FINITE REPRESENTABILITY OF INTEGERS AS 2-SUMS

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

Zach Higgins
Dept. of Mathematics, University of California San Diego, La Jolla, California

Zoe Koch
Department of Mathematics, University of Vermont, Burlington, Vermont

Received: 5/15/17, Revised: 1/2/18, Accepted: 2/12/18, Published: 4/27/18

Abstract

A set $\mathcal{A} = \mathcal{A}_{h,n} \subset [n] \cup \{0\}$ is said to be an additive h-basis if each element in $\{0, 1, \ldots, hn\}$ can be written as an h-sum of elements of $\mathcal{A}$ in *at least* one way. We seek *multiple* representations as sums of h not necessarily distinct elements of $\mathcal{A}$, and, in this paper we make a start by restricting ourselves to $h = 2$. We say that $\mathcal{A}$ is a truncated $(\alpha, 2, g)$ additive basis if each $j \in [\alpha n, (2 - \alpha)n]$ can be represented as a 2-sum of elements of $\mathcal{A}$ in at least g ways. In this paper, we provide sharp asymptotics for the event that a randomly selected set from $\{1, 2, \ldots, n\}$ is a truncated $(\alpha, 2, g)$ additive basis with high or low probability.

1. Introduction and Statement of Results

1.1. Balls in Boxes

We start by introducing results from the classical theory of the random allocation of balls to boxes. We will be seeing, in the rest of the paper, how and to what extent the results apply to situations such as ours, i.e., "representing integers as sums of h, not necessarily distinct, integers from a random set of integers."

Suppose that we are trying to place balls in boxes so that each box contains at most one ball. This is the so-called "birthday problem", and it is well-known, e.g. [2], that if we randomly throw n balls into N boxes, then the threshold for the property to hold with high or low probability (whp or wlp) is $n = \sqrt{N}$. More precisely, if $n/\sqrt{N} \to \infty$, then the probability that each box contains at most one ball is asymptotically 0, and this probability is asymptotically 1 if $n/\sqrt{N} \to 0$.

Here and throughout this paper, we will describe these two situations by using the notation $n \gg \sqrt{N}$ and $n \ll \sqrt{N}$ respectively. There is a generalization of the birthday threshold to "at most g balls," which we rederived in [7] using Talagrand's inequality [1]:

Theorem 1.1. *Suppose n balls are randomly and uniformly distributed in N boxes, and let $X = X_g$ denote the number of boxes with at least $g+1$ balls. If $n \ll N^{g/(g+1)}$, then*

$$\mathbb{P}(X = 0) \to 1,$$

as $n \to \infty$, and if $n \gg N^{g/(g+1)}$, then

$$\mathbb{P}(X = 0) \to 0,$$

as $n, N \to \infty$.

Theorem 1.1 exhibits a progression of thresholds, which get close to $n = N$ as $g \to \infty$. It may still be the case, however, that not all boxes will have a ball in them if $n \gg N$, which leads us to the question of the coverage of each box by at least one ball, or Coupon Collection. It is well known that the expected waiting time $\mathbb{E}(W)$ for each of the boxes to be filled is $N(\ln N + \gamma + o(1))$, where γ is Euler's constant, and that the variance of the waiting time is $\sim N^2$. Together with these facts, Chebychev's inequality can be used to prove the following result:

Theorem 1.2. *Suppose n balls are randomly and uniformly distributed in N boxes, and let X and W denote, respectively, the number of empty boxes, and the waiting time until each box is filled. Denote by $\omega(1)$ any function of n that tends to ∞ as $n \to \infty$. Then, if $n = N(\ln N + \omega(1))$, we have*

$$\mathbb{P}(W > n) = \mathbb{P}(X \geq 1) \to 0,$$

and if $n = N(\ln N - \omega(1))$, we have

$$\mathbb{P}(W \leq n) = \mathbb{P}(X = 0) \to 0.$$

Various people have asked about covering each box g or more times. Generalizing work of Erdős and Rényi [4]; and Newman and Shepp [17]; Holst [13] produced the following definitive result:

