Proc. London Math. Soc. (3) 00 (2018) 1–49 © 2018 London Mathematical Society
doi:10.1112/plms.12223

A lower bound for the k-multicolored sum-free problem in $\mathbb{Z}_m^n$

László Miklós Lovász and Lisa Sauermann

Abstract

In this paper, we give a lower bound for the maximum size of a k-colored sum-free set in $\mathbb{Z}_m^n$, where $k \geqslant 3$ and $m \geqslant 2$ are fixed and n tends to infinity. If m is a prime power, this lower bound matches (up to lower order terms) the previously known upper bound for the maximum size of a k-colored sum-free set in $\mathbb{Z}_m^n$. This generalizes a result of Kleinberg–Sawin–Speyer for the case $k = 3$ and as part of our proof we also generalize a result by Pebody that was used in the work of Kleinberg–Sawin–Speyer. Both of these generalizations require several key new ideas.

1. *Introduction*

In 2016, Ellenberg and Gijswijt [13] made an enormous breakthrough on the 'cap-set problem'. This problem asks about the largest size of a subset of $\mathbb{F}_3^n$ that does not contain a three-term arithmetic progression. Ellenberg and Gijswijt [13] proved that any such set has size at most $o(2.756^n)$. Their proof uses a new polynomial method developed by Croot, Lev, and Pach [10] for the analogous problem in $\mathbb{Z}_4^n$. The preprint of Ellenberg and Gijswijt appeared just a few weeks after the one of Croot, Lev, and Pach, and subsequently a lot more activity evolved around these new ideas (see [5, 8, 11, 12, 15–17, 19, 21, 23–28, 31–35]).

Soon after the preprint of Ellenberg and Gijswijt [13] appeared, Blasiak, Church, Cohn, Grochow, Naslund, Sawin, and Umans [8] and independently Alon noticed that the argument of Ellenberg and Gijswijt can also be used to obtain an upper bound on the size of k-colored sum-free sets in $\mathbb{F}_p^n$ with $k = 3$ (see the following definition)[†].

Definition 1.1. Let G be an abelian group and let $k \geqslant 3$. A *k-colored sum-free set* in G is a collection of k-tuples $(x_{1,j}, x_{2,j}, \dots, x_{k,j})_{j=1}^{L}$ of elements of G such that for all $j_1, \dots, j_k \in \{1, \dots, L\}$

$$x_{1,j_1} + x_{2,j_2} + \cdots + x_{k,j_k} = 0 \quad \text{if and only if} \quad j_1 = j_2 = \cdots = j_k.$$

The size of a k-colored sum-free set is the number of k-tuples it consists of.

Alon, Shpilka, and Umans [2] introduced the notion of k-colored sum-free sets in the case $k = 3$ and studied its connections to certain approaches for fast matrix multiplication algorithms.

On his blog, Tao [35] published a reformulation of the proof of Ellenberg and Gijswijt, in which he introduced what was later called the slice rank of a tensor. Tao's slice rank method immediately gives upper bounds for the size of k-colored sum-free sets in $\mathbb{F}_p^n$ for any $k \geqslant 3$. In order to state these upper bounds, set

$$\Gamma_{m,k} = \min_{0<\gamma<1} \frac{1 + \gamma + \cdots + \gamma^{m-1}}{\gamma^{(m-1)/k}}$$

Received 5 May 2018; revised 21 November 2018.

2010 *Mathematics Subject Classification* 05D40, 11B30 (primary), 05E15 (secondary).

This research was supported by NSF Postdoctoral Fellowship Award DMS-1705204.

[†] Here, and whenever we write $\mathbb{F}_p$ throughout this introduction, we assume p to be prime. Since all problems we consider only use the additive structure of $\mathbb{F}_p^n$, it is not interesting to consider $\mathbb{F}_p$ for prime powers p instead of primes. We will, however, consider $\mathbb{Z}_m$ where m is a prime power, or any integer.

for integers $m \geqslant 2$ and $k \geqslant 3$. Note that $\Gamma_{m,k} < m$ (since at $\gamma = 1$ the function has value m and positive derivative). Furthermore, the function tends to infinity when $\gamma \to 0$, hence the minimum value $\Gamma_{m,k}$ is indeed attained for some $0 < \gamma_{m,k} < 1$ (one can show that there is a unique $0 < \gamma_{m,k} < 1$ where the minimum is attained, but this is not necessary for our purposes).

Tao's slice rank method [**35**], together with the arguments from Blasiak *et al.* [**8**], gives the following upper bound for the size of k-colored sum-free sets in $\mathbb{Z}_m^n$ for prime powers m. For the reader's convenience, we give a proof of Theorem 1.2 in Section 9 (see also [**26**, Theorem 4], which is a very similar theorem).

Theorem 1.2. *For every prime power m and every integer $k \geqslant 3$, the size of any k-colored sum-free set in $\mathbb{Z}_m^n$ is at most $(\Gamma_{m,k})^n$.*

Our main result is the following lower bound for the maximum size of a k-colored sum-free set in $\mathbb{Z}_m^n$, where $k \geqslant 3$ and $m \geqslant 2$ are fixed and n tends to infinity. If m is a prime power, this lower bound matches (up to lower order terms) the upper bound in Theorem 1.2. Thus, we essentially determine the maximum size of a k-colored sum-free set in $\mathbb{Z}_m^n$, if m is a fixed prime power and n tends to infinity.

Theorem 1.3. *Let $m \geqslant 2$ and $k \geqslant 3$ be fixed integers. Then, there exists a k-colored sum-free set in $\mathbb{Z}_m^n$ with size at least $(\Gamma_{m,k})^{n-O(\sqrt{n})}$.*

Note that in Theorem 1.3, the constant factor of the $O(\sqrt{n})$-term does depend on m and k.

The case $(k, m) = (3, 2)$ in Theorem 1.3 was proved by Fu and Kleinberg [**18**], building on work of Coppersmith and Winograd [**9**] in the context of fast matrix multiplication algorithms. After Blasiak *et al.* [**8**] established the upper bound, Kleinberg, Sawin, and Speyer [**24**] proved Theorem 1.3 for $k = 3$ and any $m \geqslant 2$. Their proof uses a statement that had been formulated as a conjecture in an earlier version of their paper and was then proved by Pebody [**31**]. In order to make the statement precise, we need some more notation.

For integers $r \geqslant 0$ and $\ell \geqslant 2$, set

$$T_{r,\ell} = \{(a_1, a_2, \ldots, a_\ell) \in \mathbb{Z}^\ell \mid a_1 + \cdots + a_\ell = r,\ a_1, \ldots, a_\ell \geqslant 0\}.$$

Given a probability distribution τ on $T_{r,\ell}$, one obtains ℓ probability distributions on the set $\{0, \ldots, r\}$ by taking the projections to the different coordinates. For an ℓ-tuple $(a_1, \ldots, a_\ell)$ and a permutation $\sigma \in S_\ell$, let $(a_1, \ldots, a_\ell)^\sigma = (a_{\sigma(1)}, \ldots, a_{\sigma(\ell)})$ be the ℓ-tuple obtained from $(a_1, \ldots, a_\ell)$ by permuting the coordinates according to σ. A probability distribution τ on $T_{r,\ell}$ is called *S_ℓ-symmetric* if $\tau(a_1, \ldots, a_\ell) = \tau((a_1, \ldots, a_\ell)^\sigma)$ for all $(a_1, \ldots, a_\ell) \in T_{r,\ell}$ and all $\sigma \in S_\ell$. For an S_ℓ-symmetric probability distribution τ on $T_{r,\ell}$, the ℓ projections to the individual coordinates all give the same probability distribution $\mu(\tau)$ on $\{0, \ldots, r\}$ and this distribution is called the *marginal* of τ. For every $a \in \{0, \ldots, r\}$, the distribution $\mu(\tau)$ satisfies

$$\mu(\tau)(a) = \sum_{\substack{a_2, \ldots, a_\ell \in \{0, \ldots, r\} \\ a + a_2 + \cdots + a_\ell = r}} \tau(a, a_2, \ldots, a_\ell).$$

For integers $m \geqslant 2$ and $k \geqslant 3$, let $\nu_{m,k}$ be the probability distribution on $\{0, \ldots, m-1\}$ given by

$$\nu_{m,k}(i) = \frac{\gamma_{m,k}^i}{1 + \gamma_{m,k} + \cdots + \gamma_{m,k}^{m-1}}.$$

The probability distribution $\nu_{m,k}$ has expectation $(m-1)/k$ and entropy $\log \Gamma_{m,k}$ (see Lemma 3.2). One can also show that among all probability distributions on $\{0, \ldots, m-1\}$

with expectation $(m-1)/k$, the distribution $\nu_{m,k}$ has the maximum entropy, which gives some motivation for considering this particular distribution.

The conjecture in the first version of the paper [24] of Kleinberg, Sawin, and Speyer stated that for every $m \geqslant 2$, the probability distribution $\nu_{m,3}$ occurs as the marginal of an S_3-symmetric probability distribution on $T_{m-1,3}$. As mentioned above, this was proved by Pebody [31]. Norin [29] also proposed a proof.

In order to prove Theorem 1.3, we need a generalization of the result of Pebody to $k > 3$. The following theorem generalizes a slightly stronger version of the statement to $k > 3$.

THEOREM 1.4. *For all integers $m \geqslant 2$ and $k \geqslant 3$, the probability distribution $\nu_{m,k}$ occurs as the marginal of an S_k-symmetric probability distribution $\tau_{m,k}$ on $T_{m-1,k}$ with $\tau_{m,k}(t) > 0$ for every $t \in T_{m-1,k}$.*

Our proof of Theorem 1.4 was inspired by the first version of Pebody's proof [30] of the conjecture of Kleinberg, Sawin, and Speyer. However, our proof is not a direct generalization of Pebody's work and if one restricts to $k = 3$ in our proof, one obtains a significantly different proof. In particular, we avoid the large case analysis in [30], which would become even larger when trying to generalize it to $k > 3$. Pebody later replaced the first version of his proof by yet another, much shorter proof [31]. However, it seems to be difficult to find an equally short and clean argument for the case $k > 3$.

