LOGARITHMIC REPRESENTABILITY OF INTEGERS AS K-SUMS

Anant Godbole

Department of Mathematics and Statistics, East Tennessee State University, Johnson City, Tennessee
godbolea@etsu.edu

Samuel Gutekunst

Department of Mathematics, Harvey Mudd College, Claremont, California

Vince Lyzinski

Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland

Yan Zhuang

Department of Mathematics, Brandeis University, Waltham, Massachusetts

Abstract

A set $\mathcal{A} = \mathcal{A}_{k,n} \subset [n] \cup \{0\}$ is said to be an additive k-basis if each element in $\{0, 1, \ldots, kn\}$ can be written as a k-sum of elements of $\mathcal{A}$ in *at least* one way. Seeking multiple representations as k-sums, and given any function $\phi(n) \to \infty$, we say that $\mathcal{A}$ is said to be a *truncated $\phi(n)$-representative k-basis* for $[n]$ if for each $j \in [\alpha n, (k - \alpha)n]$ the number of ways that j can be represented as a k-sum of elements of $\mathcal{A}_{k,n}$ is $\Theta(\phi(n))$. In this paper, we follow tradition and focus on the case $\phi(n) = \log n$, and show that a randomly selected set in an appropriate probability space is a truncated log-representative basis with probability that tends to one as $n \to \infty$. This result is a finite version of a result proved by Erdős and extended by Erdős and Tetali.

1. Introduction

In 1956 Erdős [4] answered a question posed in 1932 by Sidon by proving that there exists an infinite sequence of natural numbers $\mathcal{S}$ and constants c_1 and c_2 such that for large n,

$$c_1 \log n \le r_2(n) \le c_2 \log n, \tag{1}$$

where, for $k \geq 2$, $r_k(n)$ is the *number of ways* of representing the integer n as the sum of k distinct elements from $\mathcal{S}$, a so-called *log-representative basis* of order k. The result was generalized in the 1990 work of Erdős and Tetali [5] which established that there exists an infinite sequence $\mathcal{S}$ for which (1) was true for each fixed $k \geq 2$, i.e., for each large n,

$$r_k(n) = \Theta(\log n). \tag{2}$$

To achieve this result, Erdős and Tetali constructed a random sequence $\mathcal{S}$ of natural numbers by including z in $\mathcal{S}$ with probability

$$p(z) = \begin{cases} C\frac{(\log z)^{1/k}}{z^{(k-1)/k}}, & \text{if } z > z_0 \\ 0 & \text{otherwise} \end{cases}$$

where C is a determined constant and z_0 is the smallest constant such that $p(z_0) \leq 1/2$. They then showed that this random sequence is almost surely (a.s.) a log-representative basis of order k, with (2) holding a.s. for large n. We note here that a.s. in this context means "with probability one," i.e., in the sense used in measure theory.

For a natural finite variant of the above problem, we define:

Definition 1.1. With $[n] := \{1, 2, \ldots, n\}$, a set $\mathcal{A}_{k,n} \subseteq [n] \cup \{0\}$ is said to be a *log-representative k-basis for B* (or simply a representative k-basis for B) if each $j \in B \subset [kn] \cup \{0\}$ can be represented as a k-sum of distinct elements of $\mathcal{A}_{k,n}$ in $\Theta(\log n)$ ways.

Remark 1.2. To see how this is the natural finite variant of the problem tackled in Erdős and Tetali in [5], note that they teased out asymptotics for the emergence of log-representative bases for $B = [N_0, \infty)$ for some suitable N_0. They showed that in some probability space, almost all infinite sequences $\mathcal{S}$ satisfy (2). It is natural to then ask, for finite $\mathcal{A}_{k,n}$, how small $\mathcal{A}_{k,n}$ can be while still being a representative k-basis for a suitable B.

Remark 1.3. Note that a more general definition might ask that the number of representations of j equal $\Theta(\phi(n))$ for some $\phi(n) \to \infty$, but we stick close to tradition and just deal with the case $\phi(n) = \log n$; it is interesting to note, though, that Sidon's original question asked about whether it was possible to find a representative basis with $\phi(n) = o(n^\epsilon)$ for all $\epsilon > 0$.

We will use a probability model in which each integer in $[n] \cup \{0\}$ is chosen to be in $\mathcal{A}_{k,n}$ with equal (and low) probability $p = p_n$. Since, e.g., the only way to represent 1 as a 2-sum of elements of $[n] \cup \{0\}$ is as 1+0, it would be impossible for the random ensemble to form an representative basis unless we choose a target sumset, B, smaller than $[kn] \cup \{0\}$; this motivates the next definition – which was the one adopted in [7] even when we had $\phi(n) = 1$ for each n.

Definition 1.4. Let $\mathcal{A}_{k,n}$ be a subset of $[n] \cup \{0\}$. Then $\mathcal{A}_{k,n}$ is said to be a *truncated log-representative k-basis* for $[n]$ if each $j \in [\alpha n, (k-\alpha)n]$ can be represented as a k-sum of distinct elements of $\mathcal{A}_{k,n}$ in $\Theta(\log n)$ ways.

