# Sharp Threshold Asymptotics for the Emergence of Additive Bases

Anant Godbole
Department of Mathematics and Statistics
East Tennessee State University

Chang Mou Lim
Department of Mathematics
University of Chicago

Vince Lyzinski Department of Applied Mathematics and Statistics The Johns Hopkins University

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

> > November 19, 2018

#### Abstract

A set  $A \subseteq [n] \cup \{0\}$  is said to be a 2-additive basis for [n] if each  $j \in [n]$  can be written as  $j = x + y, x, y \in A, x \leq y$ . If we pick each integer in  $[n] \cup \{0\}$  independently with probability  $p = p_n \to 0$ , thus getting a random set 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  $0 < \alpha < 1$ . Under either model, the Stein-Chen method of Poisson approximation is used, in conjunction with Janson's inequalities, 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 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 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, Erdős and Tetali constructed a random sequence S of natural numbers by including z in 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 a.s. an asymptotic basis of order k and that (2) holds a.s. for large n.

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

**Definition 1.** With  $[n] := \{1, 2, ..., n\}$ , 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  $j = x_1 + x_2 + ... + x_k$ ,  $x_i \in \mathcal{A}$ , i = 1, ..., 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  $|\mathcal{A}| \geq 1.428\sqrt{n}$  (13];

 $1.463\sqrt{n}$  ([5]). The best upper bound appears to be  $|\mathcal{A}| \leq \sim 1.871\sqrt{n}$  ([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. This may be done in two ways, which lead to the following alternative definitions:

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

**Definition 3.** A set  $A \subseteq [n] \cup \{0\}$  is said to be a truncated k-additive basis for [n] if each  $j \in [\alpha n, (k - \alpha)n]$  can be written as  $j = x_1 + x_2 + \ldots + x_k$ ,  $x_i \in A$ ,  $i = 1, \ldots, k$ ;  $\alpha \in (0, 1)$ .

It turns out that definitive results using Definition 1 have been proved in the papers of Yadin [14] and Sandor [12], and we thus focus on developing results using Definition 2. 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 using the 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]), which is an alternative to the Janson inequalities and Brun's sieve used in [14],[12]. 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 the modular case, we will, accordingly, show as a corollary how our

result partially extend those in [14], [12]. In Section 4, we consider similar questions for truncated k-additive bases; at times, the Janson exponential inequalities ([1], [8]) are used to estimate some critical baseline quantities that were calculated exactly in Section 2 for k = 2.

**Remark 4.** Throughout the rest of the paper, we suppress the descriptor "additive", referring simply to "truncated k-bases," and, occasionally, "modular k-bases".

#### 2 2-Additive Bases

We begin by investigating truncated 2-bases in our random model outlined above. We will then state the analogous result for modular 2-bases and we will explain how the modular results (and their generalization to arbitrary k-bases) extends the existing work in [14] and [12]. As with all results in this paper, we seek to find the threshold value at which the bases emerge.

We begin by selecting each integer from  $\{0, 1, ..., n\}$  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 A.

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

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

it hold that  $d_{TV}(X,Y) \to 0 \ (n \to \infty)$ .

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 S denote the random sumset generated by A. The first step in the proof will be to calculate the precise asymptotics of  $\mathbb{E}(X)$ 

Proposition 6. With X as above,

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

*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:

$$2\sum_{j=\alpha n}^{n} (1-p^2)^{j/2} \leq 2\sum_{j=\alpha n}^{\infty} (1-p^2)^{j/2}$$

$$\leq \frac{2(1-p^2)^{\alpha n/2}}{1-\sqrt{1-p^2}}$$

$$\leq \frac{4}{p^2} (1-p^2)^{\alpha n/2}$$

$$\leq \frac{4}{p^2} \exp\{-np^2\alpha/2\},$$

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

$$\begin{split} 2\sum_{j=\alpha n}^{n}(1-p^2)^{j/2} & \geq & \frac{2(1-p^2)^{\alpha n/2}-2(1-p^2)^{n/2}}{1-\sqrt{1-p^2}} \\ & = & \frac{2(1-p^2)^{\alpha n/2}(1-(1-p^2)^{(2-\alpha)n/2})}{1-\sqrt{1-p^2}} \\ & \geq & \frac{2(1-p^2)^{\alpha n/2}\left(1-e^{-np^2(1-\alpha)/2}\right)}{\frac{p^2}{2}+\frac{p^4}{8}} \\ & \asymp & \frac{4}{p^2}(1-p^2)^{\alpha n/2} \\ & \asymp & \frac{4}{p^2}\exp\{-np^2\alpha/2\}. \end{split}$$