Using the upper bound on the size of 3-colored sum-free sets in $\mathbb{F}_p^n$, Fox and the first author [15] proved a polynomial bound for the arithmetic triangle removal lemma in $\mathbb{F}_p^n$. The result of Kleinberg, Sawin, and Speyer [24] then implies that the exponent in the bound is sharp. In the case of $k > 3$, polynomial bounds for the arithmetic k-cycle removal lemma in $\mathbb{F}_p^n$ were given by Fox and both authors [16]. Similarly as in the triangle case, Theorem 1.3 implies lower bounds on the possible exponents in the arithmetic k-cycle removal lemma in $\mathbb{F}_p^n$, but they do not match the exponents that Fox and the authors obtained in [16]. It would be interesting to close this gap and determine the optimal exponent for the arithmetic k-cycle removal lemma in $\mathbb{F}_p^n$.

Arithmetic removal lemmas were introduced by Green in 2005 [20], and since then the problem of improving the bounds in arithmetic removal lemmas has been widely studied [6, 7, 14, 15, 18, 22]. This is in part due to the close connection to property testing. Indeed, the construction of Fu and Kleinberg [18] establishing Theorem 1.3 in the case $(k, m) = (3, 2)$, as well as earlier work of Bhattacharya and Xie [7] in this direction, were presented as proving limits on the possible efficiency of randomized algorithms that test triangle-freeness in $\mathbb{F}_2^n$. Theorem 1.3 gives a limit on the possible efficiency of testing k-cycle-freeness in $\mathbb{F}_p^n$. For more details see [7, 18].

It is worth noting that the proof of Theorem 1.3 is not a straightforward generalization of the work of Kleinberg, Sawin, and Speyer [24] for the case $k = 3$. Their argument uses a random sampling process to find a large 3-colored sum-free set within a certain collection of 3-tuples. Although our approach for proving Theorem 1.3 is the same as in [24], serious challenges arise with the probabilistic sampling argument. The main difficulty in generalizing the work in [24] arises in proving Proposition 3.10, which states that there cannot be too many special pairs of k-tuples with increased conditional probabilities in the sampling process. This fact makes the probabilistic sampling argument work. The proposition has a much simpler proof in the case $k = 3$ than in the case $k > 3$. The most crucial tool for the proof of the proposition for $k > 3$ is an entropy inequality that we will introduce in Section 4. Furthermore, in the case of $k = 3$, one also only needs a much weaker version of Proposition 3.3, and the corresponding argument is just a single paragraph in [24].

A rough outline of the proof of Theorem 1.3 is as follows. We first reduce Theorem 1.3 to a similar statement for k-tuples of vectors in $\mathbb{Z}^n$ summing to $(m-1)\cdot \mathbb{1}^n$, the vector with each coordinate equal to $m-1$. We then start with a set X_0 of vectors in $\{0,\dots,m-1\}^n$ such that for each vector in X_0, the distribution of its entries is roughly the distribution $\nu_{m,k}$ on $\{0,\dots,m-1\}$ that we defined above. Using Theorem 1.4, we can ensure that there are k-tuples $(x_1,\dots,x_k)\in X_0^k$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$. We would like to find a large collection $(x_{1,j},x_{2,j},\dots,x_{k,j})_{j=1}^L$ of such k-tuples in X_0^k such that the only solutions to $x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k}=(m-1)\cdot\mathbb{1}^n$ are $j_1=\dots=j_k$.

In order to do so, we perform a random sampling argument as in [**24**]. More specifically, we consider carefully chosen random subsets $X_1,\dots,X_k$ of X_0. Roughly speaking, $X_1,\dots,X_k\subseteq X_0$ are obtained from a k-colored sum-free set in $\mathbb{Z}_P$ for a carefully chosen prime P via considering inverse images under certain randomly chosen affine-linear maps $\mathbb{Z}^n\to\mathbb{Z}_P$. We then consider those k-tuples $(x_1,\dots,x_k)\in X_1\times\dots\times X_k$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$. We prove that, in expectation, there is a large number of 'isolated' such k-tuples, that is, k-tuples $(x_1,\dots,x_k)\in X_1\times\dots\times X_k$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$ that do not share the same element x_i with any other such k-tuple in $X_1\times\dots\times X_k$ (for any i). These isolated k-tuples will form the desired collection $(x_{1,j},x_{2,j},\dots,x_{k,j})_{j=1}^L$ of k-tuples of vectors in $\mathbb{Z}^n$. Although this is the same strategy as for $k=3$ in [**24**], proving that the expected number of isolated k-tuples is large is much harder for $k>3$ than for $k=3$. The argument for $k>3$ crucially relies on our new entropy inequality in Section 4.

This paper is organized as follows. The first part of the paper is devoted to proving Theorem 1.3 assuming Theorem 1.4. We start with some preliminaries about entropy in Section 2. Afterward, we give the proof of Theorem 1.3 in Section 3, but we postpone several lemmas and propositions to Sections 4 and 5. In particular, the proof of Proposition 3.10 is the main difficulty in the first part of this paper and takes up all of Section 4. In the second part of the paper, starting from Section 6, we prove Theorem 1.4. Finally, in Section 9, we give a proof of Theorem 1.2 for the reader's convenience.

Notation. All logarithms are base e. The set of nonnegative real numbers is denoted by $\mathbb{R}_{\geqslant 0}$, and $\mathbb{1}^n$ denotes the all-ones vector with n entries. For any integer a, let $\mathbf{1}_a$ denote the indicator function of a. That is, $\mathbf{1}_a(x)=1$ if $x=a$ and $\mathbf{1}_a(x)=0$ otherwise.

2. *Preliminaries on entropy*

This section covers some preliminaries about entropy. Some facts are stated without proof, their proofs can be found, for example, in [**3**, Chapter 15.7]. All random variables in this section are assumed to be random variables defined on a finite ground set.

Given a probability distribution ω on a finite set S, the *entropy* of ω is defined as

$$\mathrm{H}(\omega)=\sum_{s\in S}-\omega(s)\log\omega(s).$$

Note that $\mathrm{H}(\omega)\geqslant 0$. With a slight abuse of notation, we will also write $\mathrm{H}(X)$ instead of $\mathrm{H}(\omega)$, if X is a random variable on S with distribution ω. Given several random variables $X_1,X_2,\dots,X_k$, we will write $\mathrm{H}(X_1,X_2,\dots,X_k)$ for the entropy of the joint distribution of the variables $X_1,X_2,\dots,X_k$ (which is a distribution on S^k).

Given a finite set S, the uniform distribution on S is the (unique) distribution ω on S with maximum entropy. So for every probability distribution ω on S, we have $\mathrm{H}(\omega)\leqslant\log(|S|)$. For a given finite set S, entropy is a concave function on probability distributions on S. In other words, for any two probability distributions ω_0 and ω_1 on S and any real number $0\leqslant t\leqslant 1$, we have

$$\mathrm{H}(t\omega_1+(1-t)\omega_0)\geqslant t\,\mathrm{H}(\omega_1)+(1-t)\,\mathrm{H}(\omega_0).$$

Suppose that Y is a random variable on a finite set S, and X is a random variable on any finite set. Then, the *conditional entropy* $\mathrm{H}(X \mid Y)$ is defined as

$$\mathrm{H}(X \mid Y) = \sum_{s\in S} \mathbb{P}(Y = s)\, \mathrm{H}(X \mid Y = s).$$

Here, $\mathrm{H}(X \mid Y = s)$ denotes the entropy of the conditional distribution of $(X|Y = s)$. One can show that

$$\mathrm{H}(X \mid Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y), \tag{2.1}$$

and therefore

$$\mathrm{H}(X, Y) \geqslant \mathrm{H}(Y).$$

Given a sequence of random variables $X_1, X_2, \dots, X_m$, repeatedly applying (2.1) yields

$$\mathrm{H}(X_1, X_2, \dots, X_m) = \mathrm{H}(X_1) + \mathrm{H}(X_2 \mid X_1) + \cdots + \mathrm{H}(X_m \mid X_1, X_2, \dots, X_{m-1}).$$

For any random variables X, Y, Z, we have

$$\mathrm{H}(X \mid Y, Z) \leqslant \mathrm{H}(X \mid Y).$$

If X and Y are random variables such that Y is completely determined by X, we have

$$\mathrm{H}(X) = \mathrm{H}(X, Y) \geqslant \mathrm{H}(Y) \tag{2.2}$$

and

$$\mathrm{H}(X \mid Y) = \mathrm{H}(X, Y) - \mathrm{H}(Y) = \mathrm{H}(X) - \mathrm{H}(Y). \tag{2.3}$$

If τ is an S_ℓ-symmetric probability distribution on $T_{r,\ell}$ for some $r \geqslant 0$ and $\ell \geqslant 2$, then the marginal $\mu(\tau)$ is the projection of τ to the first coordinate. So by (2.2), we have

$$\mathrm{H}(\mu(\tau)) \leqslant \mathrm{H}(\tau). \tag{2.4}$$

The following lemma is a well-known approximation of multinomial coefficients, see, for example, [**24**, Lemma 3].

