SHARP THRESHOLD ASYMPTOTICS FOR THE EMERGENCE OF ADDITIVE BASES

Anant Godbole
Department of Mathematics and Statistics, East Tennessee State University, Johnson City, Tennessee

Chang Mou Lim
Department of Mathematics, University of Chicago, Chicago, Illinois

Vince Lyzinski
Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, Maryland
lyzinski@ams.jhu.edu

Nicholas George Triantafillou
Department of Mathematics, University of Michigan, Ann Arbor, Michigan

Received: 7/5/12, Revised: 1/7/13, Accepted: 3/11/13, Published: 3/15/13

Abstract

A set $\mathcal{A} \subseteq [n] \cup \{0\}$ is said to be a 2-additive basis for $[n]$ if each $j \in [n]$ can be written as $j = x + y$, $x, y \in \mathcal{A}$, $x \leq y$. If we pick each integer in $[n] \cup \{0\}$ independently with probability $p = p_n \to 0$, thus getting a random set $\mathcal{A}$, what is the probability that we have obtained a 2-additive basis? We address this question when the target sum-set is $[(1-\alpha)n, (1+\alpha)n]$ (or equivalently $[\alpha n, (2-\alpha)n]$) for some $0 < \alpha < 1$. We use a delicate application of Janson's correlation inequalities in conjuction with the Stein-Chen method of Poisson approximation to tease out a very sharp threshold for the emergence of a 2-additive basis. Generalizations to k-additive bases are then given.

1. Introduction

In 1956, Erdős [3] answered a question posed in 1932 by Sidon by proving that there exists an infinite sequences of natural numbers $\mathcal{S}$ and constants c_1 and c_2 such that for large n,

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

where, for $k \geq 2$, $r_k(n)$ is the *number of ways* of representing the integer n as the sum of k elements from $\mathcal{S}$, a so-called *asymptotic basis* of order k. The result was generalized in the 1990 work of Erdős and Tetali [4] 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 for fixed k, 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

$$C\frac{(\log z_0)^{1/k}}{z_0^{(k-1)/k}} \leq 1/2.$$

They then showed that (2) holds a.s. for large n (implying that $\mathcal{S}$ is an asymptotic basis of order k).

A natural finite variant of the above problem concerns the notion of k-additive bases.

Definition 1. With $[n] := \{1, 2, \ldots, n\}$, a set $\mathcal{A} \subseteq [n] \cup \{0\}$ is said to be a *k-additive basis* for $[n]$ if each $j \in [n]$ can be written as $j = x_1 + x_2 + \ldots + x_k$, $x_i \in \mathcal{A}$, $i = 1, \ldots, k$.

Note that this definition allows for $x_i = x_j \ i \neq j$. In [4], Erdős and Tetali showed that in some probability space, almost all infinite sequences $\mathcal{S}$ satisfy (2) and are asymptotic bases of order k. It is natural to then ask, for finite $\mathcal{A}$, how small $\mathcal{A}$ can be while still being a k-additive basis. For $k = 2$, if $\mathcal{A}$ is a 2-additive basis we must clearly have $\binom{|\mathcal{A}|+1}{2} \geq n$, so that $|\mathcal{A}| \geq \sqrt{2n}(1 + o(1)) = 1.4142(1 + o(1))\sqrt{n}$. Extensive work, using predominantly analytic techniques, has been done to improve this trivial lower bound to $|\mathcal{A}| \geq 1.428\sqrt{n}$ ([9]); $|\mathcal{A}| \geq 1.445\sqrt{n}$ ([6]); $|\mathcal{A}| \geq 1.459\sqrt{n}$ ([13]);, and $|\mathcal{A}| \geq 1.463\sqrt{n}$ ([5]). The best upper bound appears to be $|\mathcal{A}| \lesssim 1.871\sqrt{n}(1 + o(1))$ ([10], [7]).

In this paper, we will use a probability model in which each integer in $[n] \cup \{0\}$ is chosen to be in $\mathcal{A}$ with equal (and low) probability $p = p_n$. We will then give sharp bounds on the probability that the random set $\mathcal{A}$ is a k-additive basis. It is evident that smaller numbers must be present in an additive basis, since, e.g., the only way to represent 1 in a 2-additive basis is as 1+0, and so in the random model edge effects come into play. Therefore, the random ensemble is unlikely to form an additive basis unless we adopt a different approach.

Definition 2. A set $\mathcal{A} \subseteq [n] \cup \{0\}$ is said to be a *truncated k-additive basis* for $[n]$ if each $j \in [\alpha n, (k-\alpha)n]$ can be written as $j = x_1 + x_2 + \ldots + x_k$, $x_i \in \mathcal{A}$, $i = 1, \ldots, k; \alpha \in (0,1)$.

Although our model differs from that of Erdős and Tetali (our set is constructed using constant probability p as opposed to the $p(z)$ used in [4]), the output threshold probabilities for the size of a truncated (or modular) k-additive basis end up being remarkably close to the input probabilities used by Tetali-Erdős to construct their asymptotic bases. We will stress the similarities between our results as appropriate. It also bears mentioning that the threshold size for our random $\mathcal{A}$ to be a 2-additive basis is, up to a logarithmic factor, similar to the bounds outlined above which were derived using predominantly analytic techniques.

Our work is organized as follows: We present threshold results on truncated 2-additive bases in Section 2, using the Stein-Chen method of Poisson approximation (see [2]). This method allows one to not just find the limiting probability that $\mathcal{A}$ forms an additive basis, but also to approximate the probability distribution $\mathcal{L}(X)$ of the random variable X, defined as the number of integers j that cannot be expressed as a sum of elements in $\mathcal{A}$. In Section 3, we consider similar questions for truncated k-additive bases; where we make use of the Janson exponential inequalities ([1], [8]) to estimate some critical baseline quantities that were calculated exactly in Section 2 for $k = 2$.

