Asymptotics for dependent sums of random vectors.

C. Cooper
School of Mathematical Sciences
University of North London
London N7 8DB, UK.

September 26, 2005

Abstract

We consider sequences of length m of n-tuples each with k non-zero entries chosen randomly from an Abelian group or finite field. For what values of m does there exist a subsequence which is zero-sum or linearly dependent respectively? We report some results relating to these problems.

1 Introduction

Let $V_{n,k}$ denote the vector space over the integers modulo 2, consisting of n-tuples with k non-zero entries. Given a m-sequence $\boldsymbol{\sigma} = (\mathbf{s}_i, i = 1, ..., m)$ of vectors selected uniformly at random from $V_{n,k}$, how large must m be, before the sequence $\boldsymbol{\sigma}$ contains a linearly dependent subset, with high probability (**whp**[1])?

This problem of linearly dependent subsets of random sequences of vectors over $\mathbb{Z}_2$ can be generalized in at least two ways.

Homogeneous systems of equations. (Problem I) Let $\boldsymbol{\sigma}$ be a random m-sequence consisting of n-tuples with k non-zero entries from an Abelian group $(G, +)$. How large must m be (**whp**), before there is a subsequence $\mathbf{v}_1, ..., \mathbf{v}_M$ such that $\sum_{i=1}^{M} \mathbf{v}_i = \mathbf{0}$?

Linearly dependent sets of vectors. (Problem II) Let $\boldsymbol{\sigma}$ be a random m-sequence of vectors each with k of non-zero entries from a finite field $(F, +, \times)$. How large must m be (**whp**), before there is a subsequence $\mathbf{v}_1, ..., \mathbf{v}_M$ such that $\sum_{i=1}^{M} c_i \mathbf{v}_i = \mathbf{0}$ where $(c_i \in F, c_i \neq 0)$?

[1] **whp**, with high probability. With probability tending to 1 as $n \to \infty$

The first linear dependence of a random sequence has been studied by several authors, including Balakin, Kolchin and Khokhlov [3, 4], Calkin [5, 6], Kolchin [10], and Kolchin and Khokhlov [11]. For $\mathbb{Z}_2$ and $k = 2$ the solution $m \leq n/2$ is well known. It is equivalent to the emergence of the first cycle in random graphs $G_{n,m}$. For $\mathbb{Z}_2$ and $k \geq 3$, Kolchin [10] gives typical lower bounds on m for finite k. As $k \to \infty$, Calkin [5] obtained lower bounds on m asymptotic to $n(1 - e^{-k}/\log_e 2)$. In a subsequent paper [6], Calkin obtained a similar lower bound, $n(1 - (q-1)e^{-k}/\log_e q)$, for linearly dependent sums of random vectors whose k non-zero entries are sampled uniformly from $GF(q) \setminus \{0\}$. Molloy [14] reports upper bounds based on the 2-core of random k-uniform hypergraphs. For denser matrices, Kovalenko, Levitskaya and Savchuk [12] give many results including exhaustive solutions for GF(q) and moments for finite rings. Kovalenko [13] also reports the following results for $n \times n$ random matrices $A = (a_{ij})$ with independent entries.

(i) For GF(q): If $\Pr(a_{ij} = r) = 1/q,\ r \in GF(q)$ then

$$\Pr(A \text{ is linearly independent }) \to \prod_{j=1}^{\infty}(1 - q^{-j}).$$

(ii) *'Kolmogorov's Problem.'* For GF(2): If $\Pr(a_{ij} = 1) = (1 + \Delta_{ij})/2, |\Delta_{ij}| \leq 1 - \delta$,

$$\Pr(A \text{ is linearly independent }) \to \prod_{j=1}^{\infty}(1 - 2^{-j}) \text{ as in (i) above.}$$

Although it seems natural to approach Problem II through Problem I this does not generally appear to have been the case for sparse matrices, and the successful analyses of Calkin are surprisingly indirect. We take the direct approach which gives insight into Problem I, and sharper lower bounds for Problem II. We consider the following cases.

Problem (I)(i) Each n-tuple $\mathbf{v}$ has k non-zero entries which are a fixed element γ of an Abelian group G. The order of γ is t.

Problem (I)(ii) Each n-tuple $\mathbf{v}$ has k non-zero entries selected uniformly at random from the non-zero elements of an Abelian group G, where $|G| = t$.

Problem (II) Each n-tuple has k non-zero entries selected uniformly at random from the non-zero elements of a finite field F, where $|F| = t$.

For either problem, we call such a dependent subsequence $(\mathbf{v}_1, ..., \mathbf{v}_M)$ a *zero-sum* M-subsequence of an m-sequence.

Before proceeding, we summarize the main parameters used in the paper: n is vector length, $m = \alpha n$ is sequence length, M is subsequence length, k is the number of entries per vector (constant or tending to infinity), $c = kM/n$ is the average occupancy per row of the subsequence (viewed as a matrix), t is a positive integer, all row occupancies are remaindered (modulo t).

Let $\omega = \exp 2\pi i/t$, be a t-th root of unity, let

$$f(x) = \begin{cases} \frac{1}{t}\left(e^x + e^{\omega x} + ... + e^{\omega^{t-1}x}\right) & \text{Problem (I)(i)} \\ \frac{1}{t}\left(e^x + (t-1)e^{-\frac{x}{t-1}}\right) & \text{Problems (I)(ii) and (II)} \end{cases} \tag{1}$$

and let x be given by

$$c(x) = x\frac{f'(x)}{f(x)}.$$

The following theorem gives a lower bound on the critical value of m, when k is small.

Theorem 1 *Let $t = 2$ and $k \geq 3$ or let $t \geq 3$ and $k \geq 2$. Provided $m < m^* = \alpha^* n$, then a random m-sequence contains no zero-sum subsequences,* **whp**. *Here α^* is the smallest strictly positive solution to the following system of equations.*

Problem (I)

$$\begin{aligned} \left(\left(\tfrac{c}{x}\right)^k + 1\right)^\alpha e^{-c} f(x) &= 1 \\ \tfrac{c}{k}\left(1 + \left(\tfrac{x}{c}\right)^k\right) &= \alpha \\ x\tfrac{f'(x)}{f(x)} &= c \end{aligned}$$

Problem (II)

$$\begin{aligned} \left((t-1)\left(\tfrac{c}{x}\right)^k + 1\right)^\alpha e^{-c} f(x) &= 1 \\ \tfrac{c}{k}\left(1 + \tfrac{1}{t-1}\left(\tfrac{x}{c}\right)^k\right) &= \alpha \\ x\tfrac{f'(x)}{f(x)} &= c. \end{aligned}$$

For $t = 2$, Theorem 1 gives the following lower bounds on $\alpha = m/n$ for linear dependence as a function of k. These values are similar to those given in Kolchin [10].

	$k=3$	4	5	6	7	8	9	10
m/n lower	0.8895	0.9672	0.9892	0.9962	0.9987	0.99951	0.99982	0.99993
m/n upper	0.9404	0.9801	0.9930	0.9974	0.9991	0.99966	0.99987	0.99996

The trivial upper bound (**whp**) of $\alpha \sim 1 - e^{-\alpha k}$, given above for comparison, comes from counting the number of empty rows of the matrix of the m-sequence.

If $k \to \infty$ we can simplify the lower bounds as follows.

Theorem 2 *Let $m = \alpha n$. Let α_t be the largest non-negative solution of*

$$\alpha_t = \begin{cases} \log_2 t - \log_2\left(1 + \sum_{j=1}^{t-1} \exp\left(-\frac{\alpha_t k}{2}(1 - \cos\frac{2\pi j}{t})\right)\right) & \textit{Problem (I) (i)} \\ \log_2 t - \log_2\left(1 + (t-1)e^{-\frac{\alpha_t}{2}k\frac{t}{t-1}}\right) & \textit{Problem (I) (ii)} \\ 1 - \log_t\left(1 + (t-1)e^{-\alpha_t k}\right) & \textit{Problem (II).} \end{cases}$$

There exists a positive constant $k_0(t)$ such that, if $k \geq k_0$ then provided $\alpha \leq \alpha_t$, a random m-sequence contains no zero-sum subsequence **whp**.

For Problem (II), we see that $\alpha_t \sim 1 - \frac{(t-1)e^{-k}}{\log_e t}$ which is a result obtained by Calkin [5] [6]. The approach for finite k to the asymptotic solution is extremely rapid. For $t = 2$ the results of Theorem 1, Theorem 2 and Calkin's asymptotic differ in their estimates of m only in the 7-th decimal place by the time $k = 10$.

For Problem I, provided $k \to \infty$ the lower bound for critical m tends to $n \log_2 t$. If k is order $\log n$ we can find a similar upper bound as follows. We choose $M = m/2$, a value which asymptotically maximises the expected number of M-subsequences. The problem now depends on a single parameter $c = km/2n$.

Theorem 3 *Let m be even, t be constant, $cn = km/2$, and in Problem I(i), let t divide $km/2$. Let $\tau(n) \to \infty$ arbitrarily slowly and let*

$$c_t = \begin{cases} \frac{1}{(1-\cos\frac{2\pi}{t})}(\log n + \tau(n)) & \textit{Problem (I)(i)} \\ \\ \frac{(t-1)}{t}(\log n + \tau(n)) & \textit{Problem (I)(ii).} \end{cases}$$

Provided $km/2 \geq c_t n$ (and $c \leq n^{1/7}$) then

(i) *the number W of $m/2$-subsequences of an m-sequence has expected value given by*

$$\mathbf{E}W = t\sqrt{\frac{2}{m\pi}}\frac{2^m}{t^n}(1 + o(1)),$$

(ii) *provided $m \geq n(\log_2 t) + \log n/(2\log 2) + \tau(n)$, a random m-sequence contains a zero-sum $m/2$-subsequence* **whp**.

Theorem 4 *Let $m = n(\log_2 t) + \log n/(2\log 2) + d(n)$, let $k = 2c_t n/m$. Let $\boldsymbol{\sigma}$ be a random m-sequence, then*

$$\Pr(\boldsymbol{\sigma} \textit{ has a zero-sum } m/2\textit{-subsequence }) = \begin{cases} 0 & d(n) \to -\infty \\ 1 & d(n) \to \infty. \end{cases}$$

From Theorem 2, and these theorems, we see that for Problem I, when $k \to 2c_t/(\log_2 t)$ the threshold value of m to have *any* zero-sum subsequence is about $(\log_2 t)n$. Of course, for $t = 2$ there is a deterministic upper bound of $m = n + 1$ for any k.