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  $\mathbb{P}(X \ge 1) \le \mathbb{E}(X) \to 0$  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  $\varepsilon_n$  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  $\varepsilon_n$  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

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

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

$$\mathcal{L}(J_{i_1}, J_{i_2}, \ldots) = \mathcal{L}(I_1, I_2, \ldots | I_i = 1).$$
 (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=i, 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

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

There are two terms above. The first,

$$\frac{1 - e^{-\lambda}}{\lambda} \sum_{i} \mathbb{P}^2(I_i = 1) \le \frac{1}{\lambda} \max_{i} \mathbb{P}(I_i = 1) \sum_{i} \mathbb{P}(I_j = 1) = \max_{i} \mathbb{P}(I_i = 1),$$

and thus is bounded above by  $\mathbb{P}(I_{\lceil \alpha n \rceil} = 1) \approx (1 - p^2)^{\alpha n/2}$ . 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 the sum, we begin by noting

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

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  $\mathbb{P}(B_i=k)=\binom{\lceil\frac{i+1}{2}\rceil-1}{k}p^{2k}(1-p^2)^{\lceil\frac{i+1}{2}\rceil-k-1}(1-p)+\binom{\lceil\frac{i+1}{2}\rceil-1}{k-1}p^{2k-2}(1-p^2)^{\lceil\frac{i+1}{2}\rceil-k}p$ , otherwise (when i is odd and  $f(i):=\lceil\frac{i+1}{2}\rceil$  to simplify the notation)

$$\mathbb{P}(B_i = k) = \binom{f(i)}{k} p^{2k} (1 - p^2)^{f(i) - k} 
= \binom{f(i) - 1}{k} p^{2k} (1 - p^2)^{f(i) - k} + \binom{f(i) - 1}{k - 1} p^{2k} (1 - p^2)^{f(i) - k} 
\leq \binom{f(i) - 1}{k} p^{2k} (1 - p^2)^{f(i) - k} + \binom{f(i) - 1}{k - 1} p^{2k - 2} (1 - p^2)^{f(i) - k} p 
\approx \binom{f(i) - 1}{k} p^{2k} (1 - p^2)^{f(i) - 1 - k} (1 - p) + \binom{f(i) - 1}{k - 1} p^{2k - 2} (1 - p^2)^{f(i) - k} p$$

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

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

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 A by the coupling, and analogously define  $R_y$ . Consider the following simple calculations:

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

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

$$\mathbb{P}(P_x R_x) = p.$$

From this, we see that

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

Let  $C(p) = 2p + 2p^3 + p^2$ . Our above calculations yield that  $\max_j \sum_{i \neq j} \mathbb{P}(I_i = 0, J_{j_i} = 1) \leq \Sigma_1 + \Sigma_2$ , where

$$\Sigma_1 := \max_{j} \sum_{i \neq j} \sum_{k=1}^{\left\lceil \frac{i+1}{2} \right\rceil} {\lceil \frac{i+1}{2} \rceil - 1 \choose k} p^{2k} (1 - p^2)^{\left\lceil \frac{i+1}{2} \right\rceil - k - 1} (1 - p) (C(p))^k$$

and

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

Now,

$$\begin{split} \Sigma_1 &= \max_j \sum_{i \neq j} (1-p) \left[ (1-p^2 + p^2 C(p))^{\lceil \frac{i+1}{2} \rceil - 1} - (1-p^2)^{\lceil \frac{i+1}{2} \rceil - 1} \right] \\ &\leq (1-p) \max_j \sum_{i \neq j} \left( \left\lceil \frac{i+1}{2} \right\rceil - 1 \right) C(p) p^2 (1-p^2 + C(p) p^2)^{\lceil \frac{i+1}{2} \rceil - 2} \\ &\approx \max_j \sum_{i \neq j} \lceil i/2 \rceil C(p) p^2 e^{-\lceil i/2 \rceil p^2} \\ &\leq n \cdot \frac{(2-\alpha)n}{2} C(p) p^2 e^{-\alpha n p^2/2} \\ &= O\left( n^2 p^3 e^{-\alpha n p^2/2} \right) \end{split}$$