Remark 3. An alternate way of dealing with the boundary effects encountered in finite additive bases is to define *modular k-additive bases.* A set $\mathcal{A} \subseteq [n-1] \cup \{0\}$ is said to be a modular k-additive basis for $[n]$ if each $j \in [n-1] \cup \{0\}$ can be written as $j = x_1 + x_2 + \ldots + x_k \pmod{n}$, $x_i \in \mathcal{A}$, $i = 1, \ldots, k$. Definitive results on the emergence of modular additive bases have been proved in the paper of Sandor using the method of Brun's sieve [12], and in the papers by Yadin [14] and Yu [13] using Janson's correlation inequalities. We have recovered these results using the implementation of the Janson's inequality and the Stein-Chen method outlined below though details are omitted for the sake of brevity.

We believe that the truncated basis is the more natural finite variant of the problem considered by Erdős and Tetali in [4]. They were concerned with constructing a basis with $r_k(n) = \Theta(\log n)$ for all integers greater than a fixed but arbitrary N. This allowed them great flexibility in choosing the threshold N to be large enough to achieve the desired behavior. A natural question is how small N can be while still maintaining an additive basis with $r_k(n) = \Theta(\log n)$, which is a natural analogue to the truncated basis question explored below. Also, in a sequel to this paper, we will recover the logarithmic behavior of r_k in the truncated basis setting.

Remark 4. Throughout the rest of the paper, we suppress the descriptor "additive," referring simply to "truncated k-bases." Here and in the sequal, we write $f(n) \cong g(n)$ to signify that $\frac{f(n)}{g(n)} = 1 + o(1)$. Also, we write $Y \sim \mathrm{Po}(\lambda)$ to signify that Y is a random variable distributed as Poisson(λ).

2. 2-Additive Bases

We begin by investigating truncated 2-bases in our random model outlined above. We first select each integer from $\{0, 1, \ldots, n\}$ independently with probability $p = p_n$. We thus obtain a random set $\mathcal{A}$, and denote by X the number of integers in $[\alpha n, (2-\alpha)n]$ that cannot be written in the form $x_1 + x_2$, $x_i \in \mathcal{A}$. Evidently, $\mathcal{A}$ is a truncated 2-basis if and only if $X = 0$. We can write $X = \sum_{j=\lceil \alpha n \rceil}^{\lfloor (2-\alpha)n \rfloor} I_j$, where I_j equals one or zero according as the integer j cannot or can be represented as a 2-sum of elements in $\mathcal{A}$. To simplify the notation a bit, we will write simply $\sum_{j=\alpha n}^{(2-\alpha)n} I_j$ for $\sum_{j=\lceil \alpha n \rceil}^{\lfloor (2-\alpha)n \rfloor} I_j$ in the sequel. The main results of this section are summarized in the following theorem:

Theorem 5. *Pick each integer in $[n] \cup \{0\}$ independently with probability $p = p_n \to 0$ thus getting a random set $\mathcal{A}$.*

(i) *Let $Y \sim \mathrm{Po}(\lambda = \mathbb{E}(X))$. Let $\delta > 0$; then for all*

$$p \geq \sqrt{\frac{(\frac{1}{\alpha} + \delta) \log n}{n}}, \tag{3}$$

we have that $d_{TV}(X, Y) \to 0$ $(n \to \infty)$, where

$$d_{TV}(X, Y) := \sup_{A \subseteq \mathbb{Z}^+} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

(ii) *With*

$$p = \sqrt{\frac{\frac{2}{\alpha} \log n - \frac{2}{\alpha} \log \log n + A_n}{n}}, \tag{4}$$

$$A_n \to \infty \Rightarrow \mathbb{P}(\mathcal{A} \text{ is a truncated } 2-\text{basis}) \to 1 \ \ (n \to \infty);$$

$$A_n \to -\infty \Rightarrow \mathbb{P}(\mathcal{A} \text{ is a truncated } 2-\text{basis}) \to 0 \ \ (n \to \infty);$$

and

$$A_n \to A \Rightarrow \mathbb{P}(\mathcal{A} \text{ is a truncated } 2-\text{basis}) \to \exp\left\{-2\alpha e^{-\alpha A/2}\right\} \ \ (n \to \infty).$$

Proof. Here and throughout the paper, let $\mathcal{S}$ denote the random sumset generated by $\mathcal{A}$. The first step in the proof will be to calculate the precise asymptotics of $\mathbb{E}(X)$:

Proposition 6. *With X as above,*

$$\mathbb{E}(X) = \sum_{j=\alpha n}^{(2-\alpha)n} \mathbb{P}(I_j = 1) \cong 2 \sum_{j=\alpha n}^{n} (1-p^2)^{j/2} \cong \frac{4}{p^2} \exp\{-np^2\alpha/2\}.$$