Lemma 2.1. *Let ω be a probability distribution on a finite set S, and let n be a positive integer. Assume that for every $s \in S$, the probability $\omega(s)$ is an integer multiple of $1/n$. Let M be the number of sequences $s_1, \dots, s_n$ of elements of S in which each element $s \in S$ occurs exactly $\omega(s)n$ times (this means, sampling a random element from the sequence $s_1, \dots, s_n$ recovers the probability distribution ω on S). Then, M satisfies*

$$\frac{e^{\mathrm{H}(\omega)n}}{e^{|S|}n^{|S|}} \leqslant M \leqslant e^{\mathrm{H}(\omega)n}.$$

Proof. We can assume that $\omega(s) > 0$ for every $s \in S$, because we can delete all elements $s \in S$ with $\omega(s) = 0$. First, note that M can be described as a multinomial coefficient:

$$M = \binom{n}{(\omega(s)n)_{s\in S}}.$$

For the lower bound, we use the simple Stirling approximation bounds, namely

$$\sqrt{2\pi\ell}\left(\frac{\ell}{e}\right)^{\ell} \leqslant \ell! \leqslant e\sqrt{\ell}\left(\frac{\ell}{e}\right)^{\ell}$$

for any positive integer ℓ. Now,

$$M = \binom{n}{(\omega(s)n)_{s\in S}} = \frac{n!}{\prod_{s\in S}(\omega(s)n)!} \geqslant \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\prod_{s\in S}\left(e\sqrt{\omega(s)n}\left(\frac{\omega(s)n}{e}\right)^{\omega(s)n}\right)}$$

$$= \frac{\sqrt{2\pi}}{e^{|S|}} \cdot \frac{\sqrt{n}}{\sqrt{\prod_{s\in S}\omega(s)n}} \cdot \frac{\left(\frac{n}{e}\right)^n}{\prod_{s\in S}\left(\frac{\omega(s)n}{e}\right)^{\omega(s)n}} \geqslant \frac{1}{e^{|S|}n^{|S|}} \cdot \prod_{s\in S}\omega(s)^{-\omega(s)n} = \frac{e^{\mathrm{H}(\omega)n}}{e^{|S|}n^{|S|}}.$$

For the upper bound, note that we have by the multinomial sum theorem

$$1 = \left(\sum_{s\in S}\omega(s)\right)^n \geqslant \binom{n}{(\omega(s)n)_{s\in S}} \prod_{s\in S}\omega(s)^{\omega(s)n}.$$

Here, we only considered those terms in the expansion of $(\sum_{s\in S}\omega(s))^n$ that contain each term $\omega(s)$ precisely $\omega(s)n$ times. Now, rearranging yields

$$\binom{n}{(\omega(s)n)_{s\in S}} \leqslant \prod_{s\in S}\omega(s)^{-\omega(s)n} = e^{\mathrm{H}(\omega)n},$$

as desired. □

The next lemma is also about counting sequences of elements of a set S, but under more restrictive conditions.

Lemma 2.2. *Let $f : S \to S'$ be any function between finite sets S and S'. Furthermore, let ω be a probability distribution on S and let z be a random variable on S with distribution ω. Let n be a positive integer and let us fix $s'_1, \dots, s'_n \in S'$. Now, let M be the number of sequences $s_1, \dots, s_n$ of elements of S in which each element $s \in S$ occurs exactly $\omega(s)n$ times and such that $f(s_j) = s'_j$ for $1 \leqslant j \leqslant n$. Then, M satisfies*

$$M \leqslant e^{\mathrm{H}(z|f(z))n} = e^{\mathrm{H}(z)n-\mathrm{H}(f(z))n}.$$

Proof. For every $s' \in S'$, set

$$J_{s'} = \{j \in \{1, \dots, n\} \mid s'_j = s'\}.$$

If $M = 0$, the statement is trivially true. Hence, we may assume that there exists at least one sequence $s_1, \dots, s_n \in S$ with the desired properties. Then, in particular, all the numbers $\omega(s)n$ for $s \in S$ are integers.

We claim that $|J_{s'}| = \sum_{s\in f^{-1}(s')}\omega(s)n$ for every $s' \in S'$. To see this, let us temporarily fix a sequence $s_1, \dots, s_n \in S$ satisfying all of the conditions in the lemma and let $s' \in S$. Since each $s \in f^{-1}(s')$ occurs exactly $\omega(s)n$ times in the sequence $s_1, \dots, s_n$, there are exactly $\sum_{s\in f^{-1}(s')}\omega(s)n$ choices for $j \in \{1, \dots, n\}$ such that $s'_j = f(s_j)$ equals s'. Thus, $|J_{s'}| = \sum_{s\in f^{-1}(s')}\omega(s)n$ as desired.

In order to form a sequence $s_1, \dots, s_n$ with the desired conditions, for each $s' \in S'$ we must distribute the elements of $f^{-1}(s')$ with the desired multiplicities among the index set $J_{s'}$. So, it is not hard to see that

$$M = \prod_{s'\in S'}\binom{|J_{s'}|}{(\omega(s)n)_{s\in f^{-1}(s')}}.$$

Note that for each $s \in S$, we have $\omega(s) = \mathbb{P}(z = s)$ and therefore

$$|J_{s'}| = \sum_{s \in f^{-1}(s')} \omega(s)n = \sum_{s \in f^{-1}(s')} \mathbb{P}(z = s) \cdot n = \mathbb{P}(z \in f^{-1}(s')) \cdot n = \mathbb{P}(f(z) = s') \cdot n$$

for every $s' \in S'$. Furthermore, if $s' \in S'$ and $s \in f^{-1}(s')$, then

$$\omega(s)n = \mathbb{P}(z = s) \cdot n = \frac{\mathbb{P}(z = s)}{\mathbb{P}(f(z) = s')} \cdot |J_{s'}| = \mathbb{P}(z = s \mid f(z) = s') \cdot |J_{s'}|.$$

In particular, $\mathbb{P}(z = s \mid f(z) = s')$ is an integer multiple of $1/|J_{s'}|$. Now, we obtain

$$M = \prod_{s' \in S'} \binom{|J_{s'}|}{(\mathbb{P}(z = s \mid f(z) = s') \cdot |J_{s'}|)_{s \in f^{-1}(s')}}.$$

For each $s' \in S'$, we can apply the upper bound in Lemma 2.1 to the distribution of z conditioned on $f(z) = s'$ and obtain

$$\binom{|J_{s'}|}{(\mathbb{P}(z = s \mid f(z) = s') \cdot |J_{s'}|)_{s \in f^{-1}(s')}} \leqslant \exp(\mathrm{H}(z \mid f(z) = s') \cdot |J_{s'}|).$$

Thus,

$$\begin{aligned} M \leqslant \exp\left(\sum_{s' \in S'} \mathrm{H}(z \mid f(z) = s') \cdot |J_{s'}|\right) &= \exp\left(\sum_{s' \in S'} \mathrm{H}(z \mid f(z) = s') \, \mathbb{P}(f(z) = s') \cdot n\right) \\ &= \exp(\mathrm{H}(z \mid f(z))n), \end{aligned}$$

as desired. Note that $\mathrm{H}(z \mid f(z)) = \mathrm{H}(z) - \mathrm{H}(f(z))$ by (2.3). □

Finally, we need one more lemma. Basically, this lemma states that perturbing a probability distribution slightly does not change the entropy very much.

Lemma 2.3. *Let ω_0 and ω_1 be two distributions on a finite set S, and suppose that $c > 0$ satisfies $\omega_0(s) \geqslant c$ and $\omega_1(s) \geqslant c$ for every $s \in S$. Then,*

$$|\,\mathrm{H}(\omega_1) - \mathrm{H}(\omega_0)| \leqslant \|\omega_1 - \omega_0\|_1 \log(1/c).$$

Proof. For any real number $0 \leqslant t \leqslant 1$, let $\omega_t = t\omega_1 + (1 - t)\omega_0$ (note that for $t = 0$ and $t = 1$, we indeed recover ω_0 and ω_1). Then, for each $0 \leqslant t \leqslant 1$ and each $s \in S$, we have

$$\omega_t(s) = t\omega_1(s) + (1 - t)\omega_0(s) = t(\omega_1(s) - \omega_0(s)) + \omega_0(s).$$

In particular, $\omega_t(s) = t\omega_1(s) + (1 - t)\omega_0(s) \geqslant c$. For each $0 \leqslant t \leqslant 1$, set $f(t) = \mathrm{H}(\omega_t)$, so

$$f(t) = -\sum_{s \in S} \omega_t(s) \log \omega_t(s).$$

It is easy to check that f is a continuous function on the interval [0,1] and it is differentiable on the open interval (0,1). For each $t \in (0, 1)$, we have (using that $\sum_s \omega_0(s) = \sum_s \omega_1(s) = 1$)

$$\begin{aligned} f'(t) &= -\sum_{s \in S} \left((\omega_1(s) - \omega_0(s)) \log \omega_t(s) + \omega_t(s) \frac{1}{\omega_t(s)} (\omega_1(s) - \omega_0(s))\right) \\ &= -\sum_{s \in S} (\omega_1(s) - \omega_0(s)) \log \omega_t(s). \end{aligned}$$

Hence,

$$|f'(t)| = \left|\sum_{s\in S}(\omega_1(s)-\omega_0(s))\log(1/\omega_t(s))\right| \leqslant \sum_{s\in S}|\omega_1(s)-\omega_0(s)|\log(1/c) = \|\omega_1-\omega_0\|_1\log(1/c).$$

By the mean value theorem, we now obtain

$$|\operatorname{H}(\omega_1)-\operatorname{H}(\omega_0)| = |f(1)-f(0)| \leqslant \sup_{t\in(0,1)}|f'(t)| \leqslant \|\omega_1-\omega_0\|_1\log(1/c),$$

as desired. □

3. *Proof of Theorem* 1.3

The goal of this section is to prove Theorem 1.3, which means to show that there is a sufficiently large k-colored sum-free set in $\mathbb{Z}_m^n$. We will often use the bound $|T_{m-1,k}| \leqslant m^k$, which follows from the fact that $T_{m-1,k} \subseteq \{0,\dots,m-1\}^k$. The following proposition states that there is a large collection of k-tuples in $\mathbb{Z}^n$ with certain conditions, the first of which is very similar to the condition for a k-colored sum-free set.

Proposition 3.1. *Let $m \geqslant 2$ and $k \geqslant 3$ be fixed, and let n be divisible by k and tending to infinity. Then, there exists a collection of k-tuples $(x_{1,j}, x_{2,j}, \dots, x_{k,j})_{j=1}^L$ of elements of $\mathbb{Z}^n$ with size $L \geqslant (\Gamma_{m,k})^{n-O(\sqrt{n})}$ such that*

- *for all $j_1,\dots,j_k \in \{1,\dots,L\}$*
$$x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k} = (m-1)\cdot\mathbb{1}^n \quad \text{if and only if} \quad j_1=j_2=\dots=j_k,$$
- *all coordinates of all $x_{i,j}$ are in the set $\{0,1,\dots,m-1\}$, and*
- *for each $x_{i,j}$ the sum of its n coordinates equals $(m-1)n/k$.*

Let us now prove that Proposition 3.1 implies Theorem 1.3. Afterward, we will prove Proposition 3.1.