Theorem 1.3. *If balls are randomly allotted to N boxes, we let V_g denote the waiting time until each box has at least g balls. Then, as $N \to \infty$,*

$$\mathbb{E}(V_g) = N(\ln N + (g-1)\ln\ln N + \gamma - \ln(g-1)! + o(1)),$$

where γ denotes Euler's constant. Normalizing by setting $V_g^* = V_g/N - \ln N - (g-1)\ln\ln N + \ln(g-1)! + o(1)$, we have that

$$\mathbb{P}(V_g^* \leq u) \to \exp\{-e^{-u}\}.$$

From Theorem 1.3, it is easy to derive the following result:

Theorem 1.4. *Suppose n balls are randomly and uniformly distributed in N boxes, and let X_g denote the number of boxes with at most $g-1$ balls. If $n = N(\ln N + (g-1)\ln\ln N + \omega(1))$, then*

$$\mathbb{P}(X_g = 0) \to 1,$$

and if $n = N(\ln N + (g-1)\ln\ln N - \omega(1))$, then

$$\mathbb{P}(X_g = 0) \to 0$$

as $N \to \infty$.

Of particular note is the linearity (in $\ln\ln N$) for coverings beyond the first, showing that an additional iterated logarithmic fraction suffices for each subsequent covering. We hope to show that many of these features stay intact even as dependence is introduced into the covering scenarios. As a final note, we observe that extremal behaviour in the "balls in boxes" example is trivial: The maximal number of balls that may be placed in N boxes so that each contains at most one ball is N, as is the smallest number of balls so as to guarantee at least one ball per box.

1.2. Dependence

A set $\mathcal{A} \subseteq [n]$ is said to be a B_h *set* (the totality of these for all $h \geq 2$ are known as *Sidon sets*) if each of the $\binom{|\mathcal{A}|+h-1}{h}$ sums of h elements drawn with replacement from $\mathcal{A}$ are distinct. A set $\mathcal{A} \subseteq [n] \cup \{0\}$ is said to be an *h-additive basis* if each $j \in [n]$ can be written as the sum of h, not necessarily distinct, elements in $\mathcal{A}$. Thus, a set is a B_h set or an h-additive basis if each element in the potential sumset can be obtained in at most one or at least one way respectively using elements of $\mathcal{A}$. It is known that maximal Sidon sets and minimal additive bases are both of order $n^{1/h}$: See [11] and [15] for Sidon sets and [6], [12], and [16] for the fact that minimal 2-additive bases satisfy $1.463\sqrt{n} \leq |\mathcal{A}| \leq 1.871\sqrt{n}$. See [9] and [10] for other results. In [19], and [14], there are recent improvements. We are interested, however, in random versions of these results, and we start by noting that the corresponding balls in boxes model is as follows:

The balls are the integers randomly chosen from $[n]$ by a process of sampling which includes each integer (in the non-fixed-size sample) independently with probability p. However the balls do not "go into a single box". Rather, each ball colludes with other chosen balls, including itself, generating sums with multisets of $h-1$ other balls. A ball is then placed into the box corresponding to each generated sum. For example, if $h = 2$ and the integers selected in sequence are 4, 2, and 6, then balls are placed in boxes

$$8(= 4 + 4),$$

$$4(=2+2), 6(=2+4),$$

$$12(=6+6), 8(=6+2), 10(=6+4),$$

where the numbers in the three lines indicate what occurs with integers 4, 2, and 6 respectively. There are clearly several layers of dependence in the allocation of balls to boxes.

Three known facts in the area of thresholds for the emergence of Sidon sets and additive bases are stated next. Theorem 1.5 below was proved in [9]:

Theorem 1.5. *Consider a subset A_n obtained by choosing each integer in $[n]$ independently with probability $p = p_n = \frac{k_n}{n}$. Then for any $h \geq 2$,*

$$k_n = o(n^{1/2h}) \Rightarrow \mathbb{P}(A_n \text{ is a } B_h \text{ set}) \to 1 \quad (n \to \infty)$$

and

$$n^{1/2h} = o(k_n) \Rightarrow \mathbb{P}(A_n \text{ is a } B_h \text{ set}) \to 0 \quad (n \to \infty).$$

In [7], we find the following definition that is related to the original question of Sidon (see [18]).