In the subsequent sections we give detailed proofs for Problem I(i), as these require the most care. We outline the proofs for the other cases where they do not differ substantively.

We prove the results using an occupancy model depending only on $c = km/2n$. Thus, in Problem I(i) (with $\gamma = 1$) we can fix the value of k and let $m \to \infty$ instead. The proof works equally well for equations adding to $r \pmod t$, see Theorem 9. If we consider the m-sequence as a simple hypergraph process with k-uniform edges, some small adjustments, (see end of Appendix B) give the following theorem.

Theorem 5 *Let k, t, l be positive integers, m even, and $nr + lt = km/2$. Let $m \geq 2n(\log n + \tau(n))/(k(1 - \cos\frac{2\pi}{t}))$ where $\tau(n) \to \infty$.* **whp** *random hypergraphs $G_{n,m,k}$ have a spanning subgraph on $m/2$ edges in which all vertex degrees are congruent to r (mod t).*

For example, if $t = k = 2$, then provided n, m are even, random graphs $G_{m,n}$ with $m = \frac{n}{2}(\log n + \tau(n))$ edges have both an even and an odd degree spanning subgraph with $m/2$ edges. We note that this value of m is the threshold in $G_{n,m}$ for minimum degree 1, **whp**. Of course, many of the vertices in the even degree subgraph may have degree 0. If we wish the minimum degree of the even subgraph to be 2, we can modify the proof, to prohibit vertices of degree 0. However, perhaps surprisingly this does not improve upon $m = n(\log n + \tau(n))$ which is obtained by taking $t = 4, r = 2$.

2 Some Models for Problem I(i)

2.1 Occupancy Model

Let $\mathcal{V}(n, L)$ denote the space $[n]^L$ with all elements equiprobable. A sequence $\mathbf{v} \in \mathcal{V}$ is the result of randomly placing L labelled balls in n labelled boxes. If $L = kM$, then given $\mathbf{v} \in [n]^{kM}$, we may partition $\mathbf{v}$ as follows

$$\begin{aligned} \mathbf{v} &= (v_{11}, ..., v_{1k},\ v_{21}, ..., v_{2k},\ ...,\ v_{M1}, ...,\ v_{Mk}) \\ &= (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_M) \end{aligned}$$

where $\mathbf{v}_j$ is the j-th *column* of $\mathbf{v}$. If a column has no repeated entries we say it is *simple*. If all columns are simple, we say $\mathbf{v}$ is a simple vector, and denote this subspace by $\mathcal{S}(n, k, M)$. There are $(k!)^M$ distinct elements of $\mathcal{S}(n, k, M)$ corresponding to each $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}(M)$.

Proofs in this paper use $\mathcal{V}(n, kM)$, with $\mathbf{v}$ written in column form $(\mathbf{v}_1, ..., \mathbf{v}_M)$ representing the M-sequence. To get back to $\mathcal{S}(n, k, M)$, we need the following result, proved in Appendix B.

Lemma 6 *If Theorems 1-4 are true in $\mathcal{V}(n, km)$, they are true in $\mathcal{S}(n, k, m)$ and hence $\boldsymbol{\Sigma}(m)$.*

Let $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, ..., \boldsymbol{\sigma}_m) \in \boldsymbol{\Sigma}(m)$. Define the occupancy a_j of row j of $\boldsymbol{\sigma}$ as $a_j = |\{i : \sigma_{ij} \neq 0\}|$. Let $\boldsymbol{\tau} = (\mathbf{v}_1, ..., \mathbf{v}_M)$ be a subsequence of $\boldsymbol{\sigma}$. If $\mathbf{w} = \sum_{i=1}^{M} \mathbf{v}_i$ and $\mathbf{a}(\boldsymbol{\tau})$ is the occupancy vector of $\boldsymbol{\tau}$ then

$$\mathbf{w} = \mathbf{0} \iff \mathbf{a}(\boldsymbol{\tau}) \equiv \mathbf{0} (\text{mod } t).$$

That is, the sequence $\boldsymbol{\tau}$ is zero-sum iff all occupancies in $\mathbf{a}(\boldsymbol{\tau})$ are multiples of t.

Let $\pi_t(M) = \Pr(\mathbf{a}(\boldsymbol{\tau}) \equiv \mathbf{0})$ and let $W = W(m, M)$ denote the number of zero-sum M-subsequences of an m-sequence, so that

$$\mathbf{E}W(m, M) = \binom{m}{M}\pi_t(M). \tag{2}$$

Let $\omega = \exp 2\pi i/t$, be a t-th root of unity and let $kM = cn$. Let

$$f(x) = \frac{1}{t}\Big(e^x + e^{\omega x} + ... + e^{\omega^{t-1}x}\Big)$$

and let

$$\begin{aligned} c(x) &= x\frac{f'(x)}{f(x)} \\ \sigma^2(c) &= x^2\frac{f''(x)}{f(x)} - c^2 + c. \end{aligned}$$

Theorem 7 *The probability $\pi_t(M)$, that a random M-sequence $\boldsymbol{\tau} \in \mathcal{V}(n, kM)$ is zero-sum, is*

$$\pi_t(M) = (1 + o(1))\ t\ \sqrt{\frac{c}{\sigma^2(c)}}\Big(\Big(\tfrac{c}{xe}\Big)^c f(x)\Big)^n.$$

2.2 Estimation of $\pi_t(M)$

To estimate $\pi_t(M)$, we use a technique suggested by Pittel [15], which was similarly applied in [2]. Let $T = kM$. Let $A = \{a = (a_1, ..., a_n) : \sum a_j = T, a_j \equiv 0(\text{mod } t),\ j = 1, ..., n\}$, thus

$$\pi_t(M) = \frac{1}{n^T}\sum_{a\in A}\binom{T}{a_1...a_n} = \frac{T!}{n^T}\sum_{a\in A}\Big(\prod_{j=1}^n \tfrac{1}{a_j!}\Big).$$

If $f(z) = \sum_{j=0}^\infty \frac{z^{jt}}{(jt)!}$, then provided $x > 0$,

$$\sum_A \Big(\prod \tfrac{1}{a_j!}\Big) = [z^T](f(z))^n = \frac{(f(x))^n}{x^T}[z^T]\Big(\tfrac{f(zx)}{f(x)}\Big)^n. \tag{3}$$

Let $g(z) = \frac{f(zx)}{f(x)}$, then $g(1) = 1$, and $g(z)$ is the probability generating function of a conditional Poisson $Po(x)$ random variable Y. We have

$$\mathbf{E}z^Y = \frac{f(zx)}{f(x)},\ \text{and}\ \mathbf{E}Y = \frac{xf'(x)}{f(x)},\ \text{and}\ \mathbf{E}Y(Y-1) = x^2\frac{f''(x)}{f(x)}.$$

Let $T = kM = cn$ where t divides kM. If we choose x so that $\mathbf{E}Y = c$ then

$$c = x\frac{f'(x)}{f(x)} = x\frac{\sum_{j=1}^{\infty}\frac{x^{jt-1}}{(jt-1)!}}{\sum_{j=0}^{\infty}\frac{x^{jt}}{(jt)!}}, \tag{4}$$

and

$$\sigma^2(c) = x^2\frac{f''}{f} - c^2 + c.$$

Let $Y_1, ..., Y_n$ be i.i.d copies of Y. Then

$$\begin{aligned}[z^T]\Big(\tfrac{f(zx)}{f(x)}\Big)^n &= [z^T]\,\mathbf{E}(z^{Y_1+...+Y_n}) \\ &= \Pr(Y_1 + ... + Y_n = T).\end{aligned}$$

As $T = \mathbf{E}(Y_1 + ... + Y_n) = cn$ we can use the Local Limit Theorem (see Appendix A) to deduce that

$$\Pr(Y_1 + ... + Y_n = T) = (1 + o(1))\frac{t}{\sqrt{2\pi\sigma^2(c)n}}. \tag{5}$$

Thus,

$$\pi_t = (1 + o(1))\, t\,\sqrt{\frac{c}{\sigma^2(c)}}\,\Big[\Big(\tfrac{c}{xe}\Big)^c\Big(\textstyle\sum_{j=0}^{\infty}\frac{x^{jt}}{(jt)!}\Big)\Big]^n. \tag{6}$$

2.3 A closed form for $f(x)$

Suppose $f(u)$ is the generating function of a sequence $R = (p_0, p_1, ..., p_j, ...)$. Let $\omega = \exp 2\pi i/t$, then

$$f_0(u) = \frac{1}{t}(f(u) + f(\omega u) + ... + f(\omega^{t-1}u)) \tag{7}$$

generates the subsequence $R_0 = (p_j : j \equiv 0(\text{mod } t))$. This can be seen by expanding the right hand side of $f_0(u)$.

For Problem I(i) we let $R = (\frac{x^l}{l!}, l = 0, 1, ...)$ and

$$S_{t,0}(ux) = f_0(u) = \sum_{j\geq 0}\frac{(ux)^{jt}}{(jt)!}, \tag{8}$$

so that

$$S_{t,0}(ux) = \frac{1}{t}\Big(e^{ux} + e^{\omega ux} + ... + e^{\omega^{t-1}ux}\Big), \tag{9}$$

and $f_0(1) = S_{t,0}(x)$ is the function denoted by $f(x)$ in Section 2.2.

If $t = 2$ then more simply, $S_{2,0}(x) = \frac{1}{2}(e^x + e^{-x})$.

Finally, to extract the subsequence $R_r = (p_j : j \equiv r(\text{mod } t))$ we use

$$f_r(u) = \frac{1}{t}(f(u) + \omega^{t-r} f(\omega u) + ... + (\omega^{t-1})^{t-r} f(\omega^{t-1} u)). \tag{10}$$

This completes the proof of Theorem 7.