In [7], the authors used Poisson approximation (see [2] for background) and the Janson inequality [10] to derive a sharp threshold for which values of $p = p_n$ make the set $\mathcal{A}_{k,n}$ almost never/almost surely a truncated k-basis as $n \to \infty$, i.e. every $j \in [\alpha n, (k-\alpha)n]$ could be represented *at least once* as a k-sum of elements in $\mathcal{A}_{k,n}$ with probability tending to 0 or 1 as $n \to \infty$. Here the phrases "almost never" and "almost surely" are used as is traditional in random methods. The threshold function for k=2 (for example) is roughly at $p_n = A_\alpha \sqrt{\log n / n}$, with a third order correction term controlling the actual threshold, in contrast to the fact that the minimal size of a truncated 2-basis is of magnitude $O(\sqrt{n})$ [12], [9]. The corresponding questions of maximal Sidon families (i.e., ones for which each target integer is represented *at most once*), and zero-one thresholds for the emergence of the Sidon property feature a wider gap; see [6].

The authors of [7] did not derive the asymptotics for which p determine whether $\mathcal{A}_{k,n}$ is a truncated representative k-basis as $n \to \infty$, a question which we take up presently. Our work is organized as follows: We present results on truncated representative 2-bases in Section 2, using some simple Chernoff bounds. In Section 3, we consider similar questions for truncated representative k-additive bases, and we apply Talagrand's inequality [1] to derive our desired results. The methods of proof are not dissimilar to the ones in [5], though the use of Talagrand's inequality streamlines the analysis significantly.

Remark 1.5. An alternate way of dealing with the boundary effects encountered in finite additive bases is to define *modular representative k-bases*. A set $\mathcal{A}_{k,n} \subseteq [n-1] \cup \{0\}$ is said to be a modular representative k-basis for $[n]$ if the number of ways that each $j \in [n-1] \cup \{0\}$ can be written as a $\bmod(n)$ k-sum of elements of $\mathcal{A}_{k,n}$ is $\Theta(\log n)$. Definitive results on the emergence of modular additive bases have been proved in the papers of Yadin [15] using the method of Brun's sieve and in Sandor [13] using Janson's correlation inequalities. Neither tackled the representative basis question, as we do in the present work. We believe, moreover, that the truncated basis is the more natural finite variant of the problem considered by Erdős and Tetali in [5]. They were concerned with constructing a basis with $r_k(n) = \Theta(\log n)$ for all integers greater than a fixed but arbitrary N_0. This allowed them great flexibility in choosing the threshold N_0 to be large enough to achieve the desired behavior. One might ask how small N_0 can be while still maintaining an additive basis with $r_k(n) = \Theta(\log n)$, which is the natural analogue to the truncated basis question explored below.

Throughout the rest of the paper, we suppress the descriptors "truncated" and "log", referring simply to "representative k-bases."

2. 2-Additive Representations

Consider first the case where $k = 2$. Construct the random set $\mathcal{A}_{2,n}$ by choosing each integer in $[n] \cup \{0\}$ to be in $\mathcal{A}_{2,n}$ independently with probability $p = p_n$. Let $S_2(\alpha, n) := [\alpha n, (2 - \alpha)n]$, and for each $j \in S_2(\alpha, n)$, let $Y_{j,n}$ be the number of ways that j can be represented as a 2-sum of distinct elements of $\mathcal{A}_{2,n}$; the case where summands are allowed to be equal can be proved exactly as in what follows. Let $I_{j,n} := 1\{Y_{j,n} \neq \Theta(\log n)\}$, so that $X_n := \sum_{j \in S_2(\alpha, n)} I_{j,n}$ is the number of elements of $S_2(\alpha, n)$ that are not represented order $\log n$ times in the 2-sum set.

For each $j \in [\alpha n, n]$, the maximum number of representations as 2-sums from $\mathcal{A}_{2,n}$ is given by

$$\rho_{2,n}(j) = \rho_{2,n}(2n - j) = \left\lceil \frac{j}{2} \right\rceil.$$

Fixing j, for $i = 1, ..., \rho_{2,n}(j)$ let

$$B_{i,n} = 1\{i\text{-th pair of integers in } [n] \cup \{0\} \text{ summing to } j \text{ is present in } \mathcal{A}_{2,n}\}$$

so that $Y_{j,n} = \sum_{i=1}^{\rho_{2,n}(j)} B_{i,n}$. Note that each integer in $[n] \cup \{0\}$ can be in at most one of the $\rho_{2,n}(j)$ pairs of integers summing to j, and so the associated $B_{i,n}$'s are independent $\text{Bern}(p^2)$ random variables. It follows that $Y_{j,n}$ has a $\text{Bin}(\rho_{2,n}(j), p^2)$ distribution. Two straightforward applications of Chernoff-type bounds then yield the following result:

Theorem 2.1. *Let $\alpha \in (0, 1)$ be fixed, and let $\eta > 0$ be arbitrarily small. Create the random set $\mathcal{A}_{2,n}$ by picking each integer in $[n] \cup \{0\}$ to be in $\mathcal{A}_{2,n}$ with probability*

$$p = p_n := \sqrt{\frac{\left(\frac{2}{\alpha} + \eta\right) \log n}{n}}.$$

Then

$$\lim_{n \to \infty} \mathbb{P}(X_n = 0) = 1,$$

so that w.h.p. $\mathcal{A}_{2,n}$ is an asymptotic representative 2-basis as $n \to \infty$.