Definition 1.6. We say that $\mathcal{A} \subseteq [n]$ satisfies the $B_h[g]$ property for integers $h \geq 2; g \geq 1$ if for all integers $k, h \leq k \leq nh$, k is realized in at most g ways as a sum

$$a_1 + a_2 + \ldots + a_h = k$$

for $a_1 \leq a_2 \leq \ldots \leq a_h$ and $a_i \in \mathcal{A}$ for each i.

The authors of [7] go on to generalize Theorem 1.5 as follows:

Theorem 1.7. *Let $\mathcal{A} \subseteq [n]$ be a random subset of $[n]$ in which each element of $[n]$ is selected for membership in $\mathcal{A}$ independently with probability $p := \frac{k}{n}$. Then for any $h \geq 2$, $g \geq 1$ we have:*

$$k = o\left(n^{\frac{g}{h(g+1)}}\right) \Rightarrow \mathbb{P}(\mathcal{A} \textit{ is } B_h[g]) \to 1 \quad (n \to \infty),$$

and

$$n^{\frac{g}{h(g+1)}} = o\left(k\right) \Rightarrow \mathbb{P}(\mathcal{A} \textit{ is } B_h[g]) \to 0 \quad (n \to \infty).$$

Theorems 1.5 and 1.7 are about each integer being represented at most once or at most $g \geq 2$ times. Their statements are included so as to draw a full parallel between the balls in boxes results from Section 1.1. In transitioning to the case of additive bases and Theorem 1.8 below, we first note, as in [10], that a single input probability for integer selection will cause edge effect issues. For example, for $h = 2$, since the only way to represent 1 as a 2-sum is as $1+0$, both 0 and 1 *must* be selected in order for 1 to be represented. For this reason, we say that $\mathcal{A} \subseteq [n] \cup \{0\}$

is a *truncated* (α, h) *additive basis* if each integer in $[\alpha n, (h-\alpha)n]$ can be written as an h-sum of elements in $\mathcal{A}$. Given this altered definition, the "each integer appears at least once as an h-sum" analogy is not fully preserved, since we have each integer *in a range* appearing at least once as an h-sum. Nonetheless, we have the following result from [10]:

Theorem 1.8. *If we choose elements of* $\{0\} \cup [n]$ *to be in* $\mathcal{A}$ *with probability*

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

where $K = K_{\alpha,h} = \frac{h!(h-1)!}{\alpha^{h-1}}$*, then*

$$\mathbb{P}(\mathcal{A} \textit{ is a truncated } (\alpha, h) \textit{ additive basis}) \to \begin{cases} 0 \\ 1 \\ \exp\{-\frac{2\alpha}{h-1} e^{-A/K}\} \end{cases}$$

according as A_n *tends to* $-\infty, \infty$ *with* $|A_n| = o(\log\log n)$*, or* $A \in \mathbb{R}$ *respectively.*

Even though edge effects *can* be eliminated by considering modular additive bases, here we continue to consider the truncated additive basis case, where the target sumset is reduced via the parameter α, since we are using the same probability p of selection. The case $h = 2$ is studied in greater detail in the next result, which addresses coverage of each integer as a 2-sum in at least g different ways. This is the main result of this paper.

Theorem 1.9. *If we choose elements of* $\{0\} \cup [n]$ *to be in* $\mathcal{A}$ *with probability*

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

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

$$\mathbb{P}(\mathcal{A} \textit{ is a truncated } (\alpha, 2, g) \textit{ basis}) \to \begin{cases} 0 & \textit{if } A_n \to -\infty \\ 1 & \textit{if } A_n \to \infty \\ \exp\left\{-\frac{2\alpha}{(g-1)!} e^{-A\alpha/2}\right\} & \textit{if } A_n \to A \in \mathbb{R} \end{cases},$$

where a truncated $(\alpha, 2, g)$ *basis is one for which each integer in the target set* $[\alpha n, (2-\alpha)n]$ *can be written as a 2-sum in at least* g *ways.*

Theorem 1.9 exhibits the log log phenomenon that arose in the context of Coupon Collection. Interestingly, though, the log log factor is present for the first covering with a negative contribution, disappears for the second, and then reappears with a positive sign. The paper [7] provides many more examples of this phenomenon in a variety of covering and packing situations, specifically those that arise in the context of combinatorial designs, permutations, and union free set families.