and hence  $\Sigma_1 \to 0$  if p satisfies (3). For  $\Sigma_2$ , we have

$$\Sigma_{2} = C(p)p \max_{j} \sum_{i \neq j} \sum_{k=1}^{\lceil \frac{i+1}{2} \rceil} {\lceil \frac{i+1}{2} \rceil - 1 \choose k-1} p^{2k-2} (1-p^{2})^{\lceil \frac{i+1}{2} \rceil - k} (C(p))^{k-1}$$

$$= C(p)p \max_{j} \sum_{i \neq j} (1-p^{2} + C(p)p^{2})^{\lceil \frac{i+1}{2} \rceil - 1}$$

$$\approx C(p)p \max_{j} \sum_{i \neq j} e^{-\lceil i/2 \rceil p^{2}}$$

$$\leq C(p)pne^{-\alpha np^{2}/2}$$

$$= o(\Sigma_{1})$$

and we have that  $d_{TV}(X,Y) \to 0$  if p satisfies (3), and 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.  $\square$ 

The modular 2-basis result is proved similarly, and suppressing all details we state the result:

**Theorem 7.** Pick each integer in  $[n-1] \cup \{0\}$  independently with probability  $p = p_n$  thus getting a random set A. Then,

i. Let X denote the number of missing integers in the modulo-2 sumset of A, and let  $Y \sim \text{Po}(\lambda = \mathbb{E}(X))$ . Let  $\delta > 0$ , then for all

$$p \ge \sqrt{\frac{(1+\delta)\log n}{n}}$$

it holds that  $d_{TV}(X,Y) \to 0$ .

ii. If we let

$$p = \sqrt{\frac{2\log n + A_n}{n}} \text{ where } |A_n| = o(\log n),$$

then

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

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

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

The second part of Theorem 7 was proven in [14] using Janson's correlation inequalities and in [12] using the method of Brun's sieve (see for example [1]). In [12], the author is able to derive part i. of Theorem 7 at the threshold value of p, while the Stein-Chen method allows us to derive the result in a window about the threshold, thus somewhat generalizing the previous known results.

**Remark 8.** 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  $p = \sqrt{K\log n/n}$ :

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

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 Modular k-Additive Bases

Let us briefly now turn our attention to modular k-additive bases. We define  $\mathcal{A}$  as in Section 2, and will choose integers to be in  $\mathcal{A}$  with probability  $p = p_n$ . Define

$$S := \{x_1 + x_2 + \ldots + x_k \pmod{n} | x_i \in A\}.$$

We define  $I_j$  to be 0 if  $j \in \mathcal{S}$  and 1 if  $j \notin \mathcal{S}$ . As before, let  $X = \sum_{j=0}^{n-1} I_j$ . In [12] it was proven, again using the method of Brun's sieve, that

**Theorem 9.** With X defined as above, if we choose elements of [0, n-1] to be in  $\mathcal{A}$  with probability

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

where  $|A_n| = o(\log n)$ , then

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

Also, if  $Y \sim \text{Po}(\mathbb{E}(X))$ ,

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

.

Using the Stein-Chen method we are able to reproduce the above result and prove the total variation convergence for all  $p \in [p_0, 1]$ , where  $p_0 > \sqrt[k]{\frac{Ak! \log n}{n^{k-1}}}$ , with A > (k-1)/k. We will provide proof for our contribution to the window of convergence, omitting proofs when results were derived also in [12]

With X as above, we first derive (as in [12]) that

$$\mathbb{E}(X) = n \cdot \exp\left\{-\frac{n^{k-1}p^k}{k!}(1 + O(1/np))\right\}.$$

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

$$\mathcal{L}(J_{i_1}, J_{i_2}, \ldots) = \mathcal{L}(I_1, I_2, \ldots | I_i = 1).$$
 (6)

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, the  $I_j$ 's are decreasing functions of the baseline i.i.d. random

variables  $\{Y_i\}$  that have distribution  $\mathbb{P}(Y_0 = 1) = p; \mathbb{P}(Y_0 = 0) = 1 - p$ , and a monotone coupling satisfying (6) exists. Thus with  $\lambda := \mathbb{E}(X)$  and  $Y \sim \text{Po}(\lambda)$ , we can apply Theorem 2.E and Corollary 2.C.4 in [2] to get

$$d_{TV}(X,Y) \le \frac{1 - e^{-\lambda}}{\lambda} \left( \mathbb{V}(X) - \lambda + 2 \sum_{j=0}^{n-1} \mathbb{P}^2(I_j = 1) \right)$$
$$\le \frac{\mathbb{V}(X)}{\lambda} - 1 + 2\mathbb{P}(I_0 = 1)$$

Now, from the asymptotic formula of  $\mathbb{E}(X)$ , we have that  $2\mathbb{P}(I_0 = 1) \approx 2\exp\left\{-\frac{n^{k-1}p^k}{k!}\right\}$  and this goes to 0 as  $n \to \infty$  if

$$p = \frac{k!^{1/k}\phi(n)}{n^{(k-1)/k}}$$

where  $\phi(n) \to \infty$  as  $n \to \infty$  is arbitrary.