Proof. This is a critical computation and needs to be justified for the entire range of p's that we encounter, in particular for all p satisfying (3). The first direction is easy:

$$\begin{aligned} 2 \sum_{j=\alpha n}^{n} (1-p^2)^{j/2} &\leq 2 \sum_{j=\alpha n}^{\infty} (1-p^2)^{j/2} \\ &\leq \frac{2(1-p^2)^{\alpha n/2}}{1-\sqrt{1-p^2}} \\ &\leq \frac{4}{p^2}(1-p^2)^{\alpha n/2} \\ &\leq \frac{4}{p^2}\exp\{-np^2\alpha/2\}, \end{aligned}$$

using the inequalities $1 - x \leq e^{-x}$ and $\sqrt{1-p^2} \leq 1 - \frac{p^2}{2}$. For the other direction, we have

$$\begin{aligned} 2 \sum_{j=\alpha n}^{n} (1-p^2)^{j/2} &\geq \frac{2(1-p^2)^{\alpha n/2} - 2(1-p^2)^{n/2}}{1-\sqrt{1-p^2}} \\ &= \frac{2(1-p^2)^{\alpha n/2}(1-(1-p^2)^{(1-\alpha)n/2})}{1-\sqrt{1-p^2}} \\ &\cong \frac{2(1-p^2)^{\alpha n/2}\left(1-e^{-np^2(1-\alpha)/2}\right)}{\frac{p^2}{2}} \\ &\cong \frac{4}{p^2}(1-p^2)^{\alpha n/2} \\ &\cong \frac{4}{p^2}\exp\{-np^2\alpha/2\}, \end{aligned}$$

as desired. □

With p defined as in (4) we see that $\mathbb{E}(X) = (1+o(1))(2\alpha \cdot \exp\{-\alpha A_n/2\})$. As is often the case with threshold phenomena, Markov's inequality can be used to easily establish the first part of (ii), as we have $\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \to 0$ provided that $A_n \to \infty$ in (4).

To establish the second and third statements in part (ii) of the theorem, we go beyond estimating the point probability $\mathbb{P}(X = 0)$, using instead the Stein-Chen method of Poisson approximation [2] to establish a total variation approximation for the distribution $\mathcal{L}(X)$ of X. If the distribution of X is approximately Poisson, then we will have $e^{-\lambda} - \varepsilon_n \leq \mathbb{P}(X = 0) \leq e^{-\lambda} + \varepsilon_n$, where $\lambda = \mathbb{E}(X)$ is the mean of the approximating Poisson variable and ε_n is the total variation error bound for the approximation. We have that $\mathbb{E}(X) \to \infty$ if $A_n \to -\infty$; $\mathbb{E}(X) \to 0$ if $A_n \to \infty$; and $\mathbb{E}(X) \to 2\alpha e^{-\alpha A/2}$ if $A_n \to A$. Thus, if we can show that ε_n tends to zero in a window around

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

we will have that $\mathbb{P}(X = 0) \to 0$ if $A_n \to -\infty$; $\mathbb{P}(X = 0) \to 1$ if $A_n \to \infty$; $\mathbb{P}(X = 0) \to \exp\{-2\alpha e^{-\alpha A/2}\}$ if $A_n \to A$.

With $Y \sim \text{Po}(\lambda = \mathbb{E}(X))$, and throughout writing $d_{\text{TV}}(A, B)$ instead of the more appropriate $d_{\text{TV}}(\mathcal{L}(A), \mathcal{L}(B))$, we seek to bound $d_{TV}(X, Y)$. Following [2], we first need to determine, for each j separately, an auxiliary sequence of variables J_{j_i} defined on the same probability space with the property that

$$\mathcal{L}(J_{j_1}, J_{j_2}, \ldots) = \mathcal{L}(I_1, I_2, \ldots | I_j = 1). \tag{5}$$

Our explicitly constructed coupling of the J_{j_i}'s is as follows: If $I_j = 1$, set $J_{j_i} = I_i$ for all $i \neq j$. If $I_j = 0$, for all pairs $x_1, x_2 \in \mathcal{A}$ such that $x_1 \neq x_2$ and $x_1 + x_2 = j$, remove x_1 from $\mathcal{A}$ with probability $\frac{p(1-p)}{1-p^2}$, remove x_2 from $\mathcal{A}$ with probability $\frac{p(1-p)}{1-p^2}$, and remove both x_1 and x_2 from $\mathcal{A}$ with probability $\frac{p^2+1-2p}{1-p^2}$. If $x \in \mathcal{A}$ and $x + x = j$, then we remove x from $\mathcal{A}$ with probability 1. Finally, define $J_{j_i} = 1$ if $i \notin \mathcal{S}$ after the above coupling is implemented. It is clear that (5) is satisfied since we have "de-selected" offending integers based on the conditional probability of one or both integers in a pair being absent, given that both are not present. The total variation bounds derived in [2] are expressed in terms of the probability that the coupled indicator variables are different after the coupling is implemented; i.e., $I_i = 1, J_{j_i} = 0$ or $I_i = 0, J_{j_i} = 1$. Now, $\mathbb{P}(I_i = 1, J_{j_i} = 0) = 0$, since if integer i is not present in $\mathcal{S}$, it cannot magically appear after some integers have been de-selected.

The formula we need thus reduces to

$$d_{TV}(X, Y) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_j \left(\mathbb{P}^2(I_j = 1) + \mathbb{P}(I_j = 1) \cdot \sum_{i \neq j} \mathbb{P}(I_i = 0, J_{j_i} = 1) \right). \tag{6}$$

There are two terms above. The first,

$$\frac{1 - e^{-\lambda}}{\lambda} \sum_j \mathbb{P}^2(I_j = 1) \leq \frac{1}{\lambda} \max_i \mathbb{P}(I_i = 1) \sum_i \mathbb{P}(I_j = 1) = \max_i \mathbb{P}(I_i = 1),$$

and thus is bounded above by $\mathbb{P}(I_{\lceil \alpha n \rceil} = 1) \cong (1 - p^2)^{\alpha n/2}$. This bound tends to zero if $1/\sqrt{n} = o(p)$, a condition that is satisfied for p satisfying (3).