2. Proof of Theorem 1.9

Let $r_{\mathcal{A}}(j)$ be the number of pairs $a_1, a_2 \in \mathcal{A}$ with $a_1 \le a_2$ and $a_1 + a_2 = j$. We say that j is *underrepresented* if and only if $r_{\mathcal{A}}(j) \le g-1$. Note that if $B = [n]$, then $r_B(j) = r_B(2(n+1)-j)$. Let $X = X_g$ be the number of integers in $[\alpha n, (2-\alpha)n]$ that are underrepresented. The threshold we seek to establish is for $\mathbb{P}(X=0)$, and, as in so many instances where we employ the Poisson paradigm (see, e.g., [1]), this transition occurs at the level at which $\mathbb{E}(X)$ rapidly transitions from asymptotically 0 to asymptotically ∞; this is because $\mathbb{P}(X=0) \sim e^{-\mathbb{E}(X)} = e^{-\lambda}$. Towards this end we next carefully estimate λ. We have that

$$X = \sum_{j=\alpha n}^{(2-\alpha)n} I_j,$$

where I_j is the indicator of the event that the integer j is underrepresented as defined above. By linearity of expectation,

$$\begin{aligned}\lambda = \mathbb{E}(X) &= \sum_{j=\alpha n}^{(2-\alpha)n} \mathbb{P}(j \text{ is underrepresented})\\ &\sim 2\sum_{j=\alpha n}^{n} \mathbb{P}(j \text{ is underrepresented})\\ &= 2\sum_{j=\alpha n}^{n}\sum_{s=0}^{g-1}\binom{\lfloor j/2\rfloor}{s}p^{2s}(1-p^2)^{\lfloor j/2\rfloor - s}(1+o(1)), \qquad (1)\end{aligned}$$

where the second line in (1) is due to the fact that $r_B(j) = r_B(2(n+1)-j)$. The last equality in (1) is more subtle. If j is odd, it can be represented as the sum of $\lfloor j/2\rfloor$ *disjoint* pairs of integers. For even j, however, the number of representations is $j/2$, one of which is of the form $a+a$. The correct summand in the last line of (1) is thus, for even j, $(1-p)(1-p^2)^{\frac{j}{2}-1}$ for $s=0$ and for $s \ge 1$ it equals (by considering whether a is or is not selected)

$$\begin{aligned}&\binom{\frac{j}{2}-1}{s-1}p^{2s-1}(1-p^2)^{\frac{j}{2}-s} + \binom{\frac{j}{2}-1}{s}p^{2s}(1-p^2)^{\frac{j}{2}-1-s}(1-p)\\ =\ &\frac{2s}{jp}\binom{\frac{j}{2}}{s}p^{2s}(1-p^2)^{\frac{j}{2}-s} + \frac{1-p}{1-p^2}\binom{\frac{j}{2}}{s}p^{2s}(1-p^2)^{\frac{j}{2}-s}(1+o(1))\\ =\ &\binom{\frac{j}{2}}{s}p^{2s}(1-p^2)^{\frac{j}{2}-s}(1+o(1)),\end{aligned}$$

assuming (as we may) that $p \to 0$ and $jp \to \infty$. Whether $s = 0$ or $s \geq 1$, therefore, the expression in (1) is valid. Trivially, we have

$$\lambda \geq 2 \sum_{j=\alpha n}^{n} \binom{\lfloor j/2 \rfloor}{g-1} p^{2g-2}(1-p^2)^{\lfloor j/2 \rfloor - g+1}, \tag{2}$$

and Proposition A.2.5 (iii) in [3], which estimates the left tail of a binomial random variable with large mean by its last term, yields

$$\begin{aligned}
\lambda &\leq 2 \sum_{j=\alpha n}^{n} \frac{jp^2/2 - gp^2}{jp^2/2 + 1 - g - p^2} \binom{\lfloor j/2 \rfloor}{g-1} p^{2g-2}(1-p^2)^{\lfloor j/2 \rfloor - g+1} \\
&= 2 \sum_{j=\alpha n}^{n} \binom{\lfloor j/2 \rfloor}{g-1} p^{2g-2}(1-p^2)^{\lfloor j/2 \rfloor - g+1}(1+o(1)). \qquad (3)
\end{aligned}$$