We next need to show that  $\frac{\mathbb{V}(X)}{\lambda} = 1 + o(1)$ .

$$\mathbb{V}(X) = \mathbb{V}\left(\sum_{j=0}^{n-1} I_{j}\right)$$

$$= \sum_{j=0}^{n-1} \mathbb{V}(I_{j}) + 2 \sum_{i < j} \left(\mathbb{E}(I_{i}I_{j}) - \mathbb{E}(I_{i})\mathbb{E}(I_{j})\right)$$

$$= \sum_{j=0}^{n-1} \mathbb{E}(I_{j}) - \sum_{j=0}^{n-1} \mathbb{E}(I_{j})^{2} + 2 \sum_{i < j} \mathbb{E}(I_{i}I_{j}) - n(n-1)\mathbb{E}(I_{0})\mathbb{E}(I_{1})$$

$$= \lambda - (n+n(n-1))\mathbb{E}^{2}(I_{0}) + 2 \sum_{i < j} \mathbb{E}(I_{i}I_{j}),$$

so that

$$\frac{\mathbb{V}(X)}{\lambda} - 1 = \frac{(n-1)\mathbb{E}(I_0 I_1)}{\mathbb{E}(I_0)} - n\mathbb{E}(I_0). \tag{7}$$

Next, we need to bound the growth of  $\mathbb{E}(I_iI_j)$  for i < j. Now  $\mathbb{E}(I_iI_j)$  equals the probability then neither i nor j is in  $\mathcal{S}$ , and thus  $\mathbb{E}(I_iI_j) = \mathbb{P}(S=0)$ , where

$$S = \sum_{r=i,j} \sum_{l=1}^{k} \sum_{\mathbf{a}:=\{a_1,\dots,a_l\}:\sum a_i = r} J_{r,l,\mathbf{a}},$$

where  $J_{\mathbf{a}} := J_{r,l,\mathbf{a}}$  equals one if the l integers in  $\mathbf{a}$  that sum to r (with possible repetition) are all selected to be in  $\mathcal{A}$ . We have that (see for example Lemma 1.5 in [12])

$$\mathbb{E}(S) = 2\frac{\binom{n}{k}}{n}p^k(1 + O(1/np)).$$

Set  $\{r, l, \mathbf{a}\} \sim \{s, m, \mathbf{b}\}$  if  $\mathbf{a} \cap \mathbf{b} \neq \emptyset$ ;  $\{r, l, \mathbf{a}\} \neq \{s, m, \mathbf{b}\}$ . For the associated dependency graph, we see that

$$\Delta = \sum_{\{r,l,\mathbf{a}\}} \sum_{\{s,m,\mathbf{b}\} \sim \{r,l,\mathbf{a}\}} \mathbb{E}(J_{\mathbf{a}}J_{\mathbf{b}}) = O(n^{2k-3}p^{2k-1}) = O(n^{k-2}p^{k-1}\mathbb{E}(S)),$$

and thus by Janson's inequality we have

$$\mathbb{E}(I_0 I_1) \le e^{-\mathbb{E}(S) + \Delta} = \exp\{-2\frac{\binom{n}{k}}{n} p^k (1 + O(1/np) + O(n^{k-2} p^{k-1}))\}.$$

Returning to (7), we see that

$$\frac{\mathbb{V}(X)}{\lambda} - 1$$

$$\leq n \frac{\left(\exp\{-2\frac{\binom{n}{k}}{n}p^{k}(1 + O(1/np) + O(n^{k-2}p^{k-1})\} - \exp\{-2\frac{\binom{n}{k}}{n}p^{k}(1 + O(1/np))\}\right)}{\exp\{-\frac{\binom{n}{k}}{n}p^{k}(1 + O(1/np))\}}$$

$$= n \frac{\exp\{-2\frac{\binom{n}{k}}{n}p^{k}(1 + O(1/np))\}}{\exp\{-\frac{\binom{n}{k}}{n}p^{k}(1 + O(1/np))\}} \left(\exp\{-2\frac{\binom{n}{k}}{n}p^{k}O(n^{k-2}p^{k-1})\} - 1\right)$$

$$\sim \lambda \cdot n^{2k-3}p^{2k-1}.$$

Thus, the total variation distance goes to 0 if

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

which holds if  $p > \frac{(Ak! \log n)^{1/k}}{n^{(k-1)/k}}$ ,  $A > \frac{k-1}{k}$ . Poisson approximation is thus valid in a window that includes the threshold value of  $p = \frac{(k! \log n)^{1/k}}{n^{(k-1)/k}}$ .