2.4 Equivalent Poisson Model

We will often use the following alternative approach to estimate $\pi_t(M)$. Rather than 'distort' the distribution $f(x)$ to ensure that the expected value is c in (4), we can choose $x = c$ and use the 'equivalent Poisson model'. Specifically let X_j be a system of i.i.d Poisson(c) random variables and P_c be the probability of an event in this system.

$$\begin{aligned} \pi_t(M) = \sum_{t|a_j} \binom{kM}{a_1 ... a_n} \frac{1}{n^{kM}} &= \frac{P_c(t \text{ divides } X_i,\ i = 1, ..., n, \text{ and } \sum X_i = kM)}{P_c(\sum X_i = kM)} \quad (11) \\ &= (1 + o(1))\sqrt{2\pi cn}\Big(e^{-c}\Big(\sum_{j=0}^{\infty} \tfrac{c^{jt}}{(jt)!}\Big)\Big)^n \\ &\quad \times \Pr(\sum X_i = cn \mid X_i \equiv 0(\text{mod } t),\ i = 1, .., n). \quad (12) \end{aligned}$$

3 Proof of Theorems 1 and 2 for Problem I(i)

3.1 Case of small subsets M

Lemma 8 *Let $k \geq 3$ when $t = 2$ or $k \geq 2$ when $t \geq 3$. Let $m = \alpha n,\ \alpha \leq 1$. Let $\boldsymbol{\sigma} \in \mathcal{V}(n, km)$, then* **whp** $\boldsymbol{\sigma}$ *contains no zero-sum M-subsequence*

(i) for any M, provided $m \leq \alpha_0 n,\ \alpha_0 = 1/3k$,

(ii) for any $M \leq c_0 n/k$, where $c_0(\alpha, t)$ is a positive constant.

Proof Let $\mathbf{v} \in \mathcal{V}(n, kM)$ be zero-sum subsequence of $\boldsymbol{\sigma}$. Consider the occupancy vector $\mathbf{a}(\mathbf{v})$. Suppose there are s boxes j with $a_j > 0$, so that $s \leq kM/t \leq n$. The expected number μ of zero-sum M-subsequences of $\boldsymbol{\sigma}$ satisfies

$$\mu \leq \binom{m}{M}\binom{n}{s}\Big(\tfrac{s}{n}\Big)^{kM} \leq \Big(\tfrac{\alpha ne}{M}\Big)^M \Big(e^s \Big(\tfrac{s}{n}\Big)^{kM-s}\Big).$$

For fixed M, the right hand side is monotone increasing in s up to $s = ne^{kM/s}$ a value of $s \geq ne^t$. On substituting $s = kM/t$ and $M = \theta nt/k$ where $\theta \leq \alpha k/t$ we find

$$\mu \leq \left[\left(\frac{\alpha k}{t}\right)^{\frac{t}{(t-1)k}} e^{\frac{t}{(t-1)k} + \frac{1}{(t-1)}} \theta^{1 - \frac{t}{(t-1)k}} \right]^{(1-\frac{1}{t})kM}.$$

To prove (i) note that the upper bound on μ is maximized when $\theta = \alpha k/t$.

To prove (ii) we note that constant values of $c_0 = t\theta$ may be found for constant k below which $\mu = o(1/n)$. If $k \to \infty$ then we may choose $\theta = e^{-1/(t-1)}(1-\delta)$. □

3.2 Proof of Theorem 1

Let $m = \alpha n$, $kM = cn$. From (2) the expected number $\mathbf{E}W(m, M)$ of zero-sum M-subsequences of an m sequence satisfies

$$\mathbf{E}W(\alpha n, \frac{cn}{k}) = \binom{\alpha n}{\frac{cn}{k}} \pi_t\Big(\tfrac{cn}{k}\Big) = ((1 + o(1))G(c, \alpha))^n,$$

where $G(c, \alpha) = \mu(c, \alpha)g(c)$ and

$$\mu(c, \alpha) = \frac{\alpha^\alpha}{(c/k)^{c/k}(\alpha - c/k)^{\alpha - c/k}}, \tag{13}$$

$$g(c) = \left(\tfrac{c}{xe}\right)^c f(x),$$

and

$$c = x\frac{f'(x)}{f(x)}. \tag{14}$$

For fixed α,

$$\frac{d\mu}{dc} = \mu(c) \log\left(\tfrac{\alpha k - c}{c}\right)^{1/k},$$

and

$$\frac{dg}{dc} = \frac{\partial g}{\partial c} + \frac{\partial g}{\partial x}\frac{dx}{dc} = g(c)\,\log(c/x),$$

as $\frac{\partial g}{\partial x} = 0$ follows from (14), so that

$$\frac{dG(c)}{dc} = G(c) \log\left[\left(\tfrac{\alpha k - c}{c}\right)^{1/k} \frac{c}{x}\right].$$

For fixed α the extrema of $G(c, \alpha)$ are at values of c satisfying

$$\alpha = \frac{c}{k}\Big(1 + \left(\tfrac{x}{c}\right)^k\Big). \tag{15}$$

Substituting (15) into $G(c,\alpha)=1$ gives the equations

$$\left(\left(\tfrac{c}{x}\right)^k+1\right)^\alpha e^{-c}f(x)=1. \tag{16}$$

As $c\to 0, x\to((t-1)!c)^{1/t}$ (from (4)). For small c,

$$\left(\tfrac{\alpha k-c}{c}\right)^{1/k}\frac{c}{x}=c^{1-1/t-1/k}\frac{(\alpha k)^{1/k}}{((t-1)!)^{1/t}},$$

which tends to zero provided $1/t+1/k<1$. Thus for small $c, G'(c)<0$. Moreover, $G(0,\alpha)=1$. Thus, for fixed α as c increases away from 0, the first extremum, *if any*, is a minimum. From Lemma 8 we need only consider $c\geq c_0,\ \alpha\geq\alpha_0$, where $\alpha_0,\ c_0$ are small positive constants.

The smallest positive solution α^*, of (14), (15), (16) gives a value of $m=\alpha^* n$ above which $\mathbf{E}W\to\infty$.

3.3 Proof of Theorem 2

It is clear from equation (15) that as $x/c\to 1,\ c\to\alpha k/2$, but finding a good estimate of the error is somewhat messy. The probability of a cell occupancy divisible by t in the Poisson(c) model of Section 2.4 is

$$\begin{aligned}e^{-c}\sum_{j\geq 0}\frac{c^{jt}}{(jt)!} &= \frac{1}{t}\Big(1+e^{-c}\Big(e^{c\omega}+e^{c\omega^2}+\ldots+e^{c\omega^{t-1}}\Big)\Big)\\ &\leq \frac{1}{t}\Big(1+\textstyle\sum_{p=1}^{t-1}e^{-c(1-\cos\frac{2\pi p}{t})}\Big),\end{aligned}$$

where $\omega=e^{2\pi i/t}$. This inequality follows because $\omega^{t-p}=\overline{\omega^p}$ so that

$$e^{c\omega^p}+e^{c\overline{\omega^p}}=2\cos\Big(c\ \sin\tfrac{2\pi p}{t}\Big)e^{c\cos\frac{2\pi p}{t}}\ \leq\ 2e^{c\cos\frac{2\pi p}{t}}.$$

From (12) an upper bound $E_U(W)$ on $\mathbf{E}W(mM)$, the expected number W of m-sequences with a zero-sum subset of size M, is

$$\begin{aligned}E_U(W) &= \binom{\alpha n}{\frac{cn}{k}}\sqrt{2\pi cn}\Big(\tfrac{1}{t}\Big(1+\textstyle\sum_{p=1}^{t-1}e^{-c(1-\cos\frac{2\pi p}{t})}\Big)\Big)^n \\ &= ((1+o(1))F(\epsilon))^n,\end{aligned} \tag{17}$$

where $c=\frac{\alpha k}{2}(1-\epsilon)$ and $\epsilon\in[-1,1]$. From (13), we substitute $\mu(c,\alpha)$ so that

$$F(\epsilon)=\left(\frac{2}{(1-\epsilon)^{\frac{1}{2}(1-\epsilon)}(1+\epsilon)^{\frac{1}{2}(1+\epsilon)}}\right)^\alpha\frac{1+\sum_{p=1}^{t-1}\exp-\beta_p\frac{\alpha k}{2}(1-\epsilon)}{t},$$

and where $\beta_p = \beta_{t-p} = 1 - \cos 2\pi p/t$.

Let $\frac{d\log F(\epsilon)}{d\epsilon} = \frac{\alpha}{2}H(\epsilon)$, then

$$H(\epsilon) = k\frac{\sum \beta_p e^{-\beta_p \frac{\alpha k}{2}(1-\epsilon)}}{1+\sum e^{-\beta_p \frac{\alpha k}{2}(1-\epsilon)}} - \log\frac{1+\epsilon}{1-\epsilon}. \tag{18}$$

$H(0) > 0$ always, and there are two relevant solutions to $H(\epsilon) = 0$, namely ϵ_0 (a maximum) and ϵ_1 (a minimum) whose approximate values, when k is suitably large, are given below.

(i) If $\epsilon \to 0$ then $\log\frac{1+\epsilon}{1-\epsilon} \sim 2\epsilon$ and we find that

$$\epsilon_0 \sim \frac{k}{2}\frac{\sum \beta_p e^{-\beta_p \frac{\alpha k}{2}}}{1+\sum e^{-\beta_p \frac{\alpha k}{2}}}. \tag{19}$$

(ii) Let $\epsilon = 1 - \lambda\delta$ and $\delta = \frac{2\log 2k\beta_1}{\beta_1 \alpha k}$, (or $\frac{1}{\alpha k}\log k$ if $t = 2$), then

$$H(\epsilon) = \frac{1}{(2k\beta_1)^{\lambda-1}}\Big(1 + \Theta\Big(\Big(\tfrac{1}{k}\Big)^{\lambda(\beta_2/\beta_1 - 1)}\Big)\Big) - \log\Big(\tfrac{\beta_1\alpha k}{\lambda \log 2k\beta_1}(1 - \tfrac{\lambda\delta^*}{2})\Big).$$

If $\lambda \gg 1$ then and $H(\epsilon) < 0$, and when $\lambda \to 1 - o(1)$, at approximately

$$1 - \epsilon_1 = \begin{cases} \frac{\log k}{\alpha k}\Big(1 - \frac{\log\log 2\alpha k}{\log k}\Big) & t = 2 \\ \\ \frac{2\log 2\beta_1 k}{\beta_1 \alpha k}\Big(1 - \frac{\log\log \beta_1\alpha k}{\log 2\beta_1 k}\Big) & t \geq 3, \end{cases}$$

we see that $H(\epsilon) \geq 0$, and remains so as λ decreases, provided $\lambda \gg \lambda^* = e^{-\Theta(k)}$. By Lemma 8 the largest value of ϵ we need to consider is $\epsilon^* = 1 - \frac{2c_0}{\alpha k}$, which implies a value of $\lambda = \frac{\beta_1 c_0}{\log 2k\beta_1} \gg \lambda^*$. The value of $F(\epsilon^*)$ in (17) is itself bounded above by

$$\Big(\tfrac{\alpha e k}{c_0}\Big)^{c_0/k}\Big(\tfrac{1+(t-1)e^{-c_0(1-\cos\frac{2\pi}{t})}}{t}\Big) < 1,$$

if k is larger than some constant $k_0(t)$.