To bound the second term in (6), we begin by noting

$$\frac{1-e^{-\lambda}}{\lambda}\sum_j\left[\mathbb{P}(I_j=1)\sum_{i\neq j}\mathbb{P}(I_i=0,J_{j_i}=1)\right]\leq \max_j\sum_{i\neq j}\mathbb{P}(I_i=0,J_{j_i}=1).$$

We further bound this second term by conditioning on the number of 2-sums of i that are present pre-coupling, denoting this number by B_i. Denote the sumset of $\mathcal{A}$ post-coupling by $\mathcal{S}^*$. We see that

$$\mathbb{P}(I_i=0,J_{j_i}=1)=\sum_{k=1}^{\lceil\frac{i+1}{2}\rceil}\mathbb{P}(i\notin\mathcal{S}^*|B_i=k)\cdot\mathbb{P}(B_i=k).$$

Note that if there exists x such that $i=2x$, then

$$\begin{aligned}\mathbb{P}(B_i=k)=&\binom{\lceil\frac{i+1}{2}\rceil-1}{k}p^{2k}(1-p^2)^{\lceil\frac{i+1}{2}\rceil-k-1}(1-p)\\&+\binom{\lceil\frac{i+1}{2}\rceil-1}{k-1}p^{2k-2}(1-p^2)^{\lceil\frac{i+1}{2}\rceil-k}p,\end{aligned}$$

otherwise (when i is odd and $f(i):=\lceil\frac{i+1}{2}\rceil$ to simplify the notation)

$$\begin{aligned}&\mathbb{P}(B_i=k)=\binom{f(i)}{k}p^{2k}(1-p^2)^{f(i)-k}\\&=\binom{f(i)-1}{k}p^{2k}(1-p^2)^{f(i)-k}+\binom{f(i)-1}{k-1}p^{2k}(1-p^2)^{f(i)-k}\\&\leq\binom{f(i)-1}{k}p^{2k}(1-p^2)^{f(i)-k}+\binom{f(i)-1}{k-1}p^{2k-2}(1-p^2)^{f(i)-k}p\\&\cong\binom{f(i)-1}{k}p^{2k}(1-p^2)^{f(i)-1-k}(1-p)+\binom{f(i)-1}{k-1}p^{2k-2}(1-p^2)^{f(i)-k}p.\end{aligned}$$

Our next step is to bound $\mathbb{P}(i\notin\mathcal{S}^*|B_i=k)$ via

$$\mathbb{P}(i\notin\mathcal{S}^*|B_i=k)\leq\left(\mathbb{P}(\text{a given 2-sum of } i \text{ removed}|B_i=k)\right)^k.$$

To bound the above term, we will assume all elements of the 2-sums of i are part of 2-sums of j and thus have positive probability of being removed by the coupling process. Fix x,y $x\neq y$, such that $x+y=i$. For ease of notation, define P_x to be the event that x is part of a pre-coupling 2-sum of j, and analogously define P_y. Further define R_x to be the event x is removed from $\mathcal{A}$ by the coupling, and analogously define R_y. Consider the following simple calculations:

$$\mathbb{P}(P_x,\ P_y^c,\ R_x)=p(1-p)\left(1-\frac{p(1-p)}{1-p^2}\right)\leq p,$$

since the only undesirable outcome is if only the “other” component of the 2-sum of 0 is removed in lieu of x;

$$\mathbb{P}(P_x,\ P_y,\ R_x,\ R_y^c) = p^2 \frac{p(1-p)}{1-p^2}\left(1 - \frac{p(1-p)}{1-p^2}\right) \le p^3;$$

and

$$\mathbb{P}(P_x,\ P_y,\ R_x,\ R_y) = p^2\left(1 - \frac{p(1-p)}{1-p^2}\right)^2 \le p^2.$$

In the case where $x = y$, the only case we would need to consider is

$$\mathbb{P}(P_x,\ R_x) = p\left(1 - \frac{p(1-p)}{1-p^2}\right) \le p.$$

From this, we see that

$$\mathbb{P}(\text{a given 2-sum of } i \text{ removed} \mid B_i = k) \le 2p + 2p^3 + p^2.$$

Let $C(p) := 2p + 2p^3 + p^2$. Our above calculations yield that

$$\max_j \sum_{i \ne j} \mathbb{P}(I_i = 0, J_{j_i} = 1) \le \Sigma_1 + \Sigma_2, \tag{7}$$

where

$$\Sigma_1 := \max_j \sum_{i \ne j} \sum_{k=1}^{\lceil \frac{i+1}{2} \rceil} \binom{\lceil \frac{i+1}{2} \rceil - 1}{k} p^{2k}(1-p^2)^{\lceil \frac{i+1}{2} \rceil - k - 1}(1-p)(C(p))^k,$$

and

$$\Sigma_2 := \max_j \sum_{i \ne j} \sum_{k=1}^{\lceil \frac{i+1}{2} \rceil} \binom{\lceil \frac{i+1}{2} \rceil - 1}{k-1} p^{2k-2}(1-p^2)^{\lceil \frac{i+1}{2} \rceil - k} p(C(p))^k.$$