Proof. First note that $\rho_{2,n}(j)$ is maximized by $j = n$, so that for any constant K, it follows that $\mathbb{P}(Y_{n,n} \geq K \log n) \geq \mathbb{P}(Y_{j,n} \geq K \log n)$ for all $j \in S_2(\alpha, n)$. We have that

$$\mathbb{E}(Y_{n,n}) = \frac{n}{2} \left(\frac{\frac{2}{\alpha} + \eta}{n} \right) \log n + o(1) = \left(\frac{1}{\alpha} + \frac{\eta}{2} \right) \log n + o(1).$$

An application of Chernoff's bound, see for example [3, Theorem 2.15], gives that

for any $\delta > 0$, $j \in S_2(\alpha, n)$:

$$\mathbb{P}[Y_{j,n} \geq (1+\delta)\mathbb{E}(Y_{n,n})] \leq \mathbb{P}\left[Y_{n,n} \geq (1+\delta)\mathbb{E}(Y_{n,n})\right]$$

$$= \mathbb{P}\left[Y_{n,n} \geq (1+\delta)\left(\left[\frac{1}{\alpha} + \frac{\eta}{2}\right]\log n + o(1)\right)\right]$$

$$\leq (1+o(1))\exp\left\{-\left(\frac{1}{\alpha} + \frac{\eta}{2}\right)(\log n)[(1+\delta)\log(1+\delta) - \delta]\right\}.$$

Letting $f(\delta) = \left(\frac{1}{\alpha} + \frac{\eta}{2}\right)[(1+\delta)\log(1+\delta) - \delta]$, we see that f is unbounded and monotonically increasing for $\delta > 0$, and so for any $\lambda > 0$, an appropriate δ_0 can be chosen such that $f(\delta_0) = \lambda + 1$ giving that

$$\mathbb{P}\left[Y_{j,n} \geq (1+\delta_0)\left(\left[\frac{1}{\alpha} + \frac{\eta}{2}\right]\log n + o(1)\right)\right] \leq n^{-\lambda-1}.$$

Next note that $\rho_{2,n}(j)$ is minimized for $j = \alpha n = (2-\alpha)n$, so that for any constant K', it follows that $\mathbb{P}(Y_{\alpha n,n} \leq K'\log n) \geq \mathbb{P}(Y_{j,n} \leq K'\log n)$ for all $j \in S_2(\alpha, n)$. Now

$$\mathbb{E}(Y_{\alpha n,n}) = \frac{\alpha n}{2}\left(\frac{\frac{2}{\alpha} + \eta}{n}\right)\log n + o(1) = \left(1 + \frac{\eta\alpha}{2}\right)\log n + o(1).$$

Another application of Chernoff's bound, see [3, Theorem 2.17], gives then that for any $0 \leq \varepsilon \leq e^{-1}$ and $j \in S_2(\alpha, n)$:

$$\mathbb{P}(Y_{j,n} \leq \varepsilon\mathbb{E}(Y_{\alpha n,n})) \leq \mathbb{P}(Y_{\alpha n,n} \leq \varepsilon\mathbb{E}(Y_{\alpha n,n}))$$

$$= \mathbb{P}\left(Y_{\alpha n,n} \leq \varepsilon\left[\left(1 + \frac{\eta\alpha}{2}\right)\log n + o(1)\right]\right)$$

$$\leq (1+o(1))\exp\left\{-(1 - 2\varepsilon + 2\varepsilon\log\varepsilon)\left(1 + \frac{\eta\alpha}{2}\right)\log n\right\}.$$

Let $g(\varepsilon) = (1 - 2\varepsilon + 2\varepsilon\log\varepsilon)\left(1 + \frac{\eta\alpha}{2}\right)$. We see that $\lim_{\varepsilon\to 0} g(\varepsilon) = 1 + \frac{\eta\alpha}{2}$ so that there exists a $\gamma > 0$ and a $\varepsilon_0 > 0$ such that $g(\varepsilon_0) = 1 + \gamma$ and

$$\mathbb{P}(Y_{j,n} \leq \varepsilon_0\mathbb{E}(Y_{\alpha n,n})) \leq n^{-\gamma-1}$$

for all $j \in S_2(\alpha, n)$. Next, note that

$$\mathbb{P}(X_n = 0) = \mathbb{P}\left(\cap_{j\in S_2(\alpha,n)}\{Y_{j,n} = \Theta(\log n)\}\right)$$

$$= 1 - \mathbb{P}\left(\cup_{j\in S_2(\alpha,n)}\{Y_{j,n} \neq \Theta(\log n)\}\right)$$

$$\geq 1 - \sum_{j\in S_2(\alpha,n)} \mathbb{P}(Y_{j,n} \neq \Theta(\log n))$$

$$\geq 1 - n(n^{-\gamma-1} + n^{-\lambda-1})$$

$$= 1 - n^{-\gamma} - n^{-\lambda}$$

$$\to 1 \text{ as } n \to \infty,$$

which finishes the proof. $\square$

Remark: Note that, with the notation as in Theorem 2.1, if for any constants K, $\varepsilon > 0$ we have

$$p = p_n := \sqrt{\frac{K \log^{1+\varepsilon} n}{n}},$$

then

$$\mathbb{E}(Y_{n,n}) = \frac{n}{2} \frac{K \log^{1+\epsilon} n}{n} + o(1) = \Theta(\log^{1+\varepsilon} n).$$