Proof of Theorem 1.3 *assuming Proposition* 3.1. Let $m \geqslant 2$ and $k \geqslant 3$ be fixed. Assume for now that n is divisible by k. Let us consider a collection of k-tuples $(x_{1,j}, x_{2,j}, \dots, x_{k,j})_{j=1}^L$ of elements of $\mathbb{Z}^n$ as in Proposition 3.1 (in particular, $L \geqslant (\Gamma_{m,k})^{n-O(\sqrt{n})}$). Using this collection, we will construct a k-colored sum-free set $(y_{1,j}, y_{2,j}, \dots, y_{k,j})_{j=1}^L$ in $\mathbb{Z}_m^n$. For $1 \leqslant i \leqslant k-1$, and for any j, let $y_{i,j}$ be the projection of $x_{i,j} \in \mathbb{Z}^n$ to $\mathbb{Z}_m^n$. For $i=k$, let $y_{k,j}$ be the projection of $x_{k,j}-(m-1)\cdot\mathbb{1}^n$ to $\mathbb{Z}_m^n$.

Let us now check that $(y_{1,j}, y_{2,j}, \dots, y_{k,j})_{j=1}^L$ is indeed a k-colored sum-free set in $\mathbb{Z}_m^n$. Let $j_1,\dots,j_k \in \{1,\dots,L\}$ be such that $y_{1,j_1}+y_{2,j_2}+\dots+y_{k,j_k}=0$ in $\mathbb{Z}_m^n$. By the choice of the $y_{i,j}$, each of the n coordinates of the vector

$$x_{1,j_1}+x_{2,j_2}+\dots+x_{k-1,j_{k-1}}+x_{k,j_k}-(m-1)\cdot\mathbb{1}^n$$

is divisible by m. Thus, every coordinate of $x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k} \in \mathbb{Z}^n$ has remainder $m-1$ upon division by m. Since each coordinate is nonnegative, this implies, in particular, that each coordinate of $x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k}$ is at least $m-1$. On the other hand, the sum of the coordinates of each x_{i,j_i} equals $(m-1)n/k$, so the sum of the coordinates of $x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k}$ equals $(m-1)n$. Since each is at least $m-1$, this implies that all the coordinates of $x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k}$ are equal to $m-1$. Hence,

$$x_{1,j_1}+x_{2,j_2}+\dots+x_{k,j_k} = (m-1)\cdot\mathbb{1}^n.$$

So by the first property listed in Proposition 3.1, we must have $j_1=j_2=\dots=j_k$.

For the converse, assume $j_1 = j_2 = \cdots = j_k$. Then,

$$x_{1,j_1} + x_{2,j_2} + \cdots + x_{k-1,j_{k-1}} + x_{k,j_k} - (m-1)\cdot \mathbb{1}^n = 0$$

in $\mathbb{Z}^n$, hence $y_{1,j_1} + y_{2,j_2} + \cdots + y_{k,j_k} = 0$ in $\mathbb{Z}_m^n$ as well.

Thus, we have constructed a k-colored sum-free set in $\mathbb{Z}_m^n$ of size

$$L \geqslant (\Gamma_{m,k})^{n-O(\sqrt{n})},$$

if n is divisible by k. Note that by embedding $\mathbb{Z}_m^n$ into $\mathbb{Z}_m^{n'}$ for $n' > n$, any k-colored sum-free set in $\mathbb{Z}_m^n$ gives rise to a k-colored sum-free set in $\mathbb{Z}_m^{n'}$ of the same size. Thus, for all n we obtain a k-colored sum-free set in $\mathbb{Z}_m^n$ of size at least $(\Gamma_{m,k})^{n-O(\sqrt{n})}$. □

The rest of this section will be devoted to proving Proposition 3.1. However, in some sense this section covers only the general steps for the proof, the major difficulty lies in the proofs of several lemmas and propositions that are postponed to the next two sections.

Proof of Proposition 3.1. Let $m \geqslant 2$ and $k \geqslant 3$ be fixed and let n be divisible by k and sufficiently large (in terms of m and k). Our goal is to find a sufficiently large collection of k-tuples with the properties listed in Proposition 3.1. Recall that in Section 1, we defined a specific probability distribution $\nu_{m,k}$ on $\{0, \dots, m-1\}$.

LEMMA 3.2. *The probability distribution $\nu_{m,k}$ has expectation $\mathbb{E}(\nu_{m,k}) = (m-1)/k$, and its entropy is $\mathrm{H}(\nu_{m,k}) = \log \Gamma_{m,k}$.*

We will prove Lemma 3.2 in Section 5. In fact, one can also prove that $\nu_{m,k}$ has the highest entropy among all probability distributions on $\{0, \dots, m-1\}$ with expectation $(m-1)/k$. Although we will not need this fact for our argument, it still provides motivation as to why the distribution $\nu_{m,k}$ is relevant.

In order to use $\nu_{m,k}$ for constructing the desired collection of k-tuples in $\mathbb{Z}^n$, we first need to round the probabilities in $\nu_{m,k}$ to rational numbers with denominator n. The following proposition basically states that a suitable rounding of $\nu_{m,k}$ exists.

PROPOSITION 3.3. *Let $m \geqslant 2$, $k \geqslant 3$ and let n be divisible by k and sufficiently large (in terms of m and k). Then, there exist probability distributions ν on $\{0, \dots, m-1\}$ and τ on $T_{m-1,k}$ with the following conditions.*

- *Each probability $\nu(i)$ for $i \in \{0, \dots, m-1\}$ is an integer multiple of $1/n$.*
- *ν has expectation $\mathbb{E}(\nu) = (m-1)/k$.*
- $\mathrm{H}(\nu) \geqslant \log \Gamma_{m,k} - C_{m,k}/n$.
- *Each probability $\tau(t)$ for $t \in T_{m-1,k}$ is an integer multiple of $1/n$.*
- *τ is S_k-symmetric and has marginal ν.*
- *We have $\mathrm{H}(\tau') \leqslant \mathrm{H}(\tau) + (D_{m,k} \log n)/n$ for every probability distribution τ' on $T_{m-1,k}$ with the property that all of the k coordinate projections of τ' are equal to ν.*

Here, $C_{m,k} > 0$ and $D_{m,k} > 0$ are constants only depending on m and k.

The key ingredients for the proof of Proposition 3.3 are Theorem 1.4 and Lemma 3.2. Starting from there, one needs to perform several rounding steps in order to obtain Proposition 3.3. We will postpone the technical details of these rounding steps to Section 5. The proof of Theorem 1.4 will be given in the second part of this paper, starting from Section 6.

Let us fix ν and τ as in Proposition 3.3. Now, let $X_0 \subseteq \{0, \dots, m-1\}^n$ consist of those vectors $x \in \{0, \dots, m-1\}^n$ that have exactly $\nu(i)n$ coordinates equal to i for every $i = 0, \dots, m-1$ (that is, that contain exactly $\nu(0)n$ zeros, exactly $\nu(1)n$ ones, and so on).

Then for any $x \in X_0$, choosing one of the coordinates of x uniformly at random recovers the probability distribution ν on $\{0, \dots, m-1\}$.

Note that for each $x \in X_0$, all of its n coordinates are in the set $\{0, \dots, m-1\}$ and their sum is equal to $0 \cdot \nu(0)n + \dots + (m-1) \cdot \nu(m-1)n = \mathbb{E}(\nu)n = n(m-1)/k$. So, if we choose our desired collection of k-tuples in such a way that $x_{i,j} \in X_0$ for all i and j, then the second and third condition in Proposition 3.1 will automatically be satisfied.

In order to obtain the desired collection of k-tuples, we will use a probabilistic sampling argument. To describe this sampling, let us first (using Bertrand's Postulate) choose a prime number P with

$$4n^{D_{m,k}+2m^k} \exp\left(\frac{\mathrm{H}(\tau) - \mathrm{H}(\nu)}{k-2} \cdot n\right) \leqslant P \leqslant n^{D_{m,k}+3m^k} \exp\left(\frac{\mathrm{H}(\tau) - \mathrm{H}(\nu)}{k-2} \cdot n\right). \tag{3.1}$$

By (2.4), we have $\mathrm{H}(\tau) \geqslant \mathrm{H}(\mu(\tau)) = \mathrm{H}(\nu)$. Hence, as long as n is large enough, we have

$$P \geqslant n^{m^k} \geqslant (2k)!m^{2k}.$$

On the other hand, $\mathrm{H}(\tau) \leqslant \log(|T_{m-1,k}|) \leqslant \log(m^k) = k \log m$. Therefore, as long as n is sufficiently large, we have $P \leqslant \exp(2k(\log m)n)$, so

$$\log P \leqslant 2k(\log m)n. \tag{3.2}$$

We now define functions $\tilde{g}_1, \dots, \tilde{g}_k : X_0 \to \mathbb{Z}^{n+k-1}$. For any $x \in X_0$, consider its coordinates $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$. For $1 \leqslant i \leqslant k-1$, we define $\tilde{g}_i(x) = (\tilde{g}_i(x)^{(1)}, \tilde{g}_i(x)^{(2)}, \dots, \tilde{g}_i(x)^{(n+k-1)}) \in \mathbb{Z}^{n+k-1}$ by

$$\tilde{g}_i(x)^{(j)} = \begin{cases} x^{(j)} & \text{if } j \leqslant n \\ 1 & \text{if } j = n+i \\ 0 & \text{if } j > n \text{ and } j \neq n+i. \end{cases}$$

In other words, for $1 \leqslant i \leqslant k-1$, the vector $\tilde{g}_i(x)$ can be obtained from x by attaching the ith standard basis vector in $\mathbb{Z}^{k-1}$ at the end.