Thus the first relevant value of $\epsilon \leq \epsilon^*$ for which $F(\epsilon) = 1$ occurs as $\epsilon \to 0$. Expanding $F(\epsilon)$ about $\epsilon = 0$, we have

$$F(\epsilon) \sim \frac{2^\alpha}{t}e^{-\frac{\alpha\epsilon^2}{2} + O(\epsilon^4)}\Big(1 + \textstyle\sum_{p=1}^{t-1} \exp -\beta_p \frac{\alpha k}{2}(1-\epsilon)\Big).$$

Choosing $F(\epsilon_0) = 1$, where ϵ_0 is given by (19), and solving for α we see that

$$\alpha = \Big(\log_2 t - \log_2\Big(1 + \textstyle\sum_{p=1}^{t-1} \exp -\beta_p \frac{\alpha k}{2}(1-\epsilon_0)\Big)\Big)/(1 - \epsilon_0^2/2 + O(\epsilon_0^4)). \tag{20}$$

This value of α is strictly greater than α_t in the statement of Theorem 2.

4 Proof of Theorem 3 for Problem I(i).

We use the Equivalent Poisson Model of Section 2.4, and estimate (12) with suitable accuracy.

4.1 Critical values of c

Generalizing the definitions (8) and (9) of Section 2.3, let

$$S_{t,r}(x) = \sum_{j\geq 0} \frac{x^{jt+r}}{(jt+r)!}, \tag{21}$$

from (10) we see that

$$S_{t,r}(x) = \frac{1}{t}\Big(e^x + \omega^{t-r}e^{\omega x} + ... + \omega^{(t-1)(t-r)}e^{\omega^{t-1}x}\Big). \tag{22}$$

If the cell occupancy is Poisson $Po(c)$, the conditional distribution for the case where the occupancy is congruent to $r(\text{mod } t)$, has expected value d_r, where

$$d_r = c\frac{S_{t,r-1}(c)}{S_{t,r}(c)}. \tag{23}$$

As usual, let $c = kM/n$. We will say that c is *critical* if $d_r = c(1+o(\frac{1}{n}))$, $(0 \leq r \leq t-1)$. Once c is critical,

$$S_{t,r}(c) = (1+o(\frac{t}{n}))S_{t,0}(c) = \frac{1}{t}e^c(1+o(\frac{t}{n})) \tag{24}$$

as, intuitively, the occupancy (mod t) is (asymptotically) uniform on $\{0,1,...,t-1\}$.

The critical values of c are, for $t=2$

$$c_2 = \frac{\log n + \tau(n)}{2}$$

where $\tau(n) \to \infty$ arbitrarily slowly. For $t \geq 3$, d_r oscillates and we choose

$$c_t = \frac{\log n + \tau(n)}{1-\cos\frac{2\pi}{t}}.$$

This follows because

$$(d_r - c) = c\frac{\omega^{t-r}(1-\omega)e^{-(1-\omega)c} + ... + \omega^{(t-1)(t-r)}(1-\omega^{t-1})e^{-(1-\omega^{t-1})c}}{1+\omega^{t-r}e^{-(1-\omega)c} + ... + \omega^{(t-1)(t-r)}e^{-(1-\omega^{t-1})c}},$$

so that

$$\sup|d_r - c| \leq (t-1)ce^{-(1-\cos 2\pi/t)c}(1+\Theta(e^{-(1-\cos 2\pi/t)c})).$$

4.2 Zero-sum M-subsequences

Let $W(M)$ denote the number of zero-sum M-subsequences in a random m-sequence.

To calculate $\pi_t(M)$, we use equation (12). If X_j are i.i.d $Po(c)$ random variables, then provided c is critical, we show in Appendix A that

$$\Pr(\sum_{j=1}^{n} X_j = cn \mid X_j \equiv 0(\text{mod } t), j = 1, ..., n) = \frac{t}{\sqrt{2\pi nc}}(1 + o(1)). \tag{25}$$

We see from equation (24) that $\pi_t = t(\frac{1}{t}(1 + o(\frac{t}{n}))^n$ and thus from (2), (12) and (25) that

$$\mathbf{E}W(M) = \binom{m}{M}\frac{t}{t^n}(1 + o(1)). \tag{26}$$

4.3 Cell occupancies of intersecting subsequences

We consider the joint distribution of cell occupancy for two intersecting zero-sum M-subsequences. Specifically, let S_1, S_2 be M-subsequences with intersection B, so that $S_1 = (A, B),\ S_2 = (B, C)$ say. Let $k|A| = an,\ k|B| = bn,\ k|C| = an$, be fixed and let

$$\rho_t(a, b, a) = \Pr(S_1 = (A, B) \text{ is zero-sum and } S_2 = (B, C) \text{ is zero-sum}).$$

Let (a_j, b_j, c_j) be the occupancy of the j-th row of (A, B, C). As $a_j + b_j \equiv 0 \pmod t$ and $b_j + c_j \equiv 0 \pmod t$ we have

$$(a_j, b_j, c_j) \equiv (r, t - r, r) \pmod t,$$

where $r \in \{0, 1, .., t - 1\}$ are the integer remainders on division by t. In fact,

$$\rho_t(a, b, a) = \sum_{\Omega} \frac{1}{n^{an}}\binom{an}{a_1 \dots a_n}\frac{1}{n^{bn}}\binom{bn}{b_1 \dots b_n}\frac{1}{n^{an}}\binom{an}{c_1 \dots c_n},$$

where $\Omega = \{(a_j, b_j, c_j) \equiv (r, t - r, r)(\text{mod } t) :\ j = 1, ..., n,\ \ r \in \{0, 1, .., t - 1\}\}$.

By a slight generalization of (11) we will replace ρ_t by an equivalent Poisson event. Let $X \sim Po(a),\ Y \sim Po(b),\ Z \sim Po(a)$. Let $X_j,\ (j = 1, ...n)$ be i.i.d copies of X.
Let $\mathcal{E}_j = \{(a_j, b_j) \text{ is zero-sum and } (b_j, c_j) \text{ is zero-sum}\}$. Let $\mathcal{F}$ denote the event that $\{\sum X_j = an,\ \sum Y_j = bn,\ \sum Z_j = an\}$ then

$$\rho_t(a, b, a) = \frac{\Pr(\bigwedge_{j=1,...,n} \mathcal{E}_j)\Pr(\mathcal{F} \mid \bigwedge_{j=1,...,n} \mathcal{E}_j)}{\Pr(\mathcal{F})}.$$

The $\mathcal{E}_j$ are independent, and

$$\begin{aligned}\Pr(\mathcal{E}_j) &= \sum_{r=0}^{t-1}\Big(\sum_{i=0}^{\infty}\tfrac{a^{it+r}}{(it+r)!}\Big)\times\Big(\sum_{i=0}^{\infty}\tfrac{b^{it+(t-r)}}{(it+(t-r))!}\Big)\times\Big(\sum_{i=0}^{\infty}\tfrac{a^{it+r}}{(it+r)!}\Big)e^{-(a+b+a)}\\ &= \Big(\sum_{r=0}^{t-1}S_{t,r}(a)S_{t,t-r}(b)S_{t,r}(a)\Big)e^{-(a+b+a)},\end{aligned}$$

in the notation of (21), and where $(t-0)\equiv 0$ for the case $(0,0,0)$. We note from (22) that the leading non-constant terms in the expansion of the right hand side are $e^{-x+\omega^j x}(1+\omega+...+\omega^{t-1})$, $(x=a,b)$ which vanishes exactly, and $e^{-(1-\omega)(a+b)}$, $e^{-2a+\omega a+\omega^2 a}$.

Let $a=\frac{kM}{2n}(1-\epsilon), b=\frac{kM}{2n}(1+\epsilon)$ and $|\epsilon|\le\theta/\sqrt{M}$, where $\theta\to\infty$ arbitrarily slowly. Thus

$$\Pr(\mathcal{E})=\frac{1}{t^3}\Big(t+2te^{-(1-\omega)c}+O(e^{-c(1\pm\epsilon)(1-(\omega+\omega^2)/2)}+...)\Big),$$

where $c=a+b$ and if c is critical, then $|e^{-(1-\omega)c}|=o(1/n)$, and

$$\Pr(\mathcal{E})=\frac{1}{t^2}(1+o\Big(\tfrac{t}{n}\Big)).$$

As the probabilities are calculated in the Poisson system,

$$\Pr(\mathcal{F})=\frac{(1+o(1))}{\sqrt{(2\pi n)^3 aba}}.$$

We prove in Appendix A (Lemma A2) that

$$\Pr(\mathcal{F}\mid\mathcal{E})=\frac{t^2(1+o(1))}{\sqrt{(2\pi n)^3 aba}}.$$

Hence

$$\rho_t(a,b,a)=t^2\Big(\tfrac{1}{t^2}\Big)^n(1+o(1)). \tag{27}$$

4.4 Proof of Theorem 3: Second Moment method

If two $m/2$-subsequences $S_1,\ S_2$ intersect in $|B|=L$ columns, then

$$\mathbf{E}W(W-1)=\sum_L\binom{m}{\frac{m}{2}-L,\ L,\ \frac{m}{2}-L,\ L}\rho_t(\frac{k}{n}\Big(\tfrac{m}{2}-L\Big),\frac{kL}{n},\frac{k}{n}\Big(\tfrac{m}{2}-L\Big)).$$

The multinomial coefficient above is maximized at $L=m/4$ and has total value $\binom{m}{m/2}^2$. The values of L contributing significantly to this sum occur in the range $m/4\pm\theta\sqrt{m}$ where $\theta\to\infty$ arbitrarily slowly. Under these assumptions, as $km/(2n)$ is critical

$$\mathbf{E}W(W-1)=\binom{m}{\frac{m}{2}}^2\frac{t^2}{t^{2n}}(1+o(1)).$$

Thus provided $\mathbf{E}W \to \infty$ $\Pr(W = 0) = o(1)$. This follows from (26), the statement of Theorem 3 and the Chebychev inequality.