As we have that $Y_{n,n} \approx \mathrm{Bin}(\frac{n}{2}, p^2)$, it follows that

$$\mathrm{Var}(Y_{n,n}) = \Theta(\log^{1+\varepsilon} n),$$

and a simple application of Chebyshev's inequality gives

$$\mathbb{P}(|Y_{n,n} - \mathbb{E}(Y_{n,n})| \le \log n) \ge 1 - \Theta([\log^{1-\varepsilon} n]^{-1}) \to 1 \text{ as } n \to \infty.$$

Therefore $\mathbb{P}(X_n = 0) \to 0$ and $\mathcal{A}_{2,n}$ is not an asymptotic representative 2-basis.

In [7], the authors were able to show that if

$$p = p_n := \sqrt{\frac{\frac{2}{\alpha} \log n - \frac{2}{\alpha} \log \log n + A_n}{n}}$$

for an arbitrary sequence $A_n = o(\log \log n)$, then

$$\mathbb{P}(\mathcal{A}_{2,n} \text{ is a truncated 2-basis}) = \begin{cases} 1 & \text{if } A_n \to \infty \\ 0 & \text{if } A_n \to -\infty \\ \exp\{-2\alpha e^{-\alpha A/2}\} & \text{if } A_n \to A. \end{cases}$$

It follows immediately that if

$$p = p_n := \sqrt{\frac{K \log n}{n}},$$

for some $0 < K < 2$, then with probability converging to 1, $\mathcal{A}_{2,n}$ will not be a k-basis and hence cannot be a representative 2-basis. At the threshold value

$$p = p_n \approx \sqrt{\frac{\frac{2}{\alpha} \log n}{n}},$$

we have that the behavior of lower order terms controls whether $\mathcal{A}_{2,n}$ represents each integer at least once, and so it is reasonable to expect that there are integers that are only represented a few times in the 2-sum set of $\mathcal{A}_{2,n}$ and therefore $\mathcal{A}_{2,n}$ will not form a representative 2-basis, though we have no concrete proof of this conjecture beyond our heuristic reasoning. Note however the similarity between the $p = p_n$ of Theorem 2.1 and the $p(z)$ used in [5] to construct their infinite representative basis.

3. k-Additive Representations

The problem is complicated further if we consider the representation question in the k-additive basis case, as different k-sums summing to an integer j are not necessarily disjoint. We shall begin, as before, by creating the random set $\mathcal{A}_{k,n}$ by choosing each integer in $[n] \cup \{0\}$ to be in $\mathcal{A}_{k,n}$ independently with probability $p = p_n$. Fix $\alpha \in (0,1)$, and let $S_k(\alpha, n) := [\alpha n, (k - \alpha)n]$, and for each $j \in S_k(\alpha, n)$, let $Y_{k,n}(j)$ be the number of ways that j can be represented as a k-sum of distinct elements of $\mathcal{A}_{k,n}$. Let $I_{j,k,n} := 1\{Y_{k,n}(j) \neq \Theta(\log n)\}$, so that $X_{k,n} := \sum_{j \in S_k(\alpha, n)} I_{j,k,n}$ is the number of elements of $S_k(\alpha, n)$ that are not represented order $\log n$ times in the k-sum set of $\mathcal{A}_{k,n}$. The following theorem will constitute the main result of the section, and the remainder of the section will be dedicated to its proof.

Theorem 3.1. *Let $\varepsilon > 0$ be fixed, and let $k \geq 3$ be fixed. Create the random set $\mathcal{A}_{k,n}$ by independently picking each integer in $[n] \cup \{0\}$ to be in $\mathcal{A}_{k,n}$ with probability*

$$p = p_n := \sqrt[k]{\frac{K \log n}{n^{k-1}}},$$

with

$$K = K_{\alpha,k} := \frac{(4 + \varepsilon)(k!)^2}{\alpha^{k-1}}.$$

Then

$$\lim_{n \to \infty} P(X_{k,n} = 0) = 1,$$

so that w.h.p. $\mathcal{A}_{k,n}$ is an asymptotic representative k-basis as $n \to \infty$.