#### 4 Truncated k-bases

The analysis for truncated bases is more subtle than that for modular bases, since edge effects need to be handled carefully. 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.

**Theorem 10.** 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

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

Also for

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

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

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

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 11. For  $j \in [\alpha n, n]$ ,

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

*Proof.* The number of ordered k-tuples of nonnegative integers that sum to j is  $\binom{j+k-1}{j} \sim \frac{j^{k-1}}{(k-1)!}$ . 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.  $\square$ 

Claim 12. If  $1 \le j \le kn/2$ , then  $|S_j| = |S_{kn-j}|$ 

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

Lastly, we have:

Claim 13. For  $j \in [n+1, (k-1)n]$ ,

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

*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  $j \le nk/2$ . Claim 11 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 12.  $\square$ 

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

Claim 14. With  $S_i$  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

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

*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

$$\mathbb{P}(I_{j} = 1) = \mathbb{P}(\cap_{i} B_{(j,i)}^{C}) 
\geq (1-p)^{O(1)} (1-p^{2})^{O(j)} (1-p^{3})^{O(j^{2})} \dots (1-p^{k-1})^{O(j^{k-2})} (1-p^{k})^{|T_{j}|} 
\geq \exp(-\frac{O(1)p}{1-p} - \frac{O(j)p^{2}}{1-p^{2}} - \dots \frac{|T_{j}|p^{k}}{1-p^{k}}) 
= \exp(-|T_{j}|p^{k} \left(\frac{1}{1-p^{k}} + \frac{O(j^{k-2})}{(1-p^{k-1})|T_{j}|p} + \dots + \frac{O(1)}{(1-p)|T_{j}|p^{k-1}}\right)) 
= \exp\{-|T_{j}|p^{k}(1+o(1))\}$$

as for  $x \in (0,1)$ ,  $1-x \ge e^{-x/(1-x)}$ , 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

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

Now  $\Delta_j = O(n^{2k-3}p^{2k-1})$  since the maximal contribution to  $\Delta_j$  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

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

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

We are now ready to begin the proof of Theorem 10.

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

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

The first summation, which we will call  $\Sigma_1$  (correspondingly calling the second sum  $\Sigma_2$ ) can be bounded above as follows, recalling the definition of K

from the statement of the theorem, and setting  $B = \alpha^{k-2}/k!(k-2)!$ :

$$\Sigma_{1} \leq 2 \sum_{j=\alpha n}^{\infty} \exp\{-p^{k} j^{k-1}/(k-1)!k!\}$$

$$\leq 2 \exp\{-n^{k-1} p^{k}/K\} \left(\sum_{j=0}^{\infty} \left(\exp\{-(k-1)(n\alpha)^{k-2} p^{k}/(k-1)!k!\}\right)^{j}\right)$$

$$= \frac{2 \exp\{-n^{k-1} p^{k}/K\}}{1 - \exp\{-Bn^{k-2} p^{k}\}}$$

$$\leq \frac{2 \exp\{-n^{k-1} p^{k}/K\}}{Bn^{k-2} p^{k}} (1 + Bn^{k-2} p^{k})$$

$$= \frac{2 \exp\{-n^{k-1} p^{k}/K\}}{Bn^{k-2} p^{k}} (1 + o(1))$$

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

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

For  $\Sigma_2$  we have

$$\Sigma_2 = \sum_{j=n}^{(k-1)n} \exp\{-p^k | T_j | \}$$

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

if

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

A lower bound on  $\Sigma_1$  (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

$$\mathbb{E}(X) \geq 2 \sum_{j=\alpha n}^{n} \exp\{-p^{k} j^{k-1} / (k-1)! k!\}$$

$$\sim 2 \int_{\alpha n}^{n} \exp\{-p^{k} x^{k-1} / (k-1)! k!\} dx$$

$$= 2 \int_{\alpha n}^{\infty} \exp\{-p^{k} x^{k-1} / (k-1)! k!\} dx$$

$$-2 \int_{n}^{\infty} \exp\{-p^{k} x^{k-1} / (k-1)! k!\} dx.$$
(8)