We also have the following generalization for Problem I(i).

Theorem 9 *Let k be constant. Let $km/2 \geq c_t n$. Let r be integer, $0 \leq r \leq t-1$. Let $\boldsymbol{\sigma}$ be a random m-sequence of $0,1$ vectors each with k non-zero entries, and let the addition be (modulo t). Provided there exists a natural number l such that $nr + lt = km/2$, then* **whp** *$\boldsymbol{\sigma}$ contains an $m/2$-subsequence whose sum in each row is congruent to r (mod t).*

5 Proofs for Problem I(ii)

5.1 Models for Problem I(ii)

In this problem, the k non-zero entries are selected uniformly at random from among the non-zero members of an Abelian group $(G, +)$. We express G as the direct sum of cyclically generated subgroups,

$$G = \mathbb{Z}_{t_1} \oplus \mathbb{Z}_{t_2} \oplus \cdots \oplus \mathbb{Z}_{t_l},$$

so that $|G| = t_1 t_2 \cdots t_l = t$.

We regard the rows of a sequence as boxs B as divided into sub-cells indexed by $g \in G \setminus \{0\} = \{g_1, ..., g_{t-1}\}$. If the occupancy of sub-cell B_g is n_g, then element g was selected n_g times in box B. In external direct sum form, $G \cong \Omega = \{(j_1, ..., j_l) : 0 \leq j_q \leq t_q - 1,\ q = 1, ..., l\}$, so that $g = (j_1, ..., j_l)$. The *value* of sub-cell B_g is $n_g g = (n_g j_1, ..., n_g j_l)$.

For brevity, we only consider the Poissonized version. Thus the occupancy $X_{s,g},\ s = 1..n$ of each of the $t-1$ sub-cells in row s is i.i.d Poisson $Po(c/(t-1))$. The joint probability generating function for the occupancy and value of sub-cell $(j_1, j_2, ..., j_l)$ is

$$h_g(c, z, \mathbf{u}) = \exp -\frac{c}{t-1} + \frac{cz}{t-1} u_1^{j_1} u_2^{j_2} \cdots u_l^{j_l}. \tag{28}$$

The probability generating function for the occupancy and total value of box B is

$$h(c, z, \mathbf{u}) = \exp -c + \frac{cz}{t-1}\Big(\prod_{q=1}^{l} \ \Big(1 + u_q + u_q^2 + \cdots + u_q^{t_q - 1}\Big) + (-1)\Big),$$

where the (-1) term removes the sub-cell $(0, 0, ..., 0)$. The assumption here, which is correct for a multinomial model, is that conditional on the occupancy of B being a, each of the a labelled points independently and uniformly selects a sub-cell label g, resulting in a sample for the row which is an a-vector of $g \in G \setminus \{0\}$.

Let $f_{\mathbf{r}}(c,z)$ generate the probability of a cell occupancy whose total value $(\nu_1,...,\nu_l)$ satisfies

$$(\nu_1,\nu_2,...,\nu_l) \equiv (r_1,r_2,...,r_l) \text{ (mod } (t_1,t_2,...,t_l)).$$

Setting $\omega_j = \exp 2\pi i/t_j$ and $\mathbf{r} = (r_1,r_2,...,r_l)$, we find, by a generalization of (10) that

$$\begin{aligned} f_{\mathbf{r}}(c,z) &= \frac{1}{t}\sum_{\Omega}\Big(\prod_{q=1}^{l}\omega_q^{(t_q-r_q)j_q}\Big)h(c,z,\omega_1^{j_1},...,\omega_l^{j_l}) \\ &= \begin{cases} \frac{1}{t}\Big(e^{-c+cz}+(t-1)e^{-c-\frac{cz}{t-1}}\Big) & \mathbf{r}=\mathbf{0} \\ \\ \frac{1}{t}\Big(e^{-c+cz}-e^{-c-\frac{cz}{t-1}}\Big) & \mathbf{r}\neq\mathbf{0}. \end{cases} \end{aligned} \tag{29}$$

5.2 Problem I(ii): Proof of Theorems 1-3

Very much as before, we have

$$f(x) = \frac{1}{t}\Big(e^x+(t-1)e^{-\frac{x}{t-1}}\Big)$$

and let

$$\begin{aligned} c(x) &= x\frac{f'(x)}{f(x)} \\ \sigma^2(c) &= x^2\frac{f''(x)}{f(x)} - c^2 + c. \end{aligned}$$

Theorem 10 *The probability $\pi_t(M)$, that a random M-sequence $\boldsymbol{\tau} \in \mathcal{V}(n,kM)$ is zero-sum, is*

$$\pi_t(M) = (1+o(1))\sqrt{\frac{c}{\sigma^2(c)}}\Big(\Big(\frac{c}{xe}\Big)^c f(x)\Big)^n.$$

Let $\mathbf{t} = (t_1,...,t_l)$ and $\mathbf{B} = \{\boldsymbol{\beta} = (\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,...,\boldsymbol{\beta}_n) : \boldsymbol{\nu}_j \equiv \mathbf{0}(\text{mod } \mathbf{t}),\ (j=1,...,n)\}$ where $\boldsymbol{\beta}_j = (b_{j1},...b_{j(t-1)})$ and $\sum_{j=1}^{n}\sum_{q=1}^{t-1} b_{jq} = kM$ is the total number of balls. Thus

$$\begin{aligned} \pi_t(M) &= \Pr(\mathbf{B}) = \frac{1}{(n(t-1))^{kM}}\sum_{\boldsymbol{\beta}\in\mathbf{B}}\binom{kM}{\boldsymbol{\beta}} \\ &= \frac{(kM)!}{n^{kM}}\sum_{\boldsymbol{\beta}\in\mathbf{B}}\prod_{j=1}^{n}\left[\prod_{g=1}^{t-1}\frac{1}{(t-1)^{b_{jg}}b_{jg}!}\right]. \end{aligned}$$

The individual terms of the product in square brackets above are generated by

$$f_g(z,\mathbf{u}) = \sum_{b\geq 0} \frac{1}{b!}\left(\frac{zu_1^{j_1}...u_l^{j_l}}{t-1}\right)^b,$$

which is $e^{1/(t-1)}h_g(1,z,\mathbf{u})$ in (28). Thus from (29), the product in square brackets is generated by

$$G(z) = f_{\mathbf{0}}(1,z)\, e^1 = \frac{1}{t}\Big(e^z + (t-1)e^{-\frac{z}{t-1}}\Big).$$

If $T = kM = cn$, we proceed as in Section 2.2 from equation (3) downwards, so that

$$\pi_t(M) = \frac{T!}{n^T}\frac{(G(x))^n}{x^T}[z^T]\Big(\tfrac{G(zx)}{G(x)}\Big)^n. \tag{30}$$

Set $c = xG'(x)/G(x)$ and $\sigma^2 = x^2G''(x)/G(x) - c^2 + c$. We use the Local Limit Theorem to deduce that

$$[z^T]\Big(\tfrac{G(zx)}{G(x)}\Big)^n = (1+o(1))\frac{1}{\sqrt{2\pi\sigma^2 n}},$$

for provided $t > 2$, the span of the sequence generated by $G(x)$ is 1. This can be seen by expanding $G(x)$.

The remaining details of the proofs of Theorems 1,2 are much the same as for Problem I(i).

5.3 Proof of Theorem 3

From (29) we see that the expected occupancy of a box with value

$$(\nu_1,...,\nu_l) \equiv (r_1,...,r_l)(\text{mod } (t_1,...,t_l))$$

is

$$d_{\mathbf{r}} = \begin{cases} c\dfrac{\Big(1-e^{-\frac{ct}{t-1}}\Big)}{\Big(1+(t-1)e^{-\frac{ct}{t-1}}\Big)} & \mathbf{r}=\mathbf{0} \\ \\ c\dfrac{\Big(1+\frac{1}{t-1}e^{-\frac{ct}{t-1}}\Big)}{\Big(1-e^{-\frac{ct}{t-1}}\Big)} & \mathbf{r}\neq\mathbf{0}. \end{cases}$$

The critical value of c_t above which $d_{\mathbf{r}} = c(1+o(\frac{1}{n}))$ and $f_{\mathbf{r}}(c,z) \sim f_{\mathbf{0}}(c,z)$ is given by

$$c_t = \frac{t-1}{t}(\log n + \tau(n)).$$

If two zero-sum subsequences $S_1 = (A, B)$ and $S_2 = (B, C)$ have intersection B, then the value restrictions on the j-th row of (A, B, C) are $(\mathbf{r}, \mathbf{t} - \mathbf{r}, \mathbf{r})$ where $\mathbf{t} = (t_1, ..., t_l)$ and $\mathbf{r} = (r_1, ..., r_l)$.

As before let $\mathcal{E}_j$ be the event that the j-th rows of both S_1 and S_2 are zero-sum.

$$\Pr(\mathcal{E}_j) = \sum_{\mathbf{r} \in \Omega} f_{\mathbf{r}}(a, 1) f_{\mathbf{t}-\mathbf{r}}(b, 1) f_{\mathbf{r}}(a, 1),$$

where $f_{\mathbf{r}}(a, 1)$ is given by (29).