Now,

$$\begin{aligned}
\Sigma_1 &= \max_j \sum_{i \ne j}(1-p)\left[(1-p^2+p^2C(p))^{\lceil \frac{i+1}{2} \rceil - 1} - (1-p^2)^{\lceil \frac{i+1}{2} \rceil - 1}\right] \\
&\le (1-p)\max_j \sum_{i \ne j}\left(\left\lceil \frac{i+1}{2} \right\rceil - 1\right)C(p)p^2(1-p^2+C(p)p^2)^{\lceil \frac{i+1}{2} \rceil - 2} \\
&= O\left(np^3 \sum_i (1-p^2+C(p)p^2)^{\frac{i}{2}}\right) \\
&= O\left(n^2p^3e^{-\alpha np^2/2}\right), \qquad (8)
\end{aligned}$$

and hence $\Sigma_1 \to 0$ if p satisfies (3). For Σ_2, we have

$$\begin{aligned}
\Sigma_2 &= C(p)p \max_j \sum_{i\neq j} \sum_{k=1}^{\lceil \frac{i+1}{2} \rceil} \binom{\lceil \frac{i+1}{2} \rceil - 1}{k-1} p^{2k-2}(1-p^2)^{\lceil \frac{i+1}{2} \rceil - k}(C(p))^{k-1} \\
&= C(p)p \max_j \sum_{i\neq j} (1 - p^2 + C(p)p^2)^{\lceil \frac{i+1}{2} \rceil - 1} \\
&\cong C(p)p \max_j \sum_{i\neq j} e^{-\lceil i/2 \rceil p^2} \\
&= O(np^2 e^{-\alpha np^2/2}) \\
&= o(n^2p^3 e^{-\alpha np^2/2}). \qquad (9)
\end{aligned}$$

Applying (8) and (9) to (7), we see that the second term in (6) tends to 0 if p satisfies (3), and hence $d_{TV}(X, Y) \to 0$ if p satisfies (3). For p in this range

$$|\mathbb{P}(X = 0) - e^{-\lambda}| \to 0$$

which finishes the proof. Note that if we consider a p not in this range, the result will hold by monotonicity. □

Remark 7. We have so far dealt with random sets of fixed expected size, and now indicate briefly how we can easily transition to the case of random sets of fixed size. This can be done for all the results in this paper, but we indicate the method in the context of Theorem 5: Choose one of $\binom{n+1}{|\mathcal{A}|}$ families randomly, and suppose $|\mathcal{A}| = \sqrt{Kn \log n}; K > \frac{2}{\alpha}$ (this corresponds to the expected size of $\mathcal{A}$ with $p = \sqrt{K \log n/n}$). Then we reconcile the two models as follows with $p = \sqrt{K \log n/n}$:

$$\begin{aligned}
&\mathbb{P}(\mathcal{A} \text{ is not a basis}||\mathcal{A}| = \sqrt{Kn \log n}) \\
&\leq \mathbb{P}(\mathcal{A} \text{ is not a basis}||\mathcal{A}| \leq \sqrt{Kn \log n}) \\
&\leq \mathbb{P}(\mathcal{A} \text{ is not a basis})/\mathbb{P}(|\mathcal{A}| \leq \sqrt{Kn \log n}) \\
&\cong 2\mathbb{P}(\mathcal{A} \text{ is not a basis}) \to 0,
\end{aligned}$$

by Theorem 5 and the central limit theorem. A similar argument holds if $K < \frac{2}{\alpha}$, or even if $|\mathcal{A}| = \sqrt{\frac{2}{\alpha} n \log n - \frac{2}{\alpha} n \log \log n + nA_n}$, where $|A_n| \to \infty$. It follows that we can easily go back and forth from the independent model to the fixed set size model except possibly when we are at the threshold, i.e., when

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

with $A_n \to A \in \mathbb{R}$.

3. Truncated k-Bases

Following the same setup as before, we wish to represent each $j \in [\alpha n, (k-\alpha)n]$ as $j = x_1 + \ldots + x_k$ for $x_i \in \mathcal{A}$, where $x_1 \leq x_2 \leq \ldots \leq x_k$. Again, for each j, let I_j equal 1 if j cannot be expressed as a k-sum of elements of $\mathcal{A}$ and 0 otherwise. We set $X := \sum_{j=\alpha n}^{(k-\alpha)n} I_j$, and note that $X = 0 \Leftrightarrow \mathcal{A}$ is a truncated k-basis. Our main theorem is as follows:

Theorem 8. *With X defined as above, if we choose elements of $\{0\} \cup [n]$ to be in $\mathcal{A}$ with probability*

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

where $|A_n| = o(\log\log n)$ and $K = K_{\alpha,k} = \frac{k!(k-1)!}{\alpha^{k-1}}$, then

$$\mathbb{P}(\mathcal{A}\ is\ a\ truncated\ k-basis) \to \begin{cases} 0 & \text{if } A_n \to -\infty \\ 1 & \text{if } A_n \to \infty \\ \exp\{-\frac{2\alpha}{k-1} e^{-A/K}\} & \text{if } A_n \to A \in \mathbb{R} \end{cases}.$$

Also for

$$p = \sqrt[k]{\frac{\beta K \log n}{n^{k-1}}}, \tag{10}$$

$\beta > (k-1)/k$, and $Y \sim \mathrm{Po}(\mathbb{E}(X))$,

$$d_{TV}(X, Y) \to 0 \text{ as } n \to \infty.$$

Before proving this theorem, we need some preliminary work. Let S_j be the set of all unordered k-tuples of nonnegative integers in $\{0\} \cup [n]$ that sum to j.

Claim 9. *For $j \in [\alpha n, n]$,*

$$|S_j| = j^{k-1}/[(k-1)!k!] + O(j^{k-2}).$$

Proof. The number of ordered k-tuples of nonnegative integers that sum to j is $\binom{j+k-1}{j} \cong \frac{j^{k-1}}{(k-1)!}$. All such tuples will be composed entirely of numbers in $\{0\} \cup [j]$, and at most $\binom{k}{2} \cdot n \cdot \binom{j+k-3}{j} = O(j^{k-2})$ of these contain a number repeated once or more often. We can disregard these in the asymptotic analysis, and consider the remaining unordered and ordered tuples. Each remaining unordered tuple appears $k!$ times among the remaining ordered tuples, giving us the desired first order asymptotics. □