In deriving (3), we need to know that $jp^2 \to \infty$ for j's in the selected range. This is something we can assume, since we are seeking a threshold at $p \sim \sqrt{K \log n/n}$, and we can suppose up front, e.g., that $p \geq \sqrt{\log\log n/n}$. Inequalities (2) and (3) reveal that

$$\begin{aligned}
\lambda &\sim 2 \sum_{j=\alpha n}^{n} \binom{\lfloor j/2 \rfloor}{g-1} p^{2g-2}(1-p^2)^{\lfloor j/2 \rfloor - g+1} \\
&\sim \frac{2}{(g-1)!} \sum_{j=\alpha n}^{n} \left(\frac{jp^2}{2}\right)^{g-1} e^{-jp^2/2} \qquad (4)
\end{aligned}$$

where, in the second line of (4) we have used the facts that $p \to 0$ and g is finite, and that for j's in the specified range, we have $(1-p^2)^{\lfloor j/2 \rfloor} \sim e^{-jp^2/2}$, and $\binom{\lfloor j/2 \rfloor}{g-1} \sim j^{g-1}/(2^{g-1}(g-1)!)$. Since the function $x^{g-1}e^{-x}$ is decreasing for $x > g-1$, we see that the summand in (4) will also be decreasing provided, e.g., that $p^2 \geq \frac{\log\log n}{\alpha n}$, which we will assume. Thus

$$\begin{aligned}
\lambda &\leq \frac{2}{(g-1)!} \sum_{j=\alpha n}^{\infty} \left(\frac{jp^2}{2}\right)^{g-1} e^{-jp^2/2} \\
&\leq \frac{4}{p^2(g-1)!} \int_{\alpha n p^2/2}^{\infty} x^{g-1} e^{-x} dx + o(1) \\
&\sim \frac{4}{p^2(g-1)!} \left(\frac{\alpha n p^2}{2}\right)^{g-1} \exp\{-\alpha n p^2/2\}, \qquad (5)
\end{aligned}$$

where the last line of (5) follows from the simplest asymptotic estimate (i.e., without error terms) of the incomplete gamma function. Next, we return to (4) and see that

$$\begin{aligned}
\lambda &\geq \frac{4}{p^2(g-1)!}\int_{\alpha np^2/2}^{np^2/2} x^{g-1}e^{-x}dx \\
&= \frac{4}{p^2(g-1)!}\int_{\alpha np^2/2}^{\infty} x^{g-1}e^{-x}dx - \frac{4}{p^2(g-1)!}\int_{np^2/2}^{\infty} x^{g-1}e^{-x}dx \\
&\sim \frac{4}{p^2(g-1)!}\left(\left(\frac{\alpha np^2}{2}\right)^{g-1}\exp\{-\alpha np^2/2\} - \left(\frac{np^2}{2}\right)^{g-1}\exp\{-np^2/2\}\right) \\
&= \frac{4}{p^2(g-1)!}\left(\frac{\alpha np^2}{2}\right)^{g-1}\exp\{-\alpha np^2/2\}\left(1-\left(\frac{1}{\alpha}\right)^{g-1}e^{-(1-\alpha)np^2/2}\right) \\
&\sim \frac{4}{p^2(g-1)!}\left(\frac{\alpha np^2}{2}\right)^{g-1}\exp\{-\alpha np^2/2\}. \qquad (6)
\end{aligned}$$

Combining (5) and (6) we get

Lemma 2.1.

$$\lambda \sim \frac{4}{p^2(g-1)!}\left(\frac{\alpha np^2}{2}\right)^{g-1}\exp\{-\alpha np^2/2\}.$$

It follows, via a careful calculation, that with

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

$$\lambda \sim \frac{2\alpha}{(g-1)!}e^{-\alpha A_n/2},$$

so that $\lambda \to 0$ if $A_n \to \infty$, and $\lambda \to \infty$ if $A_n \to -\infty$. In the former case we have

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \to 0,$$

which establishes the first part of the theorem.