For $i = k$, we define $\tilde{g}_k(x) = (\tilde{g}_k(x)^{(1)}, \tilde{g}_k(x)^{(2)}, \dots, \tilde{g}_k(x)^{(n+k-1)}) \in \mathbb{Z}^{n+k-1}$ by

$$\tilde{g}_k(x)^{(j)} = \begin{cases} x^{(j)} - (m-1) & \text{if } j \leqslant n \\ -1 & \text{if } j > n. \end{cases}$$

Note that for each $1 \leqslant i \leqslant k$ and each $x \in X_0 \subseteq \{0, \dots, m-1\}^n$, all the coordinates of $\tilde{g}_i(x)$ have absolute value at most $m-1$. Let us also define functions $g_1, \dots, g_k : X_0 \to \mathbb{F}_P^{n+k-1}$ by taking $g_i(x)$ to be the projection of $\tilde{g}_i(x)$ to $\mathbb{F}_P^{n+k-1}$ for every $1 \leqslant i \leqslant k$ and every $x \in X_0$. Given $x_1, \dots, x_k \in X_0$ with $x_1 + \dots + x_k = (m-1) \cdot \mathbb{1}^n$ in $\mathbb{Z}^n$, we have $\tilde{g}_1(x_1) + \dots + \tilde{g}_k(x_k) = 0$ in $\mathbb{Z}^{n+k-1}$ and therefore $g_1(x_1) + \dots + g_k(x_k) = 0$ in $\mathbb{F}_P^{n+k-1}$. However, $g_1(x_1), \dots, g_{k-1}(x_{k-1})$ are linearly independent over $\mathbb{F}_P$ (because their last $k-1$ coordinates form the standard basis vectors in $\mathbb{F}_P^{k-1}$).

For our probabilistic sampling argument, we will use a large k-colored sum-free set in $\mathbb{F}_P$. The following lemma ensures the existence of such a large k-colored sum-free set. The proof relies on a lemma by Alon [**1**, Lemma 3.1], which he proved using a modification of Behrend's construction [**4**].

Lemma 3.4. *There exists a k-colored sum-free set $(y_{1,j}, y_{2,j}, \dots, y_{k,j})_{j=1}^R$ in $\mathbb{F}_P$ of size*

$$R \geqslant P \cdot \exp(-12\sqrt{\log P \log k}). \tag{3.3}$$

Proof. Recall that $P \geqslant k \geqslant 3$. By [1, Lemma 3.1], there exists a subset $Y \subseteq \{1, \dots, \lfloor P/k \rfloor\}$ of size at least

$$|Y| \geqslant \frac{\lfloor P/k \rfloor}{e^{10\sqrt{\log(\lfloor P/k \rfloor)\log(k-1)}}} \geqslant \frac{P}{2k} \cdot \exp(-10\sqrt{\log P \log k}) \geqslant P \cdot \exp(-12\sqrt{\log P \log k})$$

such that the only solutions in Y to the integer equation $y_1 + \dots + y_{k-1} = (k-1)y_k$ satisfy $y_1 = \dots = y_{k-1} = y_k$. Note that for $y_1, \dots, y_k \in Y \subseteq \{1, \dots, \lfloor P/k \rfloor\}$, both sides of the equation $y_1 + \dots + y_{k-1} = (k-1)y_k$ are integers between 1 and $P-1$. Hence, the equation $y_1 + \dots + y_{k-1} = (k-1)y_k$ holds over $\mathbb{F}_P$ if and only if it holds in the integers. So, let us interpret Y as a subset of $\mathbb{F}_P$. Then, still the only solutions in Y to the equation $y_1 + \dots + y_{k-1} = (k-1)y_k$ over $\mathbb{F}_P$ are $y_1 = \dots = y_{k-1} = y_k$.

Now, for each $y \in Y \subseteq \mathbb{F}_P$ consider the k-tuple $(y, \dots, y, -(k-1)y) \in \mathbb{F}_P^k$. Altogether these k-tuples form a k-colored sum-free set in $\mathbb{F}_P$ of size $|Y| \geqslant P \cdot \exp(-12\sqrt{\log P \log k})$.

Let $(y_{1,j}, y_{2,j}, \dots, y_{k,j})_{j=1}^R$ be a k-colored sum-free set in $\mathbb{F}_P$ as in the lemma. For $i = 1, \dots, k$ set

$$Y_i = \{y_{i,j} \mid 1 \leqslant j \leqslant R\}.$$

Since for a fixed i, the vectors $y_{i,j}$ for $1 \leqslant j \leqslant R$ are all distinct, we have $|Y_i| = R$. Furthermore, by the definition of k-colored sum-free set, any k-tuple in $Y_1 \times \dots \times Y_k$ summing to zero needs to be of the form $(y_{1,j}, y_{2,j}, \dots, y_{k,j})$ for some j.

In order to perform the desired sampling, let us now choose a random linear map $f : \mathbb{F}_P^{n+k-1} \to \mathbb{F}_P$. More precisely, we choose f uniformly at random among all linear maps $\mathbb{F}_P^{n+k-1} \to \mathbb{F}_P$.

Claim 3.5. *Suppose that $v_1, v_2, \dots, v_\ell \in \mathbb{F}_P^{n+k-1}$ are linearly independent over $\mathbb{F}_P$. Then, the images $f(v_1), \dots, f(v_\ell)$ are probabilistically independent and uniformly distributed over $\mathbb{F}_P$.*

Proof. We can extend $v_1, v_2, \dots, v_\ell$ to a basis $v_1, v_2, \dots, v_{n+k-1}$ of $\mathbb{F}_P^{n+k-1}$. Note that we can model the random choice of f by mapping each v_i to a random element of $\mathbb{F}_P$ (uniformly, and independently for all $1 \leqslant i \leqslant n+k-1$). By this description of the random experiment, the claim is clearly true. □

We now define, for $1 \leqslant i \leqslant k$,

$$X_i = \{x \in X_0 \mid f(g_i(x)) \in Y_i\}.$$

Definition 3.6. A *candidate k-tuple* is a k-tuple $(x_1, x_2, \dots, x_k) \in X_1 \times X_2 \times \dots \times X_k$ with $x_1 + x_2 + \dots + x_k = (m-1) \cdot \mathbb{1}^n$. A candidate k-tuple is *isolated*, if there is no other candidate k-tuple $(x_1', x_2', \dots, x_k')$ with $x_i' = x_i$ for some $1 \leqslant i \leqslant k$.

Let $(x_{1,j}, x_{2,j}, \dots, x_{k,j})_{j=1}^L$ be the collection of all isolated candidate k-tuples, where L is the total number of isolated candidate k-tuples. Note that since the sets X_i depend on the choice of the random map f, the notions in Definition 3.6 also depend on this choice. In particular, L is a random variable that depends on the random map f. We claim that the collection $(x_{1,j}, x_{2,j}, \dots, x_{k,j})_{j=1}^L$ will always satisfy the first condition in Proposition 3.1. For each $j \in \{1, \dots, L\}$, we have $x_{1,j} + \dots + x_{k,j} = (m-1) \cdot \mathbb{1}^n$ by the first part of Definition 3.6. Also note that $x_{i,j} \neq x_{i,j'}$ whenever $j \neq j'$, since all the $(x_{1,j}, x_{2,j}, \dots, x_{k,j})$ are isolated. Now, suppose that we have $j_1, \dots, j_k \in \{1, \dots, L\}$ with $x_{1,j_1} + x_{2,j_2} + \dots + x_{k,j_k} = (m-1) \cdot \mathbb{1}^n$. As $x_{i,j_i} \in X_i$ for each $1 \leqslant i \leqslant k$, we obtain that $(x_{1,j_1}, x_{2,j_2}, \dots, x_{k,j_k})$ is a candidate k-tuple.

But since all the candidate k-tuples $(x_{1,j},x_{2,j},\dots,x_{k,j})_{j=1}^{L}$ in the collection are isolated, $(x_{1,j_1},x_{2,j_2},\dots,x_{k,j_k})$ must be one of the k-tuples in the collection. Hence, $j_1=\dots=j_k$. So, the isolated candidate k-tuples indeed form a collection satisfying the first condition in Proposition 3.1.

Each of the vectors $x_{i,j}$ in the k-tuples in the collection satisfies $x_{i,j}\in X_i\subseteq X_0$. As we have already seen above, this implies that the n coordinates of $x_{i,j}$ are all in the set $\{0,\dots,m-1\}$ and their sum is equal to $n(m-1)/k$. Thus, the second and third condition in Proposition 3.1 are satisfied for the collection of isolated candidate k-tuples. So, it only remains to prove that for at least one choice of the random map f, the number L of isolated candidate k-tuples is sufficiently large. In particular, it suffices to show that the expected value of L is sufficiently large.

The following proposition states that certain k-tuples have a good probability of being isolated candidate k-tuples. Note that if $x_1,\dots,x_k\in\{0,\dots,m-1\}^n$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$, then for every $1\leqslant j\leqslant n$, the jth coordinates of these vectors satisfy $x_1^{(j)}+\dots+x_k^{(j)}=m-1$. Hence, $(x_1^{(j)},\dots,x_k^{(j)})\in T_{m-1,k}$ for every $1\leqslant j\leqslant n$.

Proposition 3.7. *Let $x_1,\dots,x_k\in\{0,\dots,m-1\}^n$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$. Assume that for every $t\in T_{m-1,k}$ the number of $j\in\{1,\dots,n\}$ with $(x_1^{(j)},\dots,x_k^{(j)})=t$ is precisely $\tau(t)n$. Then $x_1,\dots,x_k\in X_0$ and furthermore the probability that $(x_1,\dots,x_k)$ is an isolated candidate k-tuple is at least*

$$\exp\big(\mathrm{H}(\nu)n-\mathrm{H}(\tau)n-25km\sqrt{n}\big).$$

We postpone the proof of this proposition for a little while, in order to first finish the proof of Proposition 3.1.

