



Generating MSTD sets through Markov processes

Frank He (University of Cambridge), Karol Daniewski (Williams College)

Contact: fsh30@cam.ac.uk, kkd2@williams.edu

Advised by Steven J. Miller

SMALL REU, Williams College



Sum and Difference Sets

Definition

Let $A \subseteq S := \{0, \dots, n\}$. Define $A + A = \{a + a' \mid a, a' \in A\}$, and similarly $A - A = \{a - a' \mid a, a' \in A\}$. We call A

- ① *Sum dominant* if $|A + A| > |A - A|$
- ② *Difference dominant* if $|A + A| < |A - A|$
- ③ *Balanced* if $|A + A| = |A - A|$

Expect most sets to be *difference dominant* because addition commutes but subtraction doesn't. Still, many *sum dominant* sets occur.

Theorem (Martin, O'bryant)

Let $\rho_+^{(n)}, \rho_-^{(n)}, \rho_0^{(n)}$ be the proportion of subsets of $\{0, \dots, n\}$ that are sum dominant, difference dominant, and balanced respectively. As $n \rightarrow \infty$,

$$\liminf_{n \rightarrow \infty} \rho_\star^{(n)} > 0, \quad \text{for any } \star \in \{-, +, 0\}$$

Remark: This Theorem was extended by Zhao, who showed that

$$\rho_\star := \lim_{n \rightarrow \infty} \rho_\star^{(n)}$$

exists for each $\star \in \{-, +, 0\}$.

Remark: About 0.04% sets are sum dominant.