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

$$\Psi(t,k) \leq \frac{1}{t^{k-2}} \int_{t}^{\infty} x^{k-2} \exp\{-Cx^{k-1}\} dx 
= \frac{1}{C(k-1)t^{k-2}} \int_{Ct^{k-1}}^{\infty} e^{-u} du 
= \frac{1}{C(k-1)t^{k-2}} \exp\{-Ct^{k-1}\},$$
(9)

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

$$\Psi(t,k) = \int_{t}^{\infty} \frac{x^{k-2}}{x^{k-2}} \exp\{-Cx^{k-1}\} dx$$

$$= \frac{1}{C(k-1)t^{k-2}} \exp\{-Ct^{k-1}\} - \frac{E}{C} \int_{t}^{\infty} \frac{1}{x^{k-1}} \exp\{-Cx^{k-1}\} dx$$

$$\geq \frac{1}{C(k-1)t^{k-2}} \exp\{-Ct^{k-1}\} - \frac{E'}{C^{2}} \frac{1}{t^{2k-3}} \exp\{-Ct^{k-1}\} \tag{10}$$

using (9). Thus, since in our context  $C^2t^{2k-3}\gg Ct^{k-2}$ ,

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

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

$$\mathbb{E}(X) \sim \Sigma_1 \simeq \frac{2(k-2)!k!}{\alpha^{k-2}n^{k-2}p^k} \exp\{-\alpha^{k-1}n^{k-1}p^k/(k-1)!k!\}.$$

Let  $\lambda = \mathbb{E}(X)$  and let  $Y \sim \text{Po}(\lambda)$ . 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 10, we use Poisson approximation to show that the total variation distance between X and Y converges to 0, so that, in particular,  $\mathbb{P}(X = 0) \to e^{-\lambda}$ .

Using Theorem 2.C of [2] with  $\Gamma_{\alpha}^{0} = \Gamma_{\alpha}^{-} = \emptyset$ ,

$$d_{TV}(X,Y) \le \frac{1 - e^{-\lambda}}{\lambda} \left( \sum_{j} \mathbb{P}^{2}(I_{j} = 1) + \sum_{j} \sum_{\ell} (\mathbb{E}(I_{j}I_{\ell}) - \mathbb{E}(I_{j})\mathbb{E}(I_{\ell})) \right);$$

the above is just a variation of the bound used in the proof of Theorem 5. Now

$$\frac{1 - e^{-\lambda}}{\lambda} \sum_{j} \mathbb{P}^{2}(I_{j} = 1) \leq \max_{j} \mathbb{P}(I_{j} = 1)$$

$$\leq \max_{j} \exp\{-|T_{j}|p^{k}(1 + o(1))\}$$

$$\leq e^{-\delta n^{k-1}p^{k}(1 + o(1))},$$

where  $\delta$  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, we start by using the estimates

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

for  $i = j, \ell$ , and for the cross product term  $\mathbb{E}(I_j I_\ell)$  we use Janson's inequality as in our Stein-Chen treatment of Theorem 9. For fixed  $j, \ell$ , note that  $\mathbb{P}(I_j I_\ell = 1) = \mathbb{P}(S = 0)$ , where

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

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 then use Janson's inequality and the worst case scenario estimate for the  $\Delta$  that arises to estimate

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

and then conclude that

$$\begin{split} & \sum_{j} \sum_{\ell \neq j} (\mathbb{E}(I_{j}I_{\ell}) - \mathbb{E}(I_{j})\mathbb{E}(I_{\ell})) \\ & \leq \sum_{j} \sum_{\ell \neq j} \exp\{-(|S_{j}| + |S_{\ell}|)p^{k}(1 + O(1/np))\} \left(e^{O(n^{2k-3}p^{2k-1})} - 1\right) \\ & \sim n^{2k-3}p^{2k-1}\lambda^{2}, \end{split}$$

so that

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

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 
$$p = \sqrt[k]{\frac{[\beta K] \log n}{n^{k-1}}}$$
, where  $\beta > (k-1)/k$ , as claimed.  $\square$ 

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 plan on publishing them in a future paper.

Acknowledgment 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. We thank the anonymous referee for useful comments and for making us aware of key references (especially the work of Sandor [12]). The paper has improved significantly as a result.

### 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  $\{1, 2, ..., n\}$  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: Electronic Journal of Combinatorial Number Theory 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.