Claim 3.8. *The number of k-tuples $(x_1,\dots,x_k)$ satisfying the assumptions of Proposition 3.7 is at least $e^{\mathrm{H}(\tau)n}n^{-2m^k}$.*

Proof. We can form a k-tuple $x_1,\dots,x_k\in\{0,\dots,m-1\}^n$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$ coordinate by coordinate by choosing some $t_j\in T_{m-1,k}$ for each $1\leqslant j\leqslant n$ and then setting $(x_1^{(j)},\dots,x_k^{(j)})=t_j$. So in order for $(x_1,\dots,x_k)$ to satisfy the assumptions of Proposition 3.7, we just need to make sure that for each $t\in T_{m-1,k}$ the number of $j\in\{1,\dots,n\}$ with $t_j=t$ is exactly $\tau(t)n$. So, it suffices to prove that there are at least $e^{\mathrm{H}(\tau)n-\sqrt{n}}$ sequences $t_1,\dots,t_n$ of elements of $T_{m-1,k}$ in which each element $t\in T_{m-1,k}$ occurs exactly $\tau(t)n$ times. By Lemma 2.1 (recall from Proposition 3.3 that each probability in τ is an integer multiple of $1/n$), the number of such sequences is indeed at least $e^{\mathrm{H}(\tau)n}e^{-|T_{m-1,k}|}n^{-|T_{m-1,k}|}\geqslant e^{\mathrm{H}(\tau)n}n^{-2m^k}$. □

Combining Claim 3.8 and Proposition 3.7, we obtain that the expected value of the number L of isolated candidate k-tuples is at least

$$\begin{aligned}e^{\mathrm{H}(\tau)n}n^{-2m^k}\exp\big(\mathrm{H}(\nu)n-\mathrm{H}(\tau)n-25km\sqrt{n}\big)=e^{\mathrm{H}(\nu)n-O(\sqrt{n})}&\geqslant(\Gamma_{m,k})^n e^{-C_{m,k}-O(\sqrt{n})}\\&=(\Gamma_{m,k})^{n-O(\sqrt{n})}.\end{aligned}$$

Here, we used $\mathrm{H}(\nu)\geqslant\log\Gamma_{m,k}-C_{m,k}/n$. This finishes the proof of Proposition 3.1. □

We will now prove Proposition 3.7, apart from postponing the proofs of Lemma 3.9 and Proposition 3.10 to the next two sections. While Lemma 3.9 is a relatively easy linear algebra statement, proving Proposition 3.10 is the main difficulty in the first part of this paper. The proof will take up all of Section 4.

Proof of Proposition 3.7. Let us fix $x_1, \dots, x_k \in \{0, \dots, m-1\}^n$ with $x_1 + \dots + x_k = (m-1) \cdot \mathbb{1}^n$ and such that for every $t \in T_{m-1,k}$ the number of $j \in \{1, \dots, n\}$ with $(x_1^{(j)}, \dots, x_k^{(j)}) = t$ is precisely $\tau(t)n$.

First, we need to prove that $x_i \in X_0$ for $1 \leqslant i \leqslant k$. Let us assume that $i = 1$, the other cases are analogous. In order to show $x_1 \in X_0$, we need to check that for every $a = 0, \dots, m-1$ the vector x_1 has exactly $\nu(a)n$ coordinates equal to a. Recall that for each $(a_1, \dots, a_k) \in T_{m-1,k}$ the number of $j \in \{1, \dots, n\}$ with $(x_1^{(j)}, \dots, x_k^{(j)}) = (a_1, \dots, a_k)$ is precisely $\tau(a_1, \dots, a_k)n$. Hence, the number of $j \in \{1, \dots, n\}$ with $x_1^{(j)} = a$ is precisely

$$\sum_{\substack{a_2, \dots, a_k \in \{0, \dots, m-1\} \\ a + a_2 + \dots + a_k = m-1}} \tau(a, a_2, \dots, a_k)n = \mu(\tau)(a)n = \nu(a)n.$$

Here, we used that ν is the marginal of τ (see Proposition 3.3). So, the vector x_1 has indeed exactly $\nu(a)n$ coordinates equal to a. Thus, $x_1 \in X_0$, and analogously $x_2, \dots, x_k \in X_0$.

Now, we need to prove the desired lower bound for the probability that $(x_1, \dots, x_k)$ is an isolated candidate k-tuple. Since we already assumed $x_1 + \dots + x_k = (m-1) \cdot \mathbb{1}^n$, the k-tuple $(x_1, \dots, x_k)$ will be a candidate k-tuple if and only if $x_i \in X_i$ for $i = 1, \dots, k$. If $(x_1, \dots, x_k)$ is a candidate k-tuple, it is isolated if there does not exist another candidate k-tuple $(x_1', x_2', \dots, x_k')$ with $x_i' = x_i$ for some $1 \leqslant i \leqslant k$. Note that for any such $(x_1', x_2', \dots, x_k')$, we would need to have $(x_1', x_2', \dots, x_k') \in X_1 \times \dots \times X_k \subseteq X_0^k$. So, the probability that $(x_1, \dots, x_k)$ is an isolated candidate k-tuple is at least

$$\mathbb{P}[(x_1, \dots, x_k) \in X_1 \times \dots \times X_k] - \sum_{(x_1', \dots, x_k')} \mathbb{P}[(x_1, \dots, x_k), (x_1', \dots, x_k') \in X_1 \times \dots \times X_k], \tag{3.4}$$

where the sum is over all $(x_1', x_2', \dots, x_k') \in X_0^k$ with $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$ and $x_i = x_i'$ for some i but with $(x_1', \dots, x_k') \neq (x_1, \dots, x_k)$.

Let us first determine the first term, namely the probability that $x_i \in X_i$ for $i = 1, \dots, k$. Recall that by the definition of X_i, we have $x_i \in X_i$ if and only if $f(g_i(x_i)) \in Y_i$. Also recall that $g_1(x_1), \dots, g_{k-1}(x_{k-1}) \in \mathbb{F}_P^{n+k-1}$ are linearly independent over $\mathbb{F}_P$. Hence, by Claim 3.5 the images $f(g_1(x_1)), \dots, f(g_{k-1}(x_{k-1}))$ are probabilistically independent and uniformly distributed over $\mathbb{F}_P$. Recall that $x_1 + \dots + x_k = (m-1) \cdot \mathbb{1}^n$ in $\mathbb{Z}^n$ implies $g_1(x_1) + \dots + g_k(x_k) = 0$ in $\mathbb{F}_P^{n+k-1}$ and therefore $f(g_1(x_1)) + \dots + f(g_k(x_k)) = 0$ in $\mathbb{F}_P$. If $(x_1, \dots, x_k) \in X_1 \times \dots \times X_k$, then $(f(g_1(x_1)), \dots, f(g_k(x_k))) \in Y_1 \times \dots \times Y_k$, so $(f(g_1(x_1)), \dots, f(g_k(x_k)))$ needs to be one of the k-tuples in the multicolored sum-free set $(y_{1,j}, y_{2,j}, \dots, y_{k,j})_{j=1}^R$. There are R choices for $j \in \{1, \dots, R\}$, and for each of these choices the probability of $(f(g_1(x_1)), \dots, f(g_k(x_k))) = (y_{1,j}, \dots, y_{k,j})$ equals $(1/P)^{k-1}$ (for each $i = 1, \dots, k-1$ the probability of $f(g_i(x_i)) = y_{i,j}$ is $1/P$ and these events are all independent. If they all happen, then due to $f(g_1(x_1)) + \dots + f(g_k(x_k)) = 0$ we automatically have $f(g_k(x_k)) = y_{k,j}$). Hence,

$$\mathbb{P}[(x_1, \dots, x_k) \in X_1 \times \dots \times X_k] = \frac{R}{P^{k-1}}. \tag{3.5}$$

Now, for the second term, let $(x_1', x_2', \dots, x_k') \in X_0^k$ with $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$ and $x_i = x_i'$ for some i but with $(x_1', \dots, x_k') \neq (x_1, \dots, x_k)$. Let $d = \dim \operatorname{span}_{\mathbb{Q}}(x_1 - x_1', \dots, x_k - x_k')$. Note that $1 \leqslant d \leqslant k-2$, since $x_i - x_i' = 0$ for some i (but not for all i), and $\sum_i (x_i - x_i') = 0$. By the following lemma, the dimension of the span of $g_1(x_1), \dots, g_k(x_k), g_1(x_1'), \dots, g_k(x_k') \in \mathbb{F}_P^{n+k-1}$ equals $k-1+d$. We remark that this lemma uses the fact that P is large (in terms of m and k).

Lemma 3.9. *Suppose that* $(x_1,\dots,x_k),(x'_1,\dots,x'_k)\in X_0^k$ *satisfy* $x_1+\dots+x_k=x'_1+\dots+x'_k=(m-1)\cdot\mathbb{1}^n$. *Then,*

$$\dim\operatorname{span}_{\mathbb{F}_P}\left(g_1(x_1),\dots,g_k(x_k),g_1(x'_1),\dots,g_k(x'_k)\right)=k-1+\dim\operatorname{span}_{\mathbb{Q}}\left(x_1-x'_1,\dots,x_k-x'_k\right).$$

We postpone the proof of this lemma to Section 5. By the lemma, we can choose $k-1+d$ linearly independent vectors among $g_1(x_1),\dots,g_k(x_k),g_1(x'_1),\dots,g_k(x'_k)$. By Claim 3.5, the images under f of these $k-1+d$ linearly independent vectors will be independently uniformly distributed in $\mathbb{F}_P$.