Provided c is critical,

$$\Pr(\mathcal{E}_j) = |\Omega| \frac{1}{t^3}(1 + o(\frac{1}{n})) = \frac{1}{t^2}(1 + o(\frac{1}{n})).$$

The other calculations are virtually identical with section 4 , with the exception of the fact that the lattice span is 1 not t in (5) and is $(1, 1, 1)$ not $(1, t, t)$ in the calculation of $\Pr(\mathcal{F} \mid \mathcal{E})$.

5.4 Proofs for Problem (II)

We now consider weighted sums $c_1\boldsymbol{\sigma}_1 + ... + c_M\boldsymbol{\sigma}_M = \boldsymbol{\tau}_1 + ... + \boldsymbol{\tau}_M$, of the sequence $\boldsymbol{\sigma}$, where the c_i are non-zero. As noted by Calkin [3], if the $\sigma_{i,j}$ are selected uniformly at random from the non-zero elements of a finite field F, then so are the $\tau_{i,j}$.

If $F = \mathrm{GF}(p^l)$, then $(F, +)$ is the direct sum of l copies of $\mathbb{Z}_p$, and is a special case of Problem I(ii).

Let $W(m, M)$ denote the number of linearly dependent M-subsequences, then

$$\mathbf{E}W = \binom{m}{M}(t-1)^M \pi_t(M), \tag{31}$$

where $\pi_t(M)$ is given by (30). This leads directly to Theorem 1.

For Theorem 2 we note that $\binom{m}{l}(t-1)^l$ is maximized at $l = (t-1)m/t$ with value $\frac{t}{\sqrt{2\pi(t-1)m}} t^m$. We substitute $c = \alpha k \frac{t-1}{t}(1 - \epsilon)$ into (13) to give a similar proof to that for Problem I(i).

6 Appendix A: The Local Limit Theorem

Let X be a random variable taking values in $\{b + lk : k \in \mathbb{Z}\}$. The *span* h is the largest value of the variable l for which X can be represented in the form $b + l\mathbb{Z}$. We say that X is a *lattice* random variable with *span* h.

Theorem A 1 *[8]* **The Local Limit Theorem**

Let $X_1, ..., X_n$ be i.i.d lattice random variables with span h, $\mathbf{E}X = 0$ and finite variance σ^2. Let $Y = \frac{X_1+...+X_n}{\sigma\sqrt{n}}$ then

$$\sup \left| \frac{\sigma\sqrt{n}}{h} \Pr(Y = y) - \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \right| \to 0,$$

as $n \to \infty$.

Proofs of the Local Limit Theorem (LLT) may be found in [9], [8], and a more recent treatment in [7]. The proofs are for a one dimensional lattice with constant σ^2. Because we want $\sigma \to \infty$, albeit slowly, and also consider three dimensional lattices, we repeat the proof for the special cases in this paper.

Proofs of the LLT are based on the following observation. Let Y be a lattice random variable with span h and characteristic function $\psi(s) = \mathbf{E}e^{isY}$, then

$$\Pr(Y = y) = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-isy} \psi(s) ds.$$

Let $X_j, j = 1, ..., n$ be i.i.d $Po(c)$ random variables. If the characteristic function of conditional occupancy 0 (mod t) is $\varphi(c, s)$,

$$\Pr(\sum_{j=1}^{n} X_j = cn \mid X_j \equiv 0 (\text{mod } t), j = 1, ..., n) = \frac{t}{2\pi} \int_{-\pi/t}^{\pi/t} e^{-iscn} (\varphi(c, s))^n ds. \tag{32}$$

Lemma A 2 *Let X be a random variable with characteristic function $\varphi(c, s) = f(ce^{is})/f(c)$ where $f(x)$ is given by (1). Let $\mathbf{E}X = d = cf'(c)/f(c)$, where $d = c(1 + o(\frac{1}{n}))$, then*

$$\Pr(\sum_{j=1}^{n} X_j = cn \mid X_j \equiv 0 (mod\ t), j = 1, ..., n) = I(t, c) = \frac{t}{\sqrt{2\pi cn}} (1 + o(1)),$$

provided $c = o(\sqrt{n})$.

Proof We use the standard LLT proof, namely: Transform the integral (32) using $s = x/\sqrt{n}$. For $x \in (-A, A)$ where A is a (large) constant, approximate carefully to the central term of a normal distribution with characteristic function $\exp -cx^2/2$. For other values of x, the absolute value of the characteristic function $|\varphi(x/\sqrt{n})^n| = o(n^{-A})$.

$$\begin{aligned} \left| \frac{\sqrt{n}}{t} I(t, c) - \frac{1}{\sqrt{2\pi c}} \right| &\leq \left| \frac{1}{2\pi} \int_{-A}^{A} \left(e^{-ixc\sqrt{n}} \left(\varphi(\tfrac{x}{\sqrt{n}}) \right)^n - e^{-c\frac{x^2}{2}} \right) dx \right| \\ &+ \frac{1}{2\pi} \int_{(-\infty,\infty)\setminus(-A,A)} e^{-c\frac{x^2}{2}} dx + \frac{1}{2\pi} \int_{(-\sqrt{n}\frac{\pi}{t}, \sqrt{n}\frac{\pi}{t})\setminus(-A,A)} \left| \varphi\left(\tfrac{x}{\sqrt{n}}\right)^n \right| dx \\ &= J_1 + J_2 + J_3, \end{aligned}$$

say. To be specific, we consider Problem I(i) for $t = 2$. Thus

$$\varphi(s) = \frac{e^{ce^{is}} + e^{-ce^{is}}}{e^c + e^{-c}} = e^{-c+ce^{is}}\Big(\tfrac{1+e^{-2ce^{is}}}{1+e^{-2c}}\Big), \tag{33}$$

and (see Durret [7] (p84-86)),

$$e^{is} = 1 + is - \frac{s^2}{2} + \theta\frac{s^3}{6},$$

where $|\theta| \le 1$. Hence

$$J_1 \le \frac{1}{2\pi}\int e^{-c\frac{x^2}{2}}\left|e^{c\theta\frac{x^3}{6\sqrt{n}}}\Big(\tfrac{1+e^{-2ce^{ix/\sqrt{n}}}}{1+e^{-2c}}\Big)^n - 1\right| dx.$$

A careful approximation of $y = \Big(\tfrac{1+e^{-2ce^{ix/\sqrt{n}}}}{1+e^{-2c}}\Big)$ shows that $y = 1+e^{-2c}\Theta\Big(\tfrac{c^2x^2}{n}\Big)\exp i\Theta\Big(\tfrac{cx}{\sqrt{n}}\Big)$ provided c is critical and $cx/\sqrt{n} \to 0$. Thus

$$J_1 \le \frac{1}{2\pi}\Theta\Big(\tfrac{cA^3}{\sqrt{n}}\Big)\int_{-\infty}^{\infty} e^{-c\frac{x^2}{2}}dx = \Theta\Big(\tfrac{cA^3}{\sqrt{n}}\Big)\frac{1}{\sqrt{2\pi c}}.$$

Using $\int_A^\infty e^{-c\frac{y^2}{2}}dy \le \frac{1}{cA}e^{-}c\frac{A^2}{2}$, we see that

$$J_2 \le O\Big(n^{-A^2/4}\Big)\frac{1}{\sqrt{2\pi c}}.$$

Now

$$J_3 \le n\Big(\tfrac{e^{c\cos s}+e^{-c\cos s}}{e^c+e^{-c}}\Big)^n \le \Theta\Big(ne^{-cA^2/2}\Big).$$

□

Lemma A 3 *For Problem I(i),*

$$\Pr(\mathcal{F} \mid \mathcal{E}) = \frac{t^2(1+o(1))}{\sqrt{(2\pi n)^3 aba}}.$$

Proof The probability of a point $\mathbf{y} = (y_1, ..., y_r)$ from a lattice distribution with span $(h_1, ..., h_r)$ is

$$\Pr(\mathbf{y}) = \frac{h_1...h_r}{(2\pi)^r}\int_Q e^{-i\mathbf{s}.\mathbf{y}}\Psi(\mathbf{s})d\mathbf{s},$$

where Q is the product of the intervals $[-\pi/h_j, \pi/h_j],\ j = 1, ..., r$.

The lattice span (h_1, h_2, h_3) for the points (a_j, b_j, c_j) contributing to $(\mathcal{F} \mid \mathcal{E})$ is $(1, t, t)$. To see this, note that a_j can be any natural number, but once we have determined that

$a_j \equiv r(\mathrm{mod}\ t)$ then the coordinates b_j (resp. c_j) have span t, because $(a_j, b_j, c_j) \equiv (r, t-r, r)(\mathrm{mod}\ t)$.

Now $\Psi(\mathbf{s}) = (\psi(\mathbf{s}))^n$ where $\psi(\mathbf{s})$ is the characteristic function for the (conditional) occupancy (a_j, b_j, c_j) of cell j. Specifically,

$$\psi(\mathbf{s}) = \frac{\varphi_0 + ... + \varphi_{t-1}}{p_0 + ... + p_{t-1}},$$

where

$$p_r = S_{t,r}(a)S_{t,t-r}(b)S_{t,r}(a),$$

and

$$\varphi_r(\mathbf{s}) = S_{t,r}(ae^{is_1})S_{t,t-r}(be^{is_2})S_{t,r}(ae^{is_3}).$$

This, on expansion, is

$$e^{-(2a+b)}e^{(2a+b)e^{is}\left(\frac{1+2e^{-(1-\omega)ce^{is}}+...}{1+2e^{-(1-\omega)c}+...}\right)}$$

because the leading non-constant terms vanish exactly. □

7 Appendix B: Proof of Lemma 6

Let $L = kM$, $c = L/n$, and let $\mathcal{V}(n, L)$ be $[n]^L$ with the uniform measure. Let A be the set of occupancy vectors of $[n]^L$, thus, $A = \{\mathbf{a} \in [L]^n : \sum_{i=1}^n a_i = L\}$. The subset of $\mathcal{V}$ with occupancy $\mathbf{a}$ defines an occupancy class $\mathcal{V}_\mathbf{a}$. A property Q is an occupancy property if it is the union of occupancy classes. We use $\mathcal{Z}$ to denote 'zero-sum' properties in the sense of Problem I, and $A(\mathcal{Z})$ to denote the occupancy vectors of $\mathcal{Z}$.