Claim 10. *If $1 \leq j \leq kn/2$, then $|S_j| = |S_{kn-j}|$.*

Proof. There is a bijection between k-tuples in S_j for $1 \leq j \leq kn/2$ and those in S_j for $j \geq kn/2$ given by $\{x_1, x_2, \ldots, x_k\} \leftrightarrow \{n - x_1, n - x_2, \ldots, n - x_k\}$. □

Lastly, we have:

Claim 11. *For $j \in [n+1,(k-1)n]$,*

$$|S_j| \geq \frac{n^{k-1}}{(k-1)!k!} + O(n^{k-2}).$$

Proof. We will use the fact that $|S_j|$, i.e., the number of partitions of j of size at most k each part of which is less than or equal to n, is the coefficient of q^j in the q-binomial coefficient

$$\binom{n+k}{k}_q := \frac{(1-q)(1-q^2)\cdots(1-q^{n+k})}{(1-q)(1-q^2)\cdots(1-q^n)(1-q)(1-q^2)\cdots(1-q^k)}.$$

It is well known, see for example [11], that $\binom{n+k}{k}_q = \sum a_i q^i$ is a polynomial in q and that the coefficients are unimodal, namely $a_{j-1} < a_j$ for $j \leq nk/2$. Claim 9 yields that

$$a_n = |S_n| = \frac{n^{k-1}}{(k-1)!k!} + O(n^{k-2}),$$

and the proof now follows directly from Claim 10. □

We next begin our analysis of $\mathbb{E}(X)$ with a preliminary claim.

Claim 12. *With S_j defined as above,*

$$\mathbb{E}(X) = \sum_{j=\alpha n}^{(k-\alpha)n} \exp\left(-|T_j|p^k(1+o(1))\right),$$

where T_j, the set of k-tuples of distinct elements that add to j, satisfies

$$|T_j| \begin{cases} = j^{k-1}/(k-1)!k!, & \text{if } j \in [\alpha n, n]. \\ \geq n^{k-1}/(k-1)!k! & \text{if } j \in [n,(k-1)n]. \\ = (kn-j)^{k-1}/(k-1)!k! & \text{if } j \in [(k-1)n,(k-\alpha)n]. \end{cases}$$

Proof. Fix j, and let $\{B_{(j,i)}\}_{i=1}^{|S_j|}$ be the event that the ith unordered k-tuple of elements of $\{0\} \cup [n]$ that add to j is in $\mathcal{A}$. Then by Janson's inequality, we have

$$\begin{aligned}
\mathbb{P}(I_j = 1) &= \mathbb{P}(\cap_i B_{(j,i)}^C) \\
&\geq (1-p)^{O(1)}(1-p^2)^{O(j)}(1-p^3)^{O(j^2)} \ldots (1-p^{k-1})^{O(j^{k-2})}(1-p^k)^{|T_j|} \\
&\geq \exp(-\frac{O(1)p}{1-p} - \frac{O(j)p^2}{1-p^2} - \ldots \frac{|T_j|p^k}{1-p^k}) \\
&= \exp\biggl(-|T_j|p^k\left(\frac{1}{1-p^k} + \frac{O(j^{k-2})}{(1-p^{k-1})|T_j|p} + \ldots + \frac{O(1)}{(1-p)|T_j|p^{k-1}}\right)\biggr) \\
&= \exp\{-|T_j|p^k(1+o(1))\}
\end{aligned}$$

as for $x \in (0,1)$, $1-x \geq e^{-x/(1-x)}$, and with our choice of p, $np \to \infty$ and for all j, we have $j = O(n)$.

For an upper bound, Janson's inequality yields

$$P(I_j = 1) \leq \exp\left(-|T_j|p^{k-1}(1+o(1)) + \Delta_j/2\right).$$

Now $\Delta_j = O(n^{2k-3}p^{2k-1})$ since the maximal contribution to Δ_j is when the two tuples have distinct elements and intersect in a single element; as for all values of p in our window, $np \to \infty$. Thus

$$\begin{aligned} P(I_j = 1) &\leq \exp\left(-|T_j|p^{k-1} + O(n^{2k-3}p^{2k-1})\right) \\ &= \exp\left(-|T_j|p^k(1 + O(n^{k-2}p^{k-1}))\right). \end{aligned}$$

With $p = o(n^{-(k-2)/(k-1)})$, a condition satisfied by all of the $p's$ in our theorem, $n^{k-2}p^{k-1} \to 0$, finishing the proof. □

We are now ready to begin the proof of Theorem 8.

Proof. We now have that (with $(1+o(1))$ terms suppressed

$$\mathbb{E}(X) = 2\sum_{j=\alpha n}^{n} \exp\{-p^k j^{k-1}/(k-1)!k!\} + \sum_{j=n}^{(k-1)n} \exp\{-p^k|T_j|\}. \tag{11}$$

The first summation, which we will call Σ_1 (correspondingly calling the second sum Σ_2) can be bounded above as follows, recalling the definition of K from the statement of the theorem, and setting $B = \alpha^{k-2}/k!(k-2)!$:

$$\begin{aligned} \Sigma_1 &\leq 2\sum_{j=\alpha n}^{\infty} \exp\{-p^k j^{k-1}/(k-1)!k!\} \\ &\leq 2\exp\{-n^{k-1}p^k/K\}\left(\sum_{j=0}^{\infty}\left(\exp\{-(k-1)(n\alpha)^{k-2}p^k/(k-1)!k!\}\right)^j\right) \\ &= \frac{2\exp\{-n^{k-1}p^k/K\}}{1-\exp\{-Bn^{k-2}p^k\}} \\ &\leq \frac{2\exp\{-n^{k-1}p^k/K\}}{Bn^{k-2}p^k}(1+Bn^{k-2}p^k) \\ &= \frac{2\exp\{-n^{k-1}p^k/K\}}{Bn^{k-2}p^k}(1+o(1)) \end{aligned} \tag{12}$$

where we used the facts that $\frac{x}{x+1} \leq 1 - e^{-x}$ in the fourth line and, in the final line of the display, that for all choices of p considered in the theorem, $n^{k-2}p^k \to 0$.