The next (and critical) phase of the proof is to show that $\mathbb{P}(X = 0) \approx e^{-\lambda}$. We will exhibit this by using the Stein-Chen method of Poisson approximation (specifically Corollary 2.C.4 in [3]), which will yield that

$$d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Po}(\lambda)) = \sup_{\mathrm{A}\subseteq\mathbb{Z}^+}\left|\mathbb{P}(\mathrm{X}\in\mathrm{A}) - \sum_{\mathrm{j}\in\mathrm{A}}\frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{j}}}{\mathrm{j}!}\right| \to 0$$

for a range of p's that encompasses our threshold. (In the above $\mathcal{L}(Z)$ denotes the distribution of Z, $\mathrm{Po}(\lambda)$ the Poisson distribution with parameter λ, and d_{TV} the usual total variation distance.) Setting $A = \{0\}$ will complete the proof.

To apply Corollary 2.C.4 in [3], we must show that the variables $\{I_j : \alpha n \le j \le (2-\alpha)n\}$ are positively related. Here we make use of Theorem 2.E of [3]. Let X_i be the indicator for the event $i \notin \mathcal{A}$. Fix a $j \in [\alpha n, (2-\alpha)n]$. For any $i \in \{0\} \cup [n]$, if X_i increases from 0 to 1, then i went from being chosen to be put into $\mathcal{A}$ to being not chosen. In this case, j is at least as likely to be underrepresented as it was before, so I_j cannot decrease. Thus $\{I_j : \alpha n \le j \le (2-\alpha)n\}$ is a collection of increasing functions of the independent random variables $\{X_i : i \in \{0\} \cup [n]\}$, and so the variables $\{I_j : \alpha n \le j \le (2-\alpha)n\}$ are positively related by Theorem 2.E in [3]. Corollary 2.C.4 then gives

$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Po}(\lambda)) &\le \frac{1-e^{-\lambda}}{\lambda}\left(Var(X) - \lambda + 2\sum_j \mathbb{P}^2(I_j = 1)\right) \\ &\le \frac{1}{\lambda}\left(\sum_j \mathbb{P}^2(I_j = 1) + \sum_j \sum_\ell \{\mathbb{E}(I_j I_\ell) - \mathbb{E}(I_j)\mathbb{E}(I_\ell)\}\right) \\ &= T_1 + T_2, \quad \text{say.} \end{aligned} \tag{7}$$

Starting with T_1,

$$\begin{aligned} T_1 &\le \frac{1}{\lambda} \max_j \mathbb{P}(I_j = 1) \sum_j \mathbb{P}(I_j = 1) \\ &= \mathbb{P}(I_{\alpha n} = 1) \\ &= \sum_{j=0}^{g-1} \binom{\lfloor \alpha n/2 \rfloor}{j} p^{2j} (1-p^2)^{\lfloor \alpha n/2 \rfloor - j} \\ &\le \binom{\lfloor \alpha n/2 \rfloor}{g-1} p^{2g-2} (1-p^2)^{\lfloor \alpha n/2 \rfloor - g + 1} (1 + o(1)) \\ &\le \frac{1}{(g-1)!} \left(\frac{\alpha n p^2}{2}\right)^{g-1} e^{-\alpha n p^2/2} (1 + o(1)) \\ &\to 0 \end{aligned} \tag{8}$$

provided that $np^2 \to \infty$, which we may assume without any loss. Clearly the correlation term T_2 will dictate the closeness of the Poisson approximation. Our first lemma shows that while computing $\mathbb{P}(I_j I_\ell = 1)$, it suffices to consider the case where the sumsets for j and ℓ are disjoint.

Lemma 2.2. *For some constant K, we have that for each j, ℓ,*

$$\mathbb{P}(I_j I_\ell = 1) \le \mathbb{P}(I_j = 1)\mathbb{P}(I_\ell = 1)\left(1 + \frac{K}{np}\right).$$

Proof. We let $A_{j,\ell,r,s}$ denote the event that integers j, ℓ are represented r and s times respectively. Likewise, let $B_{j,\ell,r,s}$ denote the event that integers j, ℓ are represented in r and s entirely disjoint ways. We have that

$$\begin{aligned}
\mathbb{P}(I_j I_\ell = 1) &= \sum_{r,s=0}^{g-1} \mathbb{P}(A_{j,\ell,r,s}) \\
&= \sum_{r,s=0}^{g-1} \mathbb{P}(B_{j,\ell,r,s}) + \sum_{r,s=0}^{g-1} \mathbb{P}(B^C_{j,\ell,r,s}). \qquad (9)
\end{aligned}$$