We need to find an upper bound for the probability $\mathbb{P}[(x_1,\dots,x_k),(x'_1,\dots,x'_k)\in X_1\times\dots\times X_k]$. We have $(x_1,\dots,x_k),(x'_1,\dots,x'_k)\in X_1\times\dots\times X_k$ if and only if $(f(g_1(x_1)),\dots,f(g_k(x_k))),(f(g_1(x'_1)),\dots,f(g_k(x'_k)))\in Y_1\times\dots\times Y_k$. Since $f(g_1(x_1))+\dots+f(g_k(x_k))=0$ and $f(g_1(x'_1))+\dots+f(g_k(x'_k))=0$ in $\mathbb{F}_P$, this can only happen if both of $(f(g_1(x_1)),\dots,f(g_k(x_k)))$ and $(f(g_1(x'_1)),\dots,f(g_k(x'_k)))$ are k-tuples from the collection $(y_{1,j},y_{2,j},\dots,y_{k,j})_{j=1}^R$. Since $x_i=x'_i$ for some i, we also have $f(g_i(x_i))=f(g_i(x'_i))$. Therefore, as no two k-tuples in the collection $(y_{1,j},y_{2,j},\dots,y_{k,j})_{j=1}^R$ share the same ith vector, we must have

$$(f(g_1(x_1)),\dots,f(g_k(x_k)))=(f(g_1(x'_1)),\dots,f(g_k(x'_k)))=(y_{1,j},y_{2,j},\dots,y_{k,j})$$

for some $j\in\{1,\dots,R\}$. For each of the R choices of j, the probability of satisfying the equation above is at most $(1/P)^{k-1+d}$ (since among $g_1(x_1),\dots,g_k(x_k),g_1(x'_1),\dots,g_k(x'_k)$ there are $k-1+d$ linearly independent vectors, and each of them has probability $1/P$ to have the desired image under the map f)[†]. Hence,

$$\mathbb{P}[(x_1,\dots,x_k),(x'_1,\dots,x'_k)\in X_1\times\dots\times X_k]\leqslant\frac{R}{P^{k-1+d}}.\tag{3.6}$$

The following proposition gives an upper bound for the number of different choices of $(x'_1,x'_2,\dots,x'_k)\in X_0^k$ that we need to consider for each value of $1\leqslant d\leqslant k-2$. We postpone the proof of Proposition 3.10 to Section 4. This proof is the major part of the work of deriving Theorem 1.3 from Theorem 1.4.

Proposition 3.10. *Let* $(x_1,\dots,x_k)\in X_0^k$ *with* $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$ *be fixed such that for every* $t\in T_{m-1,k}$ *the number of* $j\in\{1,\dots,n\}$ *with* $(x_1^{(j)},\dots,x_k^{(j)})=t$ *is exactly* $\tau(t)n$. *Then for each* $1\leqslant d\leqslant k-2$, *there are at most*

$$n^{D_{m,k}+2m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)$$

different k*-tuples* $(x'_1,x'_2,\dots,x'_k)\in X_0^k$ *satisfying* $x'_1+\dots+x'_k=(m-1)\cdot\mathbb{1}^n$, $\dim\operatorname{span}_{\mathbb{Q}}(x_1-x'_1,\dots,x_k-x'_k)=d$, *and* $x_i=x'_i$ *for some* i.

For every $1\leqslant d\leqslant k-2$, we have by (3.1)

$$\frac{n^{D_{m,k}+2m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)}{P^d}\leqslant\frac{n^{D_{m,k}+2m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)}{4^dn^{dD_{m,k}+2dm^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)}\leqslant\frac{1}{4^d}.$$

Hence, by Proposition 3.10 and (3.6), the big sum in (3.4) is at most

$$\sum_{d=1}^{k-2}n^{D_{m,k}+2m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)\cdot\frac{R}{P^{k-1+d}}\leqslant\frac{R}{P^{k-1}}\sum_{d=1}^{k-2}\frac{1}{4^d}\leqslant\frac{R}{2P^{k-1}}.$$

[†] One can actually show that this is always true with equality, but we will not need this.

Recalling (3.5), (3.3), (3.1), and (3.2), we obtain that the probability that $(x_1, \dots, x_k)$ is an isolated candidate k-tuple is at least

$$\begin{aligned}\frac{R}{P^{k-1}} - \frac{R}{2P^{k-1}} &= \frac{R}{2P^{k-1}} = \frac{R}{P} \cdot \frac{1}{2P^{k-2}} \\ &\geqslant e^{-12\sqrt{\log P \log k}} \cdot \frac{1}{2} \cdot n^{-(D_{m,k}+3m^k)(k-2)} \exp(-(\mathrm{H}(\tau) - \mathrm{H}(\nu))n) \\ &\geqslant n^{-k(D_{m,k}+3m^k)} e^{-12\sqrt{2k(\log m)n \log k}} \exp(\mathrm{H}(\nu)n - \mathrm{H}(\tau)n) \\ &\geqslant \exp\left(\mathrm{H}(\nu)n - \mathrm{H}(\tau)n - 25km\sqrt{n}\right),\end{aligned}$$

where in the last step we used that n is sufficiently large in terms of m and k. This finishes the proof of Proposition 3.7.

4. *Proof of Proposition* 3.10

In this section, we will prove Proposition 3.10. The first two subsections contain preparations for the proof, and in the third subsection, we will actually prove Proposition 3.10.

4.1. *An entropy inequality*

In this subsection, we will establish an inequality between entropies which will be relevant in the process of proving Proposition 3.10.

For any $\ell \geqslant 2$, let $V_\ell \subseteq \mathbb{Q}^\ell$ denote the subspace consisting of those vectors $v = (v_1, \dots, v_\ell)$ with $v_1 + \dots + v_\ell = 0$. Note that V_ℓ is a hyperplane in $\mathbb{Q}^\ell$, so it has dimension $\ell - 1$. To any $v = (v_1, \dots, v_\ell) \in V_\ell$, we can associate a map $\tilde{v} : T_{r,\ell} \to \mathbb{Q}$ by setting $\tilde{v}(a_1, \dots, a_\ell) = v_1 a_1 + \dots + v_\ell a_\ell$ for every $(a_1, \dots, a_\ell) \in T_{r,\ell}$ (here, $r \geqslant 0$ is a nonnegative integer). Note that for fixed $r \geqslant 0$ and $\ell \geqslant 2$, any $(a_1, \dots, a_\ell) \in T_{r,\ell}$ is uniquely determined by its values $\tilde{v}(a_1, \dots, a_\ell)$ for all $v \in V_\ell$. Indeed, if we are given $\tilde{v}(a_1, \dots, a_\ell)$ for all $v \in V_\ell$, we in particular know the values $a_j - a_1$ for $j = 2, \dots, \ell$. Using $a_1 + \dots + a_\ell = r$, this determines $(a_1, \dots, a_\ell)$.

Lemma 4.1. *Given any nonzero vector $v \in V_\ell$, the vectors v^σ for $\sigma \in S_\ell$ span the entire space V_ℓ.*

Proof. Let $v = (v_1, \dots, v_\ell)$ and note that $v \neq 0$ and $v_1 + \dots + v_\ell = 0$ implies that the coordinates $v_1, \dots, v_\ell$ are not all equal. So, we may assume without loss of generality that $v_1 \neq v_2$. Let V_ℓ' be the subspace of V_ℓ spanned by the vectors v^σ for $\sigma \in S_\ell$. By definition, the subspace V_ℓ' is stable under permutations of the coordinates. We need to show that $V_\ell' = V_\ell$. Considering the transposition $\sigma = (1,2) \in S_\ell$, we have

$$\begin{aligned} v^\sigma - v &= (v_2, v_1, v_3, \dots, v_\ell) - (v_1, v_2, v_3, \dots, v_\ell) = (v_2 - v_1, v_1 - v_2, 0, \dots, 0) \\ &= (v_2 - v_1) \cdot (1, -1, 0, \dots, 0). \end{aligned}$$

Therefore, $(v_2 - v_1) \cdot (1, -1, 0, \dots, 0) \in V_\ell'$. Hence, using $v_1 \neq v_2$, we obtain $(1, -1, 0, \dots, 0) \in V_\ell'$. So, all permutations of the vector $(1, -1, 0, \dots, 0)$ are contained in V_ℓ' as well. It is easy to see that all the permutations of $(1, -1, 0, \dots, 0)$ generate V_ℓ, hence $V_\ell' = V_\ell$ as desired. □

Lemma 4.2. *Let $r \geqslant 0$ and $\ell \geqslant 2$. Suppose that π is an S_ℓ-symmetric distribution on $T_{r,\ell}$ and that z is a random variable on $T_{r,\ell}$ with distribution π. Then, for any subspace $W \subseteq V_\ell$, we have*

$$\mathrm{H}((\tilde{w}(z))_{w \in W}) \geqslant \frac{\dim W}{\ell - 1} \mathrm{H}(\pi).$$

Proof. We prove the lemma by strong induction on $\dim W$. The case $\dim W = 0$ is trivial. Therefore, suppose that $\dim W > 0$ and that we have already proved the lemma for every subspace of V_ℓ with smaller dimension.

For every $\sigma \in S_\ell$, we obtain a subspace $W^\sigma = \{w^\sigma \mid w \in W\} \subseteq V_\ell$ from W by permuting all vectors $w \in W$ according to σ. Clearly, $\dim W^\sigma = \dim W$. Given $w \in W$ and $\sigma \in S_\ell$, we have $\tilde{w}^\sigma(t) = \tilde{w}(t^{\sigma^{-1}})$ for every $t \in T_{r,\ell}$. Therefore, as π is S_ℓ-symmetric, we can conclude that $(\tilde{w}^\sigma(z))_{w\in W}$ and $(\tilde{w}(z))_{w\in W}$ have the same distribution. Hence,

$$\mathrm{H}((\tilde{w}(z))_{w\in W^\sigma}) = \mathrm{H}((\tilde{w}^\sigma(z))_{w\in W}) = \mathrm{H}((\tilde{w}(z))_{w\in W}) \tag{4.1}$$

for every $\sigma \in S_\ell$.