Fix $k \geq 3$. For $1 \leq l \leq k$, define $Y_{l,n}^*(j)$ to be the size of a maximum collection of disjoint representations of j as a l-sum of distinct elements of $\mathcal{A}_{k,n}$. The $Y_{k,n}^*$'s are significantly simpler to work with than the original $Y_{k,n}$'s, as the difficulty presented by overlapping k-sums is circumvented. This idea was exploited to great effect in our motivational paper [5]. A few simple calculations yield that for all $i \in [1, (k - \alpha)n]$

$$\mathbb{E}[Y_{l,n}(i)] = O\left(n^{l-1}p^l\right) = O\left(n^{-1+l/k}\right)n^{o(1)}.$$

The disjointness lemma (Lemma 1 in [5]) implies then that for all $l \leq k - 1$,

$$\mathbb{P}(Y_{l,n}^*(i) \geq 3k) \leq O\left(n^{-3}\right)n^{o(1)}.$$

We are ready to establish the following lemmata, the first of which is the analogue of Lemma 10 from [5]:

Lemma 3.2. *With notation as above, it follows that for all $i \in [1, (k - \alpha)n]$ we have*

$$\mathbb{P}\left(Y_{k-1,n}(i) \geq (3k-1)^{k-1}(k-1)!\right) < O\left(n^{-3}\right)n^{o(1)}.$$

Proof. We say that m sets form a Δ-system (of size m) if they have pairwise the same intersections. If $Y_{k-1,n}(i) \geq (3k-1)^{k-1}(k-1)!$, then the Δ-system lemma (Lemma 2, [5]) implies that the set system composed of the $Y_{k-1,n}(i)$ $(k-1)$-sums of i contains a Δ-system of size $3k$, and we shall denote this system via

$$\{S_1^{k-1}, \ldots, S_{3k}^{k-1}\}$$

with common pairwise intersection set R. Letting $|R| = r \leq k-2$ and letting the sum of elements of R be equal to $m < i$, it follows that if $\widehat{S}_i^{k-1} := S_i^{k-1} \setminus R$ then

$$\{\widehat{S}_1^{k-1}, \ldots, \widehat{S}_{3k}^{k-1}\}$$

is a system composed of $3k$ disjoint sets of size $k - 1 - r$ each summing to $i - m$. The probability of such a system occurring is bounded above by

$$\mathbb{P}(Y_{k-1-r,n}^*(i - m) \geq 3k) \leq O\left(n^{-3}\right) n^{o(1)}$$

as desired. $\square$

The next result is the analogue of Lemma 11 from [5]:

Lemma 3.3. *With notation as above, let $C_k := (3k-1)^{k-1}k!$. Then for each j in $[\alpha n, (k-\alpha)n]$, we have*

$$\mathbb{P}\left(Y_{k,n}(j) \geq C_k Y_{k,n}^*(j)\right) \leq O\left(n^{-2}\right) n^{o(1)}.$$

Proof. Slightly abusing notation, we shall write $x \in Y_{k,n}^*(j)$ to mean that x is in one of the maximum collection of disjoint k-sums of j counted by $Y_{k,n}^*(j)$. Then by Lemma 3.2,

$$\mathbb{P}\left(Y_{k,n}(j) \geq C_k Y_{k,n}^*(j)\right) \leq \mathbb{P}\left(\bigcup_{x \in [0,n]} \{Y_{k-1,n}(j - x) \geq \frac{C_k}{k}, \ x \in Y_{k,n}^*(j)\}\right)$$

$$\leq \sum_{x \in [0,n]} \mathbb{P}\left(Y_{k-1,n}(j - x) \geq \frac{C_k}{k}, \ x \in Y_{k,n}^*(j)\right)$$

$$\leq \sum_{x \in [0,n]} \mathbb{P}\left(Y_{k-1,n}(j - x) \geq \frac{C_k}{k}\right)$$

$$\leq \sum_{x \in [0,n]} O\left(n^{-3}\right) n^{o(1)}$$

$$= O\left(n^{-2}\right) n^{o(1)},$$

as desired. $\square$

Next, we shall use Talagrand's inequality (see Section 7.7, [1]) to show that $Y_{k,n}^*(j) = \Theta(\log n)$ with high probability for all $j \in [\alpha n, (k-\alpha)n]$. Towards that end, we prove

Lemma 3.4. *For some constant $B_k \in [-40\sqrt{k}, 40\sqrt{k}]$ we have that*

$$\mathbf{Med}(Y_{k,n}^*(j)) = \mathbb{E}(Y_{k,n}^*(j)) + B_k\sqrt{\mathbb{E}(Y_{k,n}^*(j))}.$$

Proof. First note that $Y_{k,n}^*(j)$ can be written as a function

$$Y_{k,n}^*(j) = f(J_0, J_1, \ldots, J_n)$$

of the indicator variables

$$J_i := \begin{cases} 1 & \text{if } i \in \mathcal{A} \\ 0 & \text{else.} \end{cases}$$

As the k-sums counted by $Y_{k,n}^*(j)$ are disjoint, the function $f(\cdot)$ is one-Lipschitz. Also note that f is $h-$certifiable with $h(s) = ks$, since if $f(J_0, \ldots, J_n) \geq s$, there exists s disjoint k-sums of j present in $\mathcal{A}$, and any other realization of $\mathcal{A}$ with those sk J_i's equal to 1 has $f \geq s$ as well. It immediately follows from Fact 10.1 in [11] that

$$\left|\mathbb{E}(Y_{k,n}^*(j)) - \mathbf{Med}(Y_{k,n}^*(j))\right| \leq 40\sqrt{k\mathbb{E}(Y_{k,n}^*(j))}.$$