We first calculate the contribution to (9) of the disjoint case:

$$\begin{aligned}
&\sum_{r,s=0}^{g-1} \mathbb{P}(B_{j,\ell,r,s}) \\
= \;& \sum_{r=0}^{g-1} \binom{\lfloor j/2 \rfloor}{r} p^{2r}(1-p^2)^{\lfloor j/2 \rfloor - r} \sum_{s=0}^{g-1} \binom{\lfloor \ell/2 \rfloor - D}{s} p^{2s}(1-p^2)^{\lfloor \ell/2 \rfloor - D - s}(1-p)^D \\
\leq \;& \sum_{r=0}^{g-1} \binom{\lfloor j/2 \rfloor}{r} p^{2r}(1-p^2)^{\lfloor j/2 \rfloor - r} \sum_{s=0}^{g-1} \binom{\lfloor \ell/2 \rfloor}{s} p^{2s}(1-p^2)^{\lfloor \ell/2 \rfloor - s} \left(\frac{1-p}{1-p^2}\right)^D \\
\leq \;& \mathbb{P}(I_j = 1)\mathbb{P}(I_\ell = 1). \qquad (10)
\end{aligned}$$

In the above array, we have denoted by D the number of pairs of integers, one or both of whose components overlap with the set of pairs of chosen integers that add to j, where $0 \leq D \leq 2r$. We must not choose the second of the two integers that give a sum of ℓ; this explains the $(1-p)^D$ term.

We next turn to $\sum_{r,s=0}^{g-1} \mathbb{P}(B^C_{j,\ell,r,s})$, and see that

$$\begin{aligned}
\sum_{r,s=0}^{g-1} \mathbb{P}(B^C_{j,\ell,r,s}) &= \sum_{r=0}^{g-1} \binom{\lfloor j/2 \rfloor}{r} p^{2r}(1-p^2)^{\lfloor j/2 \rfloor - r} \sum_{s=0}^{g-1} \sum_{t=1}^{D \wedge s} \binom{D}{t} p^t (1-p)^{D-t} \times \\
&\quad \binom{\lfloor \ell/2 \rfloor - D}{s-t} p^{2s-2t}(1-p^2)^{\lfloor \ell/2 \rfloor - s - D + t} \\
&= \sum_{r=0}^{g-1} \binom{\lfloor j/2 \rfloor}{r} p^{2r}(1-p^2)^{\lfloor j/2 \rfloor - r} \sum_{s=0}^{g-1} p^{2s}(1-p^2)^{\lfloor \ell/2 \rfloor - s} \times \\
&\quad \sum_{t=1}^{D \wedge s} \binom{D}{t} \binom{\lfloor \ell/2 \rfloor - D}{s-t} \frac{(1-p)^{D-t}(1-p^2)^{t-D}}{p^t} \\
&\leq \sum_{r=0}^{g-1} \binom{\lfloor j/2 \rfloor}{r} p^{2r}(1-p^2)^{\lfloor j/2 \rfloor - r} \sum_{s=0}^{g-1} p^{2s}(1-p^2)^{\lfloor \ell/2 \rfloor - s} \times \\
&\quad \sum_{t=1}^{D \wedge s} \frac{\binom{D}{t}\binom{\lfloor \ell/2 \rfloor - D}{s-t}}{p^t}. \qquad (11)
\end{aligned}$$

Consider the summand
$$\varphi(t) = \frac{\binom{D}{t}\binom{\lfloor \ell/2 \rfloor - D}{s-t}}{p^t},$$
which is the t-th term in the final sum in (11). We have
$$\frac{\varphi(t+1)}{\varphi(t)} = \frac{(D-t)(s-t)}{(t+1)p\left(\lfloor \ell/2 \rfloor - D - s + t + 1\right)} \leq 1,$$
since $p\lfloor \ell/2 \rfloor \to \infty$, and consequently $\varphi(t+1) \leq \varphi(t)$. Thus, by (11)
$$\sum_{r,s=0}^{g-1} \mathbb{P}(B^C_{j,\ell,r,s}) \leq O\left(\frac{\mathbb{P}(I_j = 1)\mathbb{P}(I_\ell = 1)}{np}\right).$$
Equation (10) thus yields, for some constant K,
$$\mathbb{P}(I_j I_\ell = 1) \leq \mathbb{P}(I_j = 1)\mathbb{P}(I_\ell = 1)\left(1 + \frac{K}{np}\right).$$
This proves Lemma 2.2. □