An element $\mathbf{s} \in \mathcal{S}(n, k, M)$ gives a unique $\boldsymbol{\sigma}(s) \in \boldsymbol{\Sigma}(M)$ and $k!$ elements of $\mathcal{S}$ map to each element of $\boldsymbol{\Sigma}$. If Q is an occupancy property of M-sequences,

$$\Pr(Q; \boldsymbol{\Sigma}) = \Pr(Q; \mathcal{S}) = \Pr(Q \mid \mathcal{S}; \mathcal{V}),$$

where $\Pr(\mathcal{E}; \Omega)$ is the probability of $\mathcal{E}$ in space Ω.

For constant k and $c \le \beta \log n$, β small, the lower bounds in Theorems 1,2 for the (non)-existence of zero-sum subsequences in $\mathcal{V}$ also hold in $\mathcal{S}$ and hence in $\boldsymbol{\Sigma}(M)$. This follows because

$$\Pr(\mathcal{Z} \mid \mathcal{S}) \le \frac{\Pr(\mathcal{Z})}{\Pr(\mathcal{S})},$$

and

$$\Pr(\mathcal{S}) = \left(\frac{(n)_k}{n^k}\right)^M \sim \exp -\frac{c(k-1)}{2},$$

so that $\Pr(\mathcal{S}) = n^{-\beta(k-1)/2}$ whereas $\Pr(\mathcal{Z}) = e^{-\Theta(n)}$ in the proofs of these theorems.

We will now show that if $c \to \infty$ then

$$\Pr(\mathcal{Z} \mid \mathcal{S}) \sim \Pr(\mathcal{Z})$$

thus proving Theorems 3,4,5. In fact the results are true for constant c, but we do not prove it here.

Let μ denote the expected occupancy of cell j, and let $\sigma^2 = \mathbf{E}(a_j - \mu)^2$.

To convert results about $\mathcal{Z}$ from $\mathcal{V}$ to $\mathcal{S}$ we use the following idea. In the space $\mathcal{V}(n, cn)$ the measure is concentrated on those occupancy vectors $\mathbf{a}$ which lie in a 'spherical shell' centre $\boldsymbol{\mu}$, radius $\sqrt{\sigma^2(\mathcal{V})n}$ and thickness $o(\sqrt{\sigma^2(\mathcal{V})n})$, and a similar result holds in $\mathcal{Z}$. Specifically

$$\mu(\mathcal{V}) = \mu(\mathcal{Z}) = c \qquad \text{by symmetry,} \tag{34}$$

$$\sigma^2(\mathcal{V}) = c - \frac{c}{n}, \tag{35}$$

$$\sigma^2(\mathcal{Z}) = c + c^2\Big(1 - \tfrac{1}{cn}\Big)\frac{\Pr(\text{ occupancy } \equiv (t-2) \pmod t)}{\Pr(\text{ occupancy } \equiv 0 \pmod t)} - c^2. \tag{36}$$

Once $\sigma(\mathcal{Z}) \to \sigma(\mathcal{V})$, most of the measure in the two spaces is concentrated within the same shell. Within this shell, the probability a sequence is simple (or graphic) does not vary significantly as a function of the occupancy $\mathbf{a}$.

Let h be a suitably large constant. We define a set $U \subset A$ of *usual* occupancy vectors, by

U(I) No occupancy a_j is greater than $h\mu$,

U(II) $\sum_{j=1}^{n}(a_j - \mu)^2 \in n\sigma^2\Big(1 \pm ch^{5/2}\sqrt{\frac{\log n}{n}}\Big)$.

Lemma B 1 *Let $L = cn$, and let $c = \beta \log n$. Provided $c \geq c_t$, property U holds in $\mathcal{V}, \mathcal{Z}$ with probability $1 - r$, where $r = o(n^{1-h})$.*

Proof We use the equivalent Poisson model to estimate conditional multinomial probabilities $\Pr(B \mid \mathcal{A}, L)$, where $\mathcal{A} = C \times ... \times C$. The random variables X_i are now independent conditional Poisson$(\mu \mid C)$, and

$$\Pr(B \mid \mathcal{A},\ L) \leq \frac{P_\mu(B \mid \mathcal{A})}{P_\mu(L \mid \mathcal{A})}, \tag{37}$$

where we will estimate $P_\mu(L \mid \mathcal{A})$ using the Local Limit Theorem.

Case of $\mathcal{V}$. The occupancy X of any cell is $B(L, 1/n)$, with expectation $\mu = c$ and $\sigma^2 = c(1 - 1/n)$. The probability that $X \geq hc$ is at most $n^{-h\beta \log h/e}$, by the Chernoff inequality. Thus $\Pr(\overline{U(I)}) = o(n^{1-h})$ provided we choose $\beta \log h/e \geq 1$.

For U(II) we condition on U(I), and use (37). The subset of $\mathcal{V}$ with occupancies bounded by hc is represented by iid truncated Poisson random variables X with expectation $c(1-o(1/n))$. Thus, by the LLT, $\Pr(\sum X_i = cn \mid U(I)) \sim \frac{1}{\sqrt{2\pi nc}}$.

Let $Y = (X-c)^2$, then $\mathbf{E}Y = \sigma^2 - o(c^2/n)$, and $0 \le Y \le (h-1)^2c^2$. Thus if $Y_1, ..., Y_n$ are iid Y, we can use the Hoeffding inequality. Let $t = n\sigma^2 h^{5/2} c\sqrt{\frac{\log n}{n}}$, then

$$\begin{aligned}\Pr\left(|\sum Y_i - n\mathbf{E}Y| \ge t \mid U(I)\right) &\le 2\exp -\frac{2t^2}{n((h-1)c)^4} \\ &= o(n^{-h}).\end{aligned}$$

Case of $\mathcal{Z}$. For U(I), the occupancy X of any cell can be modelled by a conditional Poisson distribution. In the case of Problem I(i) for example, we proceed, using (37) and $\Pr\left(\sum X_j = L \mid A(\mathcal{Z})\right) \sim \frac{1}{\sqrt{2\pi nc}}$ from the Local Limit Theorem, (25). Provided $(a-1)t \ge c$,

$$\frac{c^{at}}{at!} \le \frac{1}{t} \sum_{j=(a-1)t+1}^{at} \frac{c^j}{j!},$$

so that

$$\sum_{at \ge hc} \frac{c^{at}}{at!} \frac{e^{-c}}{e^{-c}S_{t,0}(c)} \le \frac{1}{t} \sum_{j \ge hc-t} \frac{c^j}{j!} \frac{e^{-c}}{\frac{1}{t}(1+o(1))}.$$

Thus

$$\Pr(\exists X \ge hc \mid A(\mathcal{Z})) \le n(1+o(1)) \Pr(Y \ge hc - t)$$

where $Y \sim Po(c)$ and we can use large deviation bounds for the Poisson distribution (see [1]). For U(II), we proceed exactly as in the *Case of $\mathcal{V}$*. □

Let $\mathcal{V}_{\mathbf{a}}$ be the set of vectors with occupancy $\mathbf{a}$, and $\mathcal{S}_{\mathbf{a}}$ the simple vectors with that occupancy. For *usual* occupancies $\mathbf{a} \in U$, the ratio $|\mathcal{S}_{\mathbf{a}}|/|\mathcal{V}_{\mathbf{a}}|$ is more or less constant, as we now show. We first confirm the intuitively obvious fact that more balls in one box means the proportion of simple vectors is reduced.

Lemma B 2 *Let $\mathbf{a} = (a_r)$. Given i, j such that $a_i \ge a_j$, define $\mathbf{a}'$ by $a'_i = a_i + 1$, $a'_j = a_j - 1$ and $a'_r = a_r$ if $r \ne i, j$.*

Let $\mathbf{s} \in \mathcal{S}_{\mathbf{a}}$, and let $\eta_{ij}(\mathbf{s}) = \eta(\mathbf{s})$ be the number of columns of $\mathbf{s}$ which contain both an i and j entry, and let $\overline{\eta}$ be the expected value of η in $\mathcal{S}_{\mathbf{a}}$. Then

$$\frac{|\mathcal{S}_{\mathbf{a}}|}{|\mathcal{V}_{\mathbf{a}}|}\left(1 - \tfrac{\overline{\eta}}{a_j}\right) = \frac{|\mathcal{S}_{\mathbf{a}'}|}{|\mathcal{V}_{\mathbf{a}'}|}\left(1 - \tfrac{\overline{\eta}}{a_i+1}\right).$$

Proof Consider the bipartite graph B with bipartition $(\mathcal{S}_{\mathbf{a}}, \mathcal{S}_{\mathbf{a}'})$. If $\mathbf{s} \in \mathcal{S}_{\mathbf{a}}$ and one of the j-entries of $\mathbf{s}$ can be altered to i to give a *simple* $\mathbf{t}$ in $\mathcal{S}_{\mathbf{a}'}$, then $\mathbf{s},\ \mathbf{t}$ are joined by an edge in B.

If $\mathbf{t} \in \mathcal{S}_{\mathbf{a}'}$ is a neighbour of $\mathbf{s}$, then $\mathbf{t}$ then $\eta(\mathbf{s}) = \eta(\mathbf{t})$. For if $s_l = j$ is altered to $t_l = i$, and the result $\mathbf{t}$ is simple then i was not an entry in the column of $s_l = j$ in $\mathbf{s}$.

