{16} j(2) + \frac{4}{32} j(0)\ldots$

- **Corollary.** It would suffice to show that $j(4) > \frac{j(0)}{4}$ and $j(6) < \frac{j(0)+j(2)}{4}$.

Why j?

- Comparing $j(k)$'s can tolerate larger error (than $\ell(k)$)
- $j(k)$'s already produce less error (middle more likely to be in)
- $j(k)$ only sums over ¼ the sets (thx to the conditional probability)

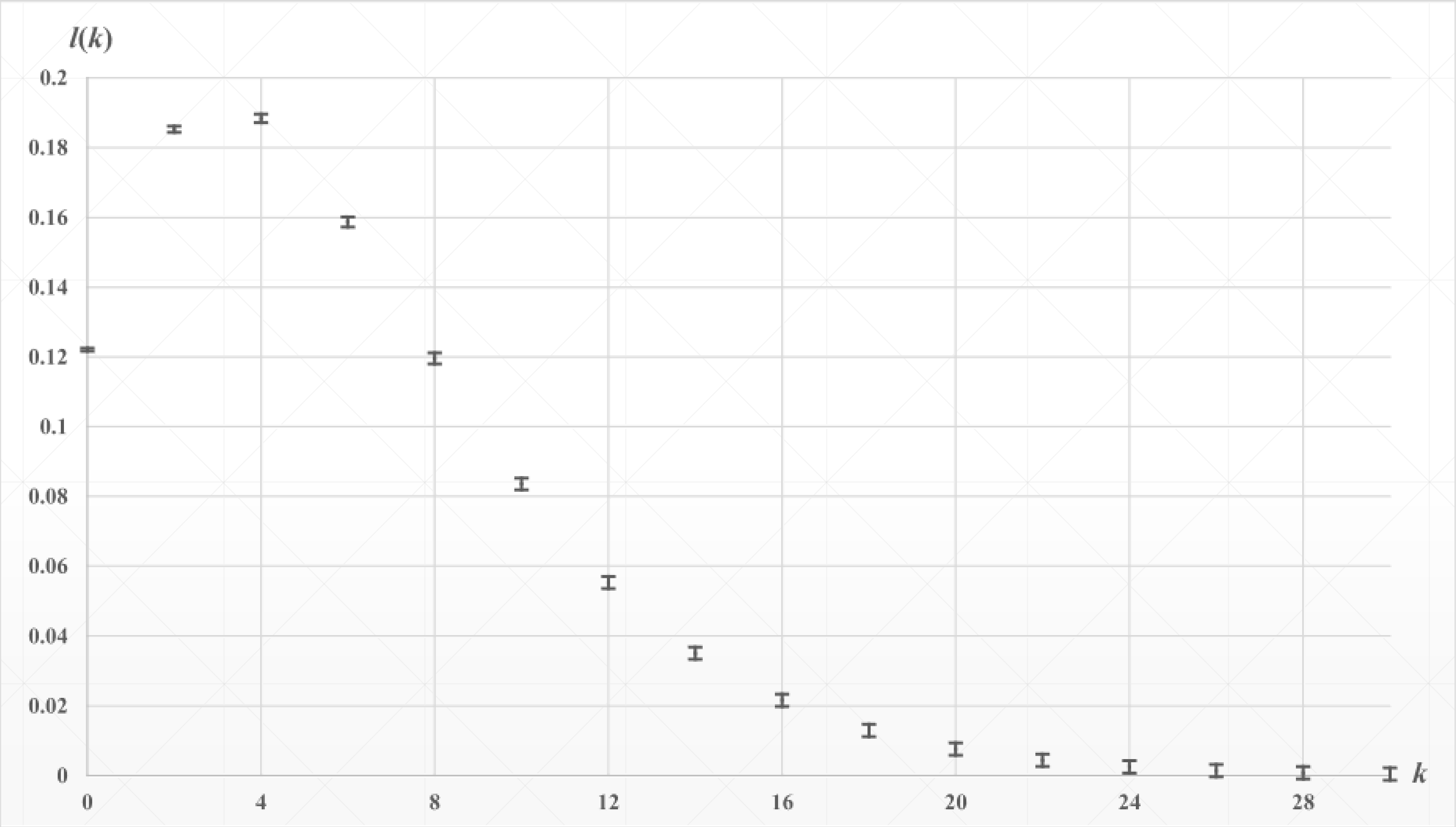
The $j(k)$ approach is [at least] 1,527,656 times faster than $\ell(k)$. ☺

Results and Conjectures

- We burned more computational power than needed, and were able to prove that

$$\ell(4) > \ell(2) > \ell(6) > \ell(0) > \ell(8) > \ell(10) > \cdots > \ell(20).$$

- It seems “obvious” that $\ell(20) > \ell(22) > \ell(24) > \cdots$, although we couldn’t prove it.


l(k)
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
4
8
12
16
20
24
28
k

Rulers

- Set of integer marks
- Complete if there's no "gap" of measurable differences
- E.g. {0, 1, 4, 6}
- One application of our results is an asymptotic bound for the number of complete rulers (basically sets with size n that miss no difference):

$$\mathrm{A103295}(n) \sim c \cdot 2^n, \text{ where } 0.2433 < c < 0.2451.$$

Of course, $c = \frac{\ell(0)}{2}$.

A remark on sums vs. differences

- It's believed that dealing with |S-S| is much harder than |S+S|:

- When the fringe has width m…
  - Diffsets: had to consider both the first and the last m numbers in [n]
  - Sumsets: could consider the two parts independently

Result: diffsets have computational complexity *squared!*

Is this entirely true?

- $\mathbb{P}(n - m \notin S - S) = 0.75^m$: the pairs are {0, n-m-1}, …, {m-1, n-1}
- $\mathbb{P}(m \notin S + S) \approx 0.75^{m/2}$: the pairs are {0, m}, {1, m-1}, …, {m/2, m/2}
- To reach the same precision, you would need the fringe to be twice the size as for the diffset, so it squared out

Bibliography

- P. Erdös and A. Rényi, *Additive properties of random sequences of positive integers*, 1960.
- O. Lazarev, S. J. Miller and K. O'Bryant, *Distribution of Missing Sums in Sumsets*, Experimental Mathematics 22 (2013), no. 2, 132–156.
- G. Martin and K. O'Bryant, *Many sets have more sums than differences*, Additive Combinatorics, Providence, RI, 2007, 287–305.
- M. B. Nathanson, *Sets with more sums than differences*, Integers 7 (2007), #A5.
- Y. Zhao, *Sets characterized by missing sums and differences*, J. Number Theory 131(2011), 2107–2134.

- We want to thank Joshua Siktar for his constructive comments.
- Credits to the CMU AFS system (which allowed time-consuming codes).