Returning to (7), using (8), we see that for another constant L,
$$\begin{aligned} d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Po}(\lambda)) &\leq L\left(\frac{\alpha np^2}{2}\right)^{g-1} e^{-\alpha np^2/2} + \frac{K}{\lambda np}\sum_j \mathbb{P}(I_j = 1)\sum_\ell \mathbb{P}(I_\ell = 1) \\ &= L\left(\frac{\alpha np^2}{2}\right)^{g-1} e^{-\alpha np^2/2} + \frac{K\lambda}{np}. \end{aligned} \tag{12}$$
Thus X may be approximated by a Poisson random variable provided that $np^2 \to \infty$ and $\lambda \ll np$. The first condition may be seen to hold if, e.g.,
$$p \gg \sqrt{\frac{\log\log n}{n}},$$
and the second if the λ given by Lemma 2.1 is (roughly speaking) of order smaller than $np \sim \sqrt{n \log n}$. We have thus established Theorem 1.9 for a range of p's that spans part of the $\lambda \to 0$ and $\lambda \to \infty$ regimes; the full theorem, including the delicate behavior at the threshold, follows easily by monotonicity (e.g., if λ is even larger than np then it is even less likely that $\mathbb{P}(X = 0)$, so that this quantity tends to zero as well).

3. Open Question

Establishing an analog of Theorem 1.9 for $h \geq 3$ would, of course, be of great interest. Combating the fact that h sums are not disjoint is the main technical hurdle we would need to overcome.

Acknowledgments. The research of all three authors was supported by NSF Grant 1263009. We are grateful for the meticulous refereeing job of the referee, which has vastly improved the paper, both in form and in content.

References

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

[2] A. Barbour and L. Holst, Some applications of the Stein-Chen method for proving Poisson convergence, *Adv. Appl. Probab.* **21** (1989), 74–90.

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

[4] P. Erdős and A. Rényi, On a classical problem of probability theory, *Magy. Tud. Akad. Mat. Kutató Int. Közl.* **6** (1961), 215–220.

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

[6] K. Ford, Unpublished work (2008).

[7] A. Godbole, T. Grubb, K. Han, and B. Kay, Threshold progressions in a variety of covering and packing contexts. Preprint (2018+).

[8] A. Godbole, S. Gutekunst, V. Lyzinski, and Y. Zhuang, Logarithmic representability of integers as k-sums, *Integers* **15A** (2015), Paper #A5.

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

[10] A. Godbole, C-M. Lim, V. Lyzinski, and N. Triantafillou, Sharp threshold asymptotics for the emergence of additive bases, *Integers* **13** (2013), Paper # A14.

[11] H. Halberstam and K. Roth, *Sequences,*. Springer-Verlag, New York, 1983.

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

[13] L. Holst, On birthday, collectors', occupancy and other classical urn problems, *Int. Stat. Rev.* **54** (1986), 15–27.

[14] J. Kohonen, An improved lower bound for finite additive 2-bases, *J. Number Theory* **174** (2017), 518–524.

[15] B. Lindström, An inequality for B_2 sequences, *J. Combin. Theory* **6** (1969), 211–212.

[16] A. Mrose, Untere schranken für die reichweiten von extremalbasen fester ordnung, *Abh. Math. Semin. Univ. Hambg.* **48** (1979), 118–124.

[17] D. Newman and L. Shepp, The double dixie cup problem *Amer. Math. Monthly* **67** (1960), 58–61.

[18] K. O'Bryant, A complete annotated bibliography of work related to Sidon sequences, *Electron. J. Combin.* (2004), Dynamic Survey #DS11.

[19] G. Yu, A new upper bound for finite additive h-bases, *J. Number Theory* **156** (2015), 95–104.