This completes the proof. □

Next we prove

Lemma 3.5. *With $p = p_n$ defined as in Theorem 3.1 and notation as above,*

$$\mathbb{E}\left(Y_{k,n}^*(j)\right) \leq \mathbb{E}\left(Y_{k,n}(j)\right) \leq \mathbb{E}\left(Y_{k,n}^*(j)\right) + o(1).$$

Proof. Let $W_{j,k,n}$ be the number of overlapping pairs of k-sums in the set of all k-sums of j using elements of $\mathcal{A}$. Then, as each k-sum not in $Y_{k,n}^*(j)$ must intersect with at least one of the k-sums of $Y_{k,n}^*(j)$, we have that

$$Y_{k,n}^*(j) \leq Y_{k,n}(j) \leq Y_{k,n}^*(j) + W_{j,k,n}.$$

Note that (writing $\sum_{l,*}$ to be the sum over all overlapping pairs of k-sums of j using elements of $\mathcal{A}$ with overlap of size l)

$$
\begin{aligned}
\mathbb{E}(W_{j,k,n}) &= \sum_{l=1}^{k-1}\sum_{l,*} p^{2k-l} = \sum_{l=1}^{k-1} O\left(n^{2k-l-2}p^{2k-l}\right) \\
&= \sum_{l=1}^{k-1} O\left(n^{-l/k}(\log n)^{(2k-l)/k}\right) \\
&= O\left(n^{-1/k}(\log n)^{(k+1)/k}\right) \\
&= o(1),
\end{aligned}
$$

as desired. □

Theorem 3.6. *With p defined as in Theorem 3.1, there exist constants $\gamma_j > 0$ and $\xi > 0$ such that*

$$\mathbb{P}\left(Y^*_{k,n}(j) \le \gamma_j \log(n) + O(\sqrt{\log n})\right) \le 2n^{-1-\xi}.$$

Proof. In [7] it was shown that the number $\rho_{k,n}(j)$ of (not necessarily disjoint) k-sums of distinct elements of $[n] \cup \{0\}$ summing to j, for $j \in [\alpha n, (k-\alpha)n]$, is bounded below by

$$\rho_{k,n}(j) \ge (1 + o(1))\frac{(\alpha n)^{k-1}}{k!(k-1)!}.$$

It is immediate that $\rho_{k,n}(j) = O(n^{k-1})$, so that there exists a constant $C(j) \ge \frac{\alpha^{k-1}}{k!(k-1)!}$ such that $\rho_{k,n}(j) = C(j)(1 + o(1))n^{k-1}$. From Lemma 3.5, we have then that for $j \in [\alpha n, (k-\alpha)n]$,

$$\mathbb{E}[Y^*_{k,n}(j)] = (1 + o(1))C(j)n^{k-1}p^k + o(1)$$
$$= (1 + o(1))C(j)K_{\alpha,k}\log n + o(1),$$

and so Lemma 3.4 gives us that

$$\mathbf{Med}(Y^*_{k,n}(j)) = (1 + o(1))C(j)K_{\alpha,k}\log n + O\left(\sqrt{\log n}\right). \qquad (3)$$

Talagrand's inequality (see Theorem 7.7.1 in [1]) gives us that for all $t, m > 0$ (where $h(s) = ks$ is the aforementioned certification function for $Y^*_{k,n}(j) = f(J_0, J_1, \ldots, J_n)$):

$$\mathbb{P}\left(Y^*_{k,n}(j) \le m - t\sqrt{h(m)}\right)\mathbb{P}\left(Y^*_{k,n}(j) \ge m\right) \le e^{-t^2/4}.$$

Let $t = \sqrt{(4 + 4\xi)\log n}$, and $m = \mathbf{Med}(Y^*_{k,n}(j))$ to see that

$$\mathbb{P}\left(Y^*_{k,n}(j) \le \mathbf{Med}(Y^*_{k,n}(j)) - \sqrt{(4+4\xi)\log n}\sqrt{k\mathbf{Med}(\rho^*_{k,n}(j))}\right) \le 2n^{-1-\xi}.$$

Using (3), we see then that

$$\mathbb{P}\left(Y^*_{k,n}(j) \le (1 + o(1))\left[C(j)K_{\alpha,k} - \sqrt{4 + 4\xi}\sqrt{kC(j)K_{\alpha,k}}\right]\log n + \cdots\right.$$
$$\left. \cdots + O(\sqrt{\log n})\right) \le 2n^{-1-\xi}.$$

Letting $\gamma_j = C(j)K_{\alpha,k} - \sqrt{4 + 4\xi}\sqrt{kC(j)K_{\alpha,k}}$, then for any $\xi < \varepsilon/4$ (where this is the ε from the definition of $K_{\alpha,k}$) it follows from $C(j) \ge \frac{\alpha^{k-1}}{k!(k-1)!}$ that $\gamma_j > 0$ as desired. $\qquad \square$

Theorem 3.7. *With $p = p_n$ defined as in Theorem 3.1, for each j there exists a constant $\delta_j > 0$ such that*

$$\mathbb{P}\left(Y^*_{k,n}(j) \ge \delta_j \log(n) + O(\sqrt{\log n})\right) \le 2n^{-5/4}.$$