Finally, the geometric bound in the second line follows from the fact that the ratio of consecutive terms in the sum satisfies

$$\exp\{-\frac{p^k}{(k-1)!k!}[(j+1)^{k-1}-j^{k-1}]\} \geq \exp\{-\frac{p^k}{(k-1)!k!}(k-1)(\alpha n)^{k-2}\}.$$

For Σ_2 we have

$$\begin{aligned} \Sigma_2 &= \sum_{j=n}^{(k-1)n} \exp\{-p^k|T_j|\} \\ &\leq (k-2)n\exp\{-p^k n^{k-1}/(k-1)!k!\} \to 0, \end{aligned} \tag{13}$$

if p satisfies

$$p \geq \left(\frac{D\log n}{n^{k-1}}\right)^{1/k}; D > (k-1)!k!.$$

A lower bound on Σ_1 (and thus on $\mathbb{E}(X)$) is obtained by using elementary integration by parts to derive a tight estimate for the integral

$$\int_x^\infty e^{-t^{k-1}}dt; x \to \infty,$$

in much the same way that Gaussian tails are analyzed (which is the $k = 3$ case.) We have

$$\begin{aligned} \mathbb{E}(X) &\geq 2\sum_{j=\alpha n}^{n} \exp\{-p^k j^{k-1}/(k-1)!k!\} \\ &\cong 2\int_{\alpha n}^{n} \exp\{-p^k x^{k-1}/(k-1)!k!\}dx \\ &= 2\int_{\alpha n}^{\infty} \exp\{-p^k x^{k-1}/(k-1)!k!\}dx \\ &\quad -2\int_{n}^{\infty} \exp\{-p^k x^{k-1}/(k-1)!k!\}dx. \end{aligned} \tag{14}$$

Setting, for $t > 0$, $\Psi(t,k) = \int_t^\infty \exp\{-Cx^{k-1}\}dx$, we see that

$$\begin{aligned} \Psi(t,k) &\leq \frac{1}{t^{k-2}}\int_t^\infty x^{k-2}\exp\{-Cx^{k-1}\}dx \\ &= \frac{1}{C(k-1)t^{k-2}}\int_{Ct^{k-1}}^\infty e^{-u}du \\ &= \frac{1}{C(k-1)t^{k-2}}\exp\{-Ct^{k-1}\}, \end{aligned} \tag{15}$$

and, for another two constants $E, E' > 0$,

$$\begin{aligned}\Psi(t,k) &= \int_t^\infty \frac{x^{k-2}}{x^{k-2}}\exp\{-Cx^{k-1}\}dx \\ &= \frac{1}{C(k-1)t^{k-2}}\exp\{-Ct^{k-1}\} - \frac{E}{C}\int_t^\infty \frac{1}{x^{k-1}}\exp\{-Cx^{k-1}\}dx \\ &\geq \frac{1}{C(k-1)t^{k-2}}\exp\{-Ct^{k-1}\} - \frac{E'}{C^2}\frac{1}{t^{2k-3}}\exp\{-Ct^{k-1}\} \qquad (16)\end{aligned}$$

using (15). Thus, since in our context $C^2t^{2k-3} \gg Ct^{k-2}$, (12) and (16) combine to give

$$\Sigma_1 \cong \frac{2(k-1)!k!}{(k-1)\alpha^{k-2}n^{k-2}p^k}\exp\{-\alpha^{k-1}n^{k-1}p^k/(k-1)!k!\}.$$

It is easy to verify from (13) that $\Sigma_2 = o(1)\Sigma_1$, and thus

$$\mathbb{E}(X) = \Sigma_1(1+o(1)) \cong \frac{2(k-2)!k!}{\alpha^{k-2}n^{k-2}p^k}\exp\{-\alpha^{k-1}n^{k-1}p^k/(k-1)!k!\}.$$

Let $\lambda = \mathbb{E}(X)$ and let $Y \sim \mathrm{Po}(\lambda)$. First we note that $\mathbb{E}(X)$ tends to zero, infinity, or $\frac{2\alpha}{k-1}e^{-A/K}$ if p is as stated as in the theorem with $A_n \to \infty$, $A_n \to -\infty$ or $A_n \to A$ respectively. To complete the proof of Theorem 8, we use Poisson approximation to show that the total variation distance between X and Y converges to 0, so that, in particular, $\mathbb{P}(X=0) \to e^{-\lambda}$.

Following [2], we first need to determine, for each j separately, an auxiliary sequence of variables J_{j_i} defined on the same probability space with the property that

$$\mathcal{L}(J_{j_1}, J_{j_2}, \ldots) = \mathcal{L}(I_1, I_2, \ldots | I_j = 1). \qquad (17)$$

Such a coupling can *probably* be described explicitly as we did for $k = 2$ but there is no need to do so: It is clear that the more integers in $\mathcal{A}$, the higher the probability that I_j is 0, as integer j is more likely to be representable as a k-sum. Therefore if we let

$$Y_i = \begin{cases} 1 & \text{if} \quad i \in \mathcal{A} \\ 0 & \text{else} \end{cases}$$

then the I_j's are decreasing functions of the baseline i.i.d. random variables $\{Y_i\}$ that have distribution $\mathbb{P}(Y_i = 1) = p; \mathbb{P}(Y_i = 0) = 1 - p$, and a monotone coupling satisfying (17) exists. We can then apply Theorem 2.C of [2] with $\Gamma_\alpha^0 = \Gamma_\alpha^- = \emptyset$,

$$d_{TV}(X,Y) \leq \frac{1-e^{-\lambda}}{\lambda}\left(\sum_j \mathbb{P}^2(I_j = 1) + \sum_j\sum_\ell (\mathbb{E}(I_jI_\ell) - \mathbb{E}(I_j)\mathbb{E}(I_\ell))\right). \qquad (18)$$