Thus B consists of components $D = (C, C')$ with η constant on each component. In D, either $a_j = \eta$ so that $C' = \emptyset$ and $d(\mathbf{s}) = 0$, or

$$d(\mathbf{s}) = a_j - \eta, \qquad d(\mathbf{t}) = a_i + 1 - \eta,$$

thus always

$$|C|(a_j - \eta) = |C'|(a_i + 1 - \eta).$$

However , $|\mathcal{V}_{\mathbf{a}}|(a_j) = |\mathcal{V}_{\mathbf{a}'}|(a_i + 1)$, so that

$$\frac{|C|}{|\mathcal{V}_{\mathbf{a}}|}\Big(1 - \tfrac{\eta}{a_j}\Big) = \frac{|C'|}{|\mathcal{V}_{\mathbf{a}'}|}\Big(1 - \tfrac{\eta}{a_i+1}\Big). \tag{38}$$

□

Lemma B 3 *Let $a_i \geq a_j$ and let $\overline{\eta_{ij}}$ be the average value of $\eta_{ij}(\mathbf{s})$ for $\mathbf{s} \in \mathcal{S}_{\mathbf{a}}$. If $\mathbf{a} \in U(I)$, then*

$$\overline{\eta_{ij}} = \frac{(k-1)a_i a_j}{L}\Big(1 + O\Big(\tfrac{kh}{n}\Big)\Big).$$

Proof We imagine the i-balls already placed in a_i distinct columns, and we sequentially add the j-balls, maintaining the simplicity of the resulting vector. Let X_l be an indicator variable for the event that the l-th j-ball lands in a column containing an i-ball. Let $N_0 = 0$ and $N_l = X_1 + ... + X_l$, then

$$\Pr(X_l = 1) = \frac{(k-1)(a_i - N_{l-1})}{L - a_i - k(l-1) + N_{l-1}}.$$

Using $N_l \leq l$ on the left hand side below, we find inductively that

$$\frac{(k-1)a_i}{L}\Big(1 - O\Big(\tfrac{kl}{L}\Big)\Big) \leq \mathbf{E}X_l \leq \frac{(k-1)a_i}{L}\Big(1 + O\Big(\tfrac{kl+a_i}{L}\Big)\Big).$$

Now $\overline{\eta_{ij}} = \mathbf{E}N_{a_j}$, and by U(I) $a_i, a_j \leq hc$. The result follows.

□

Lemma B 4 *Let $\mathbf{b} \in U$, and let $\boldsymbol{\mu} = (\mu : i = 1, ..., n)$ be the expected occupancy, suitably rounded. Provided $\mu \to \infty$,*

$$\frac{|\mathcal{S}_{\mathbf{b}}|}{|\mathcal{V}_{\mathbf{b}}|} = \frac{|\mathcal{S}_{\mu}|}{|\mathcal{V}_{\mu}|}e^{-\frac{k-1}{2}}(1 + o(1)).$$

Proof For given $\mathbf{b}$, let $I^- = \{i : b_i \le \mu\},\ I^+ = \{i : b_i > \mu\}$. We construct a path from $\boldsymbol{\mu}$ to $\mathbf{b}$ in A by moving balls one at a time from cells of $\mathcal{V}_{\boldsymbol{\mu}}$ indexed in I^- in $\mathbf{b}$ to cells of $\mathcal{V}_{\boldsymbol{\mu}}$ indexed in I^+ in any (feasible) way. The number of balls moved, ∂, is given by

$$\begin{aligned}\partial &= \sum_{i\in I^-} |\mu - b_i| \\ &= \tfrac{1}{2}\sum_{i=1}^{n} |b_i - \mu| \\ &\le \tfrac{1}{2}(n\textstyle\sum(b_i-\mu)^2)^{\frac{1}{2}} \\ &< \sqrt{nL},\end{aligned}$$

using U(II) and $\sigma^2 = c(1-o(1))$. Let $d = \frac{k-1}{L}\Big(1 + O\Big(\frac{kh}{n}\Big)\Big)$ and let $a_i' = a_i + 1,\ a_j' = a_j - 1$. By Lemma B 2,

$$\frac{|\mathcal{S}_{\mathbf{a}'}|}{|\mathcal{V}_{\mathbf{a}'}|} = \frac{|\mathcal{S}_{\mathbf{a}}|}{|\mathcal{V}_{\mathbf{a}}|}\frac{(1-da_i)}{\Big(1-da_j\Big(\frac{a_i}{a_i+1}\Big)\Big)} = \frac{|\mathcal{S}_{\mathbf{a}}|}{|\mathcal{V}_{\mathbf{a}}|}\frac{(1-da_i)}{(1-da_j)}\frac{1}{(1+\epsilon)}$$

where $\epsilon = O(d) = \frac{da_j}{(a_i+1)(1-da_j)}$ as $a_j \le a_i$ and $da_j = o(1)$. Hence

$$\begin{aligned}\frac{|\mathcal{S}_{\mathbf{b}}|}{|\mathcal{V}_{\mathbf{b}}|} &= \frac{|\mathcal{S}_{\mu}|}{|\mathcal{V}_{\mu}|}\left[\frac{\prod_{i\in I^+}\prod_{l=\mu}^{b_i-1}(1-dl)}{\prod_{i\in I^-}\prod_{l=b_i+1}^{\mu}(1-dl)}\right]\frac{1}{(1+\epsilon)^{\partial}}, \\ &= \frac{|\mathcal{S}_{\mu}|}{|\mathcal{V}_{\mu}|}\frac{P}{(1+\epsilon)^{\partial}},\end{aligned}$$

and $(1+\epsilon)^{-\partial} = \exp -(k-1)O\Big(\sqrt{\frac{n}{L}}\Big)$. We now simplify P.

$$\begin{aligned}P &= \frac{\prod_{i\in I^+}\exp -\frac{d}{2}((b_i(b_i-1)-\mu(\mu-1)) + O(d^2b_i^3)}{\prod_{i\in I^-}\exp -\frac{d}{2}(\mu(\mu+1)-(b_i(b_i+1)) + O(d^2b_i^3)} \\ &= \exp -\frac{d}{2}(\textstyle\sum_{i=1}^n(b_i^2-\mu^2) - \sum_{i=1}^n|\mu-b_i| + O(dh^3\mu^3 n)) \\ &= \exp\frac{d}{2}\Big(L\Big(1+O\Big(ch^{5/2}\sqrt{\tfrac{\log n}{n}}\Big)\Big) - O\Big(\sqrt{nL}\Big) + O(k\tfrac{L^2}{n^2})\Big) \\ &= \exp -\frac{k-1}{2}(1+o(1)).\end{aligned}$$

The third line follows from U(II) as $\sum b_i = L$ and because $\partial \le \sqrt{nL}$. □

These lemmas now allow us to switch the results of the relevant Theorems from $\mathcal{V}$ to $\mathcal{S}$ as follows.

Lemma B 5 *Provided c is critical, $\frac{|\mathcal{Z}\cap\mathcal{S}|}{|\mathcal{S}|} \sim \frac{|\mathcal{Z}|}{|\mathcal{V}|}$.*

Proof Provided $\sigma(\mathcal{Z}) \to \sigma(\mathcal{V})$ in (34) which it does when c is critical we can choose (two) values of h so that $\mathcal{Z}(U) \subset \mathcal{V}(U)$. Within $\mathcal{V}(U)$, all occupancy classes have asymptotically the same proportion of simple vectors, and thus

$$\frac{|\mathcal{Z}\cap\mathcal{S}(U)|}{|\mathcal{Z}(U)|} = (1+o(1))\frac{|\mathcal{S}(U)|}{|\mathcal{V}(U)|}.$$

so that

$$\frac{|\mathcal{Z}\cap\mathcal{S}(U)|}{|\mathcal{S}(U)|} = (1+o(1))\frac{|\mathcal{Z}|}{|\mathcal{V}|}.$$

Now

$$|\mathcal{S}(U)| \le |\mathcal{S}| \le |\mathcal{S}(U)|\Big(1+e^{k/2}\tfrac{|\mathcal{V}(\overline{U})|}{|\mathcal{V}|}\Big).$$

From Lemma B 1 $\frac{|\mathcal{V}(\overline{U})|}{|\mathcal{V}|} \le n^{1-h}$. If $k \le c \le n^{1/7}$ is fixed, we can choose $h = 2 + k/\log n$ so that both $e^{k/2}n^{1-h} = o(1)$ and $ch^{5/2}\sqrt{\frac{\log n}{n}} = o(1)$. □

Say a sequence in $\mathcal{V}(n, kM)$ is graphic if no two columns have the same underlying set. If $k \ge 3$ then m-sequences are graphic **whp** provided $m = o(n^{3/2})$.

For $k = 2$, some further analysis which follows closely the proofs above ensures the results of the appropriate theorems hold **whp** for graphic sequences. Specifically, let $\mathcal{G}$ be the sequences which are simple and graphic, then for $k = 2$, it can be shown that

$$\frac{|\mathcal{G}_{\mathbf{b}}|}{|\mathcal{V}_{\mathbf{b}}|} = \frac{|\mathcal{G}_{\mu}|}{|\mathcal{V}_{\mu}|}e^{-\frac{c+1}{2}}(1+o(1)).$$

This means that a version of Lemma B5 holds for appropriate values of h (etc), namely

$$\frac{|\mathcal{Z}\cap\mathcal{G}|}{|\mathcal{G}|} \sim \frac{|\mathcal{Z}|}{|\mathcal{V}|}.$$

References

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

[2] B. Bollobás, C. Cooper, T. Fenner and A. Frieze. *Edge disjoint Hamilton cycles in sparse random graphs of minimum degree at least k.*

[3] G. V. Balakin, V. F. Kolchin and V. I. Khokhlov. *Hypercycles in a random hypergraphs.* Diskretnaya Mathematica 3.3 (1991)

[4] G. V. Balakin, V. F. Kolchin and V. I. Khokhlov. *Hypercycles in a random hypergraph.* Discrete Maths Appl. 2 563-570 (1992)

[5] N. Calkin. *Dependent sets of constant weight binary vectors.* Combinatorics Probability and Computing (to appear)

[6] N. Calkin. *Dependent sets of constant weight vectors in $GF(q)$.* Random structures and Algorithms 9 49-54 (1996).

[7] R. Durrett. *Probability: Theory and Examples.* Wadsworth & Brooks/Cole (1991)

[8] B. V. Gnedenko. *Theory of Probability.* Chelsea (NY) (1963)

[9] B. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley (1954)

[10] V. F. Kolchin *Random graphs and systems of linear equations in finite fields.* Random Structures and Algorithms 5 135-146 (1994)

[11] V. F. Kolchin and V I. Khokhlov *On the number of cycles in a non-equiprobably random graph.* Diskretnaya Mathematica 2.3 137-145 (1990)

[12] I. Kovalenko, A. Levitskaya and Savchuk. *Selected Problems in Probabilistic Combinatorics.* Naukova Dumka (Kiev) (In Russian).

[13] I. Kovalenko. Private Communication (1997)

[14] M. Molloy. Private Communication (1995)

[15] B. Pittel. Private Communication (1993)