Proof. We will again use Talagrand's inequality, but we shall now set

$$m - t\sqrt{km} = \mathbf{Med}(Y_{k,n}^*(j)).$$

Solving for m, we get that

$$m = \left(\frac{t\sqrt{k}}{2} + \frac{1}{2}\sqrt{kt^2 + 4\mathbf{Med}((Y_{k,n}^*(j))}\right)^2.$$

As in the proof of Theorem 3.6, we have that

$$\mathbf{Med}[(Y_{k,n}^*(j)) = (1 + o(1))C(j)K_{\alpha,k}\log n + O\left(\sqrt{\log n}\right),$$

with $C(j) \geq \frac{\alpha^{k-1}}{k!(k-1)!}$ so that

$$m = \left(\frac{t\sqrt{k}}{2} + \frac{1}{2}\sqrt{kt^2 + 4(1 + o(1))C(j)K_{\alpha,k}\log n + O\left(\sqrt{\log n}\right)}\right)^2.$$

Let $t = \sqrt{5\log n}$ to arrive at $m = \delta_j \log n + O(\sqrt{\log n})$ for some constant δ_j. Apply Talagrand's inequality to see that $\mathbb{P}\left(Y_{k,n}^*(j) \geq \delta_j \log n + O(\sqrt{\log n})\right) \leq 2n^{-5/4}$, as desired. $\qquad\square$

We are now ready to prove our main result:

Proof of Theorem 3.1: Let

$$\gamma_n := \min_{j \in S_k(\alpha,n)} \gamma_j, \text{ and } \delta_n := \max_{j \in S_k(\alpha,n)} \delta_j.$$

Note that there exist strictly positive finite functions $g_1(k)$, $g_2(k)$, $g_3(k)$ and $g_4(k)$ of k, such that for all n, $g_1(k) < \gamma_n < g_2(k)$ and $g_3(k) < \delta_n < g_4(k)$. It follows that as $n \to \infty$, we have

$$0 < \lim_{n\to\infty} \gamma_n < \infty, \text{ and } 0 < \lim_{n\to\infty} \delta_n < \infty.$$

It follows from Theorems 3.6 and 3.7 that there exists a $\xi > 0$ such that for all $j \in S_k(\alpha,n)$,

$$\mathbb{P}\left(Y_{k,n}^*(j) \leq \gamma_n \log n + O\left(\sqrt{\log n}\right)\right) \leq 2n^{-1-\xi},$$

$$\mathbb{P}\left(Y_{k,n}^*(j) \geq \delta_n \log n + O\left(\sqrt{\log n}\right)\right) \leq 2n^{-5/4}.$$

It follows immediately that for any constant c,

$$\mathbb{P}\left(Y_{k,n}^*(j) \leq c\right) \geq \mathbb{P}\left(Y_{k,n}(j) \leq c\right),$$

and hence there exists a $\xi > 0$ such that for all $j \in S_k(\alpha, n)$,

$$\mathbb{P}\left(Y_{k,n}(j) \leq \gamma_n \log n + O\left(\sqrt{\log n}\right)\right) \leq 2n^{-1-\xi}.$$

Next note that by Lemma 3.3,

$$\mathbb{P}\left(Y_{k,n}(j) \geq C_k \delta_n \log n + O(\sqrt{\log n})\right)$$

$$= \mathbb{P}\left(Y_{k,n}(j) \geq C_k \delta_n \log n + O\left(\sqrt{\log n}\right), \; Y_{k,n}(j) \geq C_k Y_{k,n}^*(j)\right)$$

$$+ \mathbb{P}\left(Y_{k,n}(j) \geq C_k \delta_n \log n + O\left(\sqrt{\log n}\right), \; Y_{k,n}(j) < C_k Y_{k,n}^*(j)\right)$$

$$\leq O(n^{-2})n^{o(1)} + \mathbb{P}\left(Y_{k,n}^*(j) \geq \delta_n \log n + O\left(\sqrt{\log n}\right)\right)$$

$$= O(n^{-2})n^{o(1)} + 2n^{-5/4}.$$

Therefore, defining the event

$$A_j := \{Y_{k,n}(j) \geq C_k \delta_n \log n + O\left(\sqrt{\log n}\right)\} \cup \{Y_{k,n}(j) \leq \gamma_n \log n + O\left(\sqrt{\log n}\right)\}$$

$$\mathbb{P}(X_{k,n} \geq 1) \leq \mathbb{P}\left(\cup_j A_j\right)$$

$$\leq \sum_j \mathbb{P}\left(Y_{k,n}(j) \geq C_k \delta_n \log n + O\left(\sqrt{\log n}\right)\right)$$

$$+ \sum_j \mathbb{P}\left(Y_{k,n}(j) \leq \gamma_n \log n + O\left(\sqrt{\log n}\right)\right)$$

$$\leq kn\left(O(n^{-2})n^{o(1)} + 2n^{-5/4} + 2n^{-1-\xi}\right)$$

$$= O(n^{-\xi}) = o(1),$$

and $\mathbb{P}(X_{k,n} = 0) \to 1$ as $n \to \infty$ as desired. $\qquad \square$