Note that the above is just a variation of the bound used in the proof of Theorem 5. Now the first term in (18) is bound by

$$\begin{aligned}\frac{1-e^{-\lambda}}{\lambda}\sum_j \mathbb{P}^2(I_j=1) &\le \max_j \mathbb{P}(I_j=1)\\ &\le \max_j \exp\{-|T_j|p^k(1+o(1))\}\\ &\le e^{-\delta n^{k-1}p^k(1+o(1))},\end{aligned}$$

where δ is a constant not depending on p or n, and this term converges to 0 if $p=\sqrt[k]{\frac{G\log n}{n^{k-1}}}$ for any constant $G>0$. For the double sum in (18), we start by using the estimates

$$\mathbb{P}(I_i=1)\ge \exp\{-|S_i|p^k(1+O(1/np))\}$$

for $i=j,\ell$. For the cross product term $\mathbb{E}(I_jI_\ell)$, for fixed j,ℓ, note that $\mathbb{P}(I_jI_\ell=1)=\mathbb{P}(S=0)$, where

$$S=\sum_{r=j,\ell}\sum_{s=1}^{k}\sum_{\mathbf{a}:=\{a_1,\dots,a_s\}:\sum a_i=r} J_{r,s,\mathbf{a}},$$

and where $J_{\mathbf{a}}:=J_{r,s,\mathbf{a}}$ equals one if the s integers in $\mathbf{a}$ that sum to r (with possible repetition) are all selected to be in $\mathcal{A}$. We use Janson's inequality to get the bound

$$\mathbb{E}(I_jI_\ell)\le e^{-\mathbb{E}(S)+\Delta}, \tag{19}$$

where

$$\Delta=\sum_{\{r,l,\mathbf{a}\}}\sum_{\{s,m,\mathbf{b}\}\sim\{r,l,\mathbf{a}\}}\mathbb{E}(J_{\mathbf{a}}J_{\mathbf{b}}),$$

and $\{r,l,\mathbf{a}\}\sim\{s,m,\mathbf{b}\}$ if $\mathbf{a}\cap\mathbf{b}\ne\emptyset;\{r,l,\mathbf{a}\}\ne\{s,m,\mathbf{b}\}$. Using a worst case scenario estimate for Δ, we see

$$\Delta=O(n^{2k-3}p^{2k-1}).$$

As in the proofs of Claims 9-11, we see

$$\mathbb{E}(S)\ge(|S_j|+|S_\ell|)p^k+p^{k-1}O(n^{k-2})\ge(|S_j|+|S_\ell|)p^k(1+O(1/np)).$$

Equation (19) now yields

$$\mathbb{E}(I_jI_\ell)\le\exp\{-(|S_j|+|S_\ell|)p^k(1+O(1/np))+O(n^{2k-3}p^{2k-1})\},$$

and we then conclude that

$$\begin{aligned}&\sum_j\sum_{\ell\ne j}(\mathbb{E}(I_jI_\ell)-\mathbb{E}(I_j)\mathbb{E}(I_\ell))\\ &\le\sum_j\sum_{\ell\ne j}\exp\{-(|S_j|+|S_\ell|)p^k(1+O(1/np))\}\left(e^{O(n^{2k-3}p^{2k-1})}-1\right)\\ &=O(n^{2k-3}p^{2k-1}\lambda^2),\end{aligned}$$

so that

$$\frac{\sum_j \sum_{\ell \neq j} (\mathbb{E}(I_j I_\ell) - \mathbb{E}(I_j)\mathbb{E}(I_\ell))}{\lambda} = O(n^{2k-3} p^{2k-1} \lambda) \to 0$$

if $\lambda \ll 1/n^{2k-3}p^{2k-1}$, i.e., if

$$\frac{2(k-2)!k!n^{k-1}p^{k-1}}{\alpha^{k-2}} \exp\{-\alpha^{k-1} n^{k-1} p^k/(k-1)!k!\} \to 0,$$

i.e., if

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

for $\beta > (k-1)/k$. Applying this to (18) yields that $d_{\mathrm{TV}}(X, Y) \to 0$ as $n \to \infty$ for all p satisfying (10). □

Open Problems. In both the modular and truncated cases, the representation function question has yet to be addressed. We have preliminary results in this direction and will be publishing them in a future paper.

Acknowledgments. The research of AG, CML, and NGT was supported by NSF grant 1004624. The collaboration between AG and VL was made possible due to the support of the Acheson J. Duncan Fund for the Advancement of Research in Statistics. VL was supported by U.S. Department of Education GAANN grant P200A090128. A preliminary version of these results was presented by AG at the INTEGERS 2009 conference. We also want to acknowledge the helpful comments and suggestions of an anonymous referee.

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

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

[5] K. Ford (2008). Unpublished work.

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

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

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

[9] L. Moser (1960). On the representation of $\{1, 2, \ldots, n\}$ by sums, *Acta Arith.* **6**, 11–13.

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

[11] K. M. O'Hara (1990). Unimodality of Gaussian coefficients: a constructive proof, *J. Combin. Number Theory, Series A*.**43**, 29–52.

[12] C. Sandor (2007). Random B_h sets and additive bases in $\mathbb{Z}_n$, *Integers* **7**, Paper #A32.

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

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