Expected Difference in Sizes

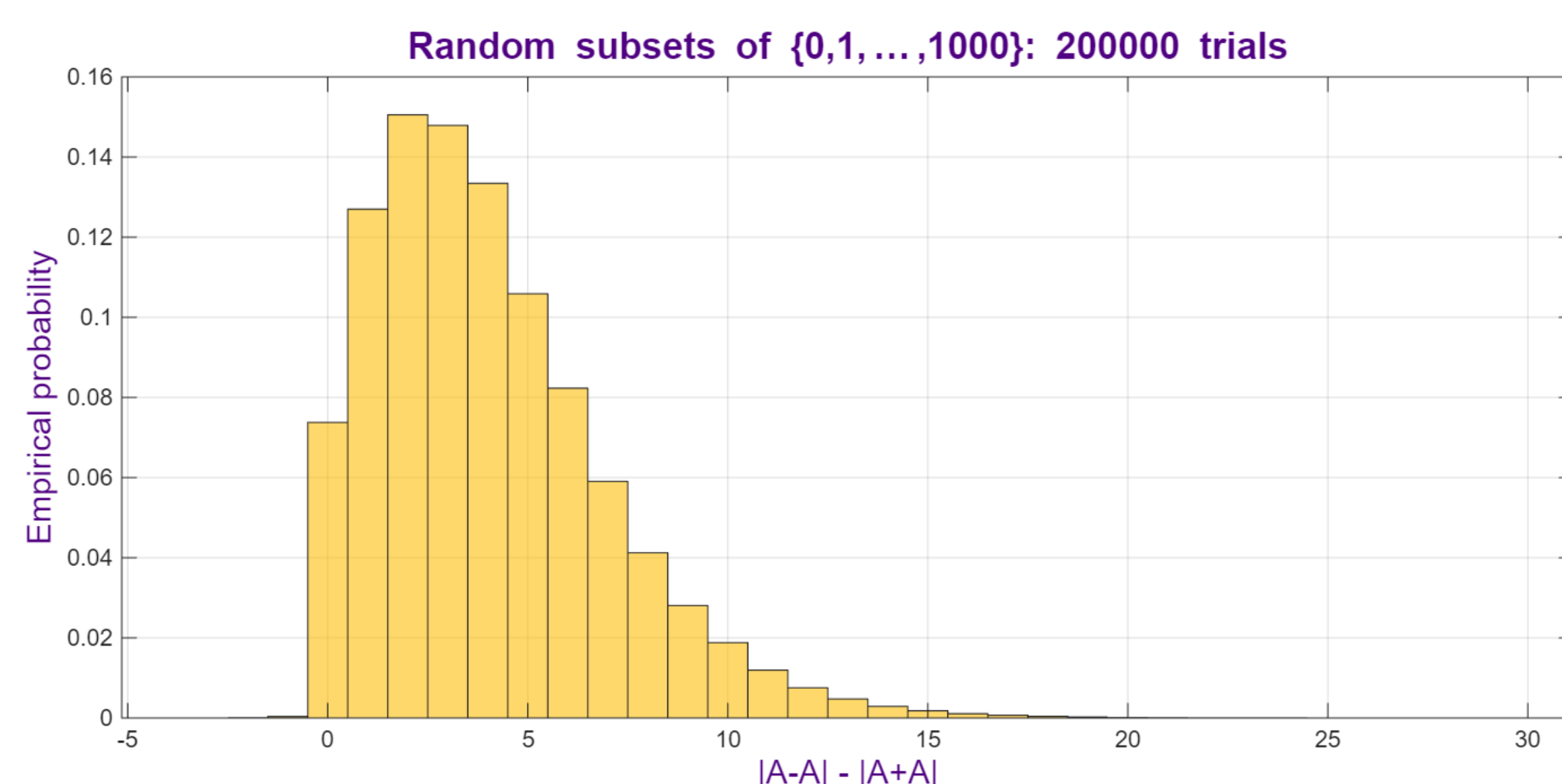


Figure 1: On average, difference sets have more terms.

Result (Martin, O'bryant): $\mathbb{E}[|A - A| - |A + A|] \rightarrow 4$ as $n \rightarrow \infty$.

Sets Generated by Markov Processes

We generalize the construction of $A \subseteq S$ by considering a Markov process that picks whether $i \in A$ depending on whether $i - 1$ is in A .

Definition

Let $(X_i)_{i=0}^n$ be a stationary two-state Markov chain with transition matrix

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}, \quad \text{where } p, q \in (0, 1)$$

and define

$$A = \{i \in \{0, \dots, n\} : X_i = 1\}$$

Remark: This process has stationary distribution

$$\pi = \left(\frac{1-q}{2-p-q}, \frac{1-p}{2-p-q} \right)$$

Many Markov Sets have More Sums than Differences

We can find the sizes of the sum and difference sets by only considering the *left* and *right fringes*, defined as:

$$L = A \cap [0, K] \quad R = (n - A) \cap [0, K]$$

We define a *fringe score*.

$$\Delta(L, R) = |(L + L) \cap [0, K]| + |(R + R) \cap [0, K]| - 2|(L + R) \cap [0, K]|$$

which counts the difference in set sizes if their middles are 'full'.

Lemma (Zhao): For fixed K , if $[K + 1, 2n - K - 1] \subseteq A + A$ and $[K + 1 - n, n - K - 1] \subseteq A - A$, then

$$\Delta = |A + A| - |A - A|$$

We prove that middles are almost surely full as $n \rightarrow \infty$, and by constructing non-negligible families of sets satisfying $\Delta \star 0$ for each $\star \in \{<, >, =\}$, we prove the following:

Theorem

Consider $\rho_+^{(n)}, \rho_-^{(n)}, \rho_0^{(n)}$ for the Markov case. As $n \rightarrow \infty$, all three have limits and

$$\lim_{n \rightarrow \infty} \rho_\star^{(n)} > 0, \quad \text{for any } \star \in \{-, +, 0\}$$

A General Formula for Expected Differences

To find the size of $A + A$ and $A - A$, it is easier to count the probabilities, for each k , of $k \notin A + A$ and $k \notin A - A$.

(Example) Even sums: Suppose $k = 2m$ is missing from $A + A$, so pairs $(m, m), (m - 1, m + 1), \dots, (0, 2m)$ must not be present in A . We count this probability using two Markov chains, $X_m, X_{m+1}, \dots, X_{2m}$ and a reversed chain $X_m, X_{m-1}, \dots, X_0$. We may then sum over all even k .

Similar counting arguments work for the odd sums case and the differences case. We arrive at the next result:

Theorem

Let $A \subseteq S := \{0, \dots, n\}$ be generated by Markov process $\{X_i\}$. As $n \rightarrow \infty$,

$$\mathbb{E}[|A - A| - |A + A|] \rightarrow \Lambda(p, q) = \frac{2(1-q)(1+p-p^2-pq)}{(1-p)^2(p+q)(2-p-q)}$$

Result: We also proved that this convergence occurs at rate

$$O(n\lambda^n), \quad \text{for some } \lambda \in (0, 1)$$

Selected References

- ① Hegarty, P. V. (2007). Some explicit constructions of sets with more sums than differences. *Acta Arithmetica*, 130(1), 61–77. <https://doi.org/10.4064/aa130-1-4>
- ② Martin, G., & O'Bryant, K. (2007). Many sets have more sums than differences. *In Additive Combinatorics* (CRM Proceedings & Lecture Notes, Vol. 43, pp. 287–305). American Mathematical Society.
- ③ Zhao, Y. (2011). Sets characterized by missing sums and differences. *Journal of Number Theory*, 131(11), 2107–2134. <https://doi.org/10.1016/j.jnt.2011.05.003>

Acknowledgements

We would like to thank our mentor Dr. Steven J. Miller and the SMALL program. This work was supported by NSF Grant DMS2241623, the Finnerty Fund, the Winston Churchill Foundation, and Williams College.