Remarks: (i) If we consider representations of integers in $[\alpha n, (k - \alpha)n]$ using k integers from $\mathcal{A}_{k,n}$ that are not necessarily distinct, we can prove a result similar to Theorem 3.1. We skip the details.
(ii) In [7], the authors showed that if

$$p := \sqrt[k]{\frac{\frac{k!(k-1)!}{\alpha^{k-1}} \log n - \frac{k!(k-1)!}{\alpha^{k-1}} \log \log n + A_n}{n^{k-1}}}$$

for $A_n = o(\log \log n)$, then

$$\mathbb{P}(\mathcal{A}_{k,n} \text{ is an asymptotic k-basis}) \to \begin{cases} 1 & \text{if } A_n \to \infty \\ 0 & \text{if } A_n \to -\infty \\ \exp\left\{-\frac{2\alpha}{k-1}e^{\frac{-A\alpha^{k-1}}{(k!(k-1)!)}}\right\} & \text{if } A_n \to A < \infty. \end{cases}$$

Therefore if we choose elements to be in $\mathcal{A}_{k,n}$ with probability

$$p = \sqrt[k]{\frac{C \log n}{n^{k-1}}}$$

for $C < \frac{k!(k-1)!}{\alpha^{k-1}}$, then w.h.p. $\mathcal{A}_{k,n}$ is not a k-basis, and so w.h.p. it is not a representative k-basis. When $C = \frac{k!(k-1)!}{\alpha^{k-1}}$, the behavior of $\mathcal{A}_{k,n}$ as a k-basis hinges on the behavior of lower order terms, and so again it is reasonable to expect some integers to be represented only a few times as k-sums of elements of $\mathcal{A}_{k,n}$. We would expect then that $\mathcal{A}_{k,n}$ is not a representative k-basis w.h.p. As the constant C increases to $K_{\alpha,k}$, our random set becomes a representative k-basis w.h.p. as $n \to \infty$. We haven't yet established any threshold behavior when $\frac{k!(k-1)!}{\alpha^{k-1}} \le C < K_{\alpha,k}$, leaving the door open for future research.

4. Further Research

It would be interesting to work out the asymptotics when $\mathcal{A}$ becomes a truncated $\phi(n)$-representative k-basis for $\phi(n) = o(n^\varepsilon)$ for $\phi(n)$ other than $\log n$, and to what extent the linearity of the target sumset can be relaxed from $[\alpha n, (k-\alpha)n]$ to perhaps $[\theta(n), kn - \theta(n)]$. We expect the most difficult challenges to present themselves for $\phi(n) = o(\log n)$. In a similar vein, improvements in Theorems 2.1 and 3.1 would be most instructive. In particular, how close to the additive basis constant $\frac{k!(k-1)!}{\alpha^{k-1}}$ can we force the constant in Theorem 3.1, as discussed in greater detail in the previous paragraph?

References

[1] N. Alon and J. Spencer (1992). *The Probabilistic Method.* Wiley, New York.

[2] A. Barbour, L. Holst, and S. Janson (1992). *Poisson Approximation.* Oxford University Press.

[3] F. Chung and L. Lu (2006). *Complex Graphs and Networks.* NSF-CBMS Lecture Notes Series, American Mathematical Society, Providence.

[4] P. Erdős (1956). Problems and results in additive number theory, in *Colloque sur la Théorie des Nombres* (CBRM), Bruxelles, 127–137.

[5] P. Erdős and P. Tetali (1990). Representations of integers as the sum of k terms, *Rand. Structures Algorithms* **1**, 245–261.

[6] A. Godbole, S. Janson, N. Locantore, and R. Rapoport (1999). Random Sidon sequences, *J. Number Theory* **75**, 7–22.

[7] A. Godbole, C-M Lim, V. Lyzinski, and N.Triantafillou (2013). Sharp threshold asymptotics for the emergence of additive bases, *Integers: Electronic Journal of Combinatorial Number Theory* **13**, Paper #

[8] C. Güntürk and M. Nathanson (2006). A new upper bound for finite additive bases, *Acta Arith.* **124**, 235–255.

[9] N. Hämmerer and G. Hofmeister (1976). Zu einer Vermutung von Rohrbach, *J. Reine Angew. Math.* **286-287**, 239–247.

[10] S. Janson (1990). Poisson approximation for large deviations, *Rand. Structures Algorithms* **1**, 221–229.

[11] M. Molloy and B. Reed (2001). *Graph Coloring and the Probabilistic Method.* Springer Verlag, New York.

[12] A. Mrose (1979). Untere Schranken für die Reichweiten von Extremalbasen fester Ordnung, *Abh. Math. Sem. Univ. Hamburg* **48**, 118–124.

[13] C. Sandor (2007). Random B_h sets and additive bases in $\mathbb{Z}_n$, *Integers: Electronic Journal of Combinatorial Number Theory* **7**, Paper #A32.

[14] G. Yu (2009). Upper bounds for finitely additive 2-bases, *Proc. Amer. Math. Soc.* **137**, 11–18.

[15] A. Yadin (2009). When Do Random Subsets Decompose a Finite Group? *Israel J. Math.* **174**, 203–219.