As $\dim W > 0$, the subspace $W \subseteq V_\ell$ contains a nonzero vector v. By Lemma 4.1, the vectors v^σ for $\sigma \in S_\ell$ span the entire space V_ℓ. In particular, we have

$$\sum_{\sigma\in S_\ell} W^\sigma = V_\ell.$$

Let $\sigma_1, \sigma_2, \dots, \sigma_q \in S_\ell$ be a sequence of permutations with $W^{\sigma_1} + \dots + W^{\sigma_q} = V_\ell$ and such that there is no shorter sequence of permutations with that property. In particular, for every $j = 1, \dots, q$, we have $W^{\sigma_j} \nsubseteq W^{\sigma_1} + \dots + W^{\sigma_{j-1}}$, because otherwise σ_j could be omitted from the sequence. Since $\ell \geqslant 2$, we have $V_\ell \neq 0$, and hence $q \geqslant 1$. For $j = 1, \dots, q$, set

$$W_j = W^{\sigma_1} + \dots + W^{\sigma_j},$$

and set $W_0 = 0$. Then, $W_q = V_l$ and $W_{j-1} \neq W_j$ for $j = 1, \dots, q$ (so W_{j-1} is a proper subspace of W_j). Let $U_j = W^{\sigma_j} \cap W_{j-1}$ for $j = 1, \dots, q$. Then,

$$\begin{aligned}\dim W_j = \dim(W_{j-1} + W^{\sigma_j}) &= \dim W_{j-1} + \dim W^{\sigma_j} - \dim U_j \\ &= \dim W_{j-1} + \dim W - \dim U_j.\end{aligned}$$

Hence, $\dim U_j = \dim W - (\dim W_j - \dim W_{j-1})$ for $j = 1, \dots, q$. In particular, $\dim U_j < \dim W$, so by the inductive assumption we have

$$\mathrm{H}((\tilde{w}(z))_{w\in U_j}) \geqslant \frac{\dim U_j}{\ell - 1}\,\mathrm{H}(\pi) = \frac{\dim W - (\dim W_j - \dim W_{j-1})}{\ell - 1}\,\mathrm{H}(\pi). \tag{4.2}$$

Recall that any $t \in T_{r,\ell}$ is uniquely determined by its images $\tilde{v}(t)$ for all $v \in V_\ell$. Hence, the random variable z is determined by $(\tilde{w}(z))_{w\in W_q}$ (recall that $W_q = V_\ell$). Conversely, $(\tilde{w}(z))_{w\in W_q}$ is clearly also determined by z. Therefore, we have $\mathrm{H}(\pi) = \mathrm{H}(z) = \mathrm{H}((\tilde{w}(z))_{w\in W_q})$. Also note that if $w = 0$, then $\mathrm{H}(\tilde{w}(z)) = 0$, hence $\mathrm{H}((\tilde{w}(z))_{w\in W_0}) = 0$ (recall that $W_0 = 0$). Using $W_0 \subseteq W_1 \subseteq \dots \subseteq W_q$, we therefore obtain

$$\mathrm{H}(\pi) = \mathrm{H}((\tilde{w}(z))_{w\in W_q}) - \mathrm{H}((\tilde{w}(z))_{w\in W_0}) = \sum_{j=1}^{q}\big[\mathrm{H}((\tilde{w}(z))_{w\in W_j}) - \mathrm{H}((\tilde{w}(z))_{w\in W_{j-1}})\big].$$

For every $j = 1, \dots, q$, we have $W_j = W_{j-1} + W^{\sigma_j}$. Hence, $(\tilde{w}(z))_{w\in W_j}$ is completely determined by $(\tilde{w}(z))_{w\in W_{j-1}}$ and $(\tilde{w}(z))_{w\in W^{\sigma_j}}$. So, $\mathrm{H}((\tilde{w}(z))_{w\in W_j}) = \mathrm{H}((\tilde{w}(z))_{w\in W_{j-1}}, (\tilde{w}(z))_{w\in W^{\sigma_j}})$ and we obtain

$$\begin{aligned}\mathrm{H}(\pi) &= \sum_{j=1}^{q}\big[\mathrm{H}((\tilde{w}(z))_{w\in W_{j-1}}, (\tilde{w}(z))_{w\in W^{\sigma_j}}) - \mathrm{H}((\tilde{w}(z))_{w\in W_{j-1}})\big] \\ &= \sum_{j=1}^{q}\mathrm{H}((\tilde{w}(z))_{w\in W^{\sigma_j}} \mid (\tilde{w}(z))_{w\in W_{j-1}}).\end{aligned}$$

Using $U_j \subseteq W_{j-1}$, this gives

$$\mathrm{H}(\pi) = \sum_{j=1}^{q} \mathrm{H}((\tilde{w}(z))_{w\in W^{\sigma_j}} \mid (\tilde{w}(z))_{w\in W_{j-1}}) \leqslant \sum_{j=1}^{q} \mathrm{H}((\tilde{w}(z))_{w\in W^{\sigma_j}} \mid (\tilde{w}(z))_{w\in U_j}).$$

Now, using $U_j \subseteq W^{\sigma_j}$, (4.1) and (4.2), we obtain

$$\begin{aligned}\mathrm{H}(\pi) &\leqslant \sum_{j=1}^{q} \big[\mathrm{H}((\tilde{w}(z))_{w\in W^{\sigma_j}}) - \mathrm{H}((\tilde{w}(z))_{w\in U_j})\big] = \sum_{j=1}^{q} \big[\mathrm{H}((\tilde{w}(z))_{w\in W}) - \mathrm{H}((\tilde{w}(z))_{w\in U_j})\big]\\ &= q\,\mathrm{H}((\tilde{w}(z))_{w\in W}) - \sum_{j=1}^{q} \mathrm{H}((\tilde{w}(z))_{w\in U_j})\\ &\leqslant q\,\mathrm{H}((\tilde{w}(z))_{w\in W}) - \sum_{j=1}^{q} \frac{\dim W - (\dim W_j - \dim W_{j-1})}{\ell-1} H(\pi)\\ &= q\,\mathrm{H}((\tilde{w}(z))_{w\in W}) - \frac{q\dim W}{\ell-1}\mathrm{H}(\pi) + \frac{\sum_{j=1}^{q}(\dim W_j - \dim W_{j-1})}{\ell-1}\mathrm{H}(\pi).\end{aligned}$$

Recall that $\dim W_0 = 0$ and $\dim W_q = \dim V_\ell = \ell - 1$. Thus,

$$\sum_{j=1}^{q}(\dim W_j - \dim W_{j-1}) = \dim W_q - \dim W_0 = \ell - 1$$

and we obtain

$$\mathrm{H}(\pi) \leqslant q\,\mathrm{H}((\tilde{w}(z))_{w\in W}) - \frac{q\dim W}{\ell-1}\mathrm{H}(\pi) + \mathrm{H}(\pi).$$

Now, rearranging gives

$$\frac{q\dim W}{\ell-1}\mathrm{H}(\pi) \leqslant q\,\mathrm{H}((\tilde{w}(z))_{w\in W}).$$

Since $q > 0$, this proves the lemma. □

We can deduce the following corollary from Lemma 4.2. Here, ν and τ are the distributions on $\{0, \dots, m-1\}$ and $T_{m-1,k}$, respectively, that we fixed earlier using Proposition 3.3. Also recall that $k \geqslant 3$.

Corollary 4.3. *Let z_τ be a random variable on $T_{m-1,k}$ with distribution τ. Then, for any subspace $W \subseteq V_k$ with $(1, \dots, 1, -(k-1)) \in W$, we have*

$$\mathrm{H}((\tilde{w}(z_\tau))_{w\in W}) \geqslant \mathrm{H}(\nu) + \frac{\dim W - 1}{k-2}(\mathrm{H}(\tau) - \mathrm{H}(\nu)).$$

Proof. Set $v = (1, \dots, 1, -(k-1)) \in W$. Furthermore, let

$$W' = \{(w_1, \dots, w_k) \in W \mid w_k = 0\}.$$

Then, W' is a subspace of W and $\dim W' = \dim W - 1$. Note that W' and v together span the entire space W. Hence, $(\tilde{w}(z_\tau))_{w\in W}$ is completely determined by $(\tilde{w}(z_\tau))_{w\in W'}$ and $\tilde{v}(z_\tau)$, so

$$\mathrm{H}((\tilde{w}(z_\tau))_{w\in W}) = \mathrm{H}((\tilde{w}(z_\tau))_{w\in W'}, \tilde{v}(z_\tau)).$$

Furthermore, note that when writing $(z_\tau^{(1)}, \dots, z_\tau^{(k)})$ for the coordinates of z_τ, we have

$$\tilde{v}(z_\tau) = z_\tau^{(1)} + \cdots + z_\tau^{(k-1)} - (k-1)z_\tau^{(k)} = (z_\tau^{(1)} + \cdots + z_\tau^{(k)}) - kz_\tau^{(k)} = (m-1) - kz_\tau^{(k)}.$$

So, the value $\tilde{v}(z_\tau)$ is in one-to-one correspondence with the last coordinate $z_\tau^{(k)}$ of z_τ. In particular, $\mathrm{H}(\tilde{v}(z_\tau)) = \mathrm{H}(z_\tau^{(k)})$. But note that the projection $z_\tau^{(k)}$ of z_τ to the last coordinate has distribution ν (since the marginal of τ is ν), so $\mathrm{H}(\tilde{v}(z_\tau)) = \mathrm{H}(z_\tau^{(k)}) = \mathrm{H}(\nu)$.

Now, we have

$$\begin{aligned}\mathrm{H}((\tilde{w}(z_\tau))_{w\in W}) - \mathrm{H}(\nu) &= \mathrm{H}((\tilde{w}(z_\tau))_{w\in W'}, \tilde{v}(z_\tau)) - \mathrm{H}(\tilde{v}(z_\tau)) = \mathrm{H}((\tilde{w}(z_\tau))_{w\in W'} \mid \tilde{v}(z_\tau))\\ &= \sum_{s\in\tilde{v}(T_{m-1,k})} \mathbb{P}(\tilde{v}(z_\tau) = s)\, \mathrm{H}((\tilde{w}(z_\tau))_{w\in W'} \mid \tilde{v}(z_\tau) = s).\end{aligned}$$

Using the one-to-one correspondence between the values $\tilde{v}(z_\tau)$ and $z_\tau^{(k)}$, the right-hand side can be rewritten in terms of $z_\tau^{(k)}$ instead of $\tilde{v}(z_\tau)$. So, we obtain

$$\mathrm{H}((\tilde{w}(z_\tau))_{w\in W}) - \mathrm{H}(\nu) = \sum_{a\in\{0,\dots,m-1\}} \mathbb{P}(z_\tau^{(k)} = a)\, \mathrm{H}((\tilde{w}(z_\tau))_{w\in W'} \mid z_\tau^{(k)} = a).$$

For each $a \in \{0, \dots, m-1\}$, the probability distribution of z_τ conditioned on $z_\tau^{(k)} = a$ gives an S_{k-1}-symmetric probability distribution on $T_{m-1-a,k-1}$ (by omitting the last coordinate $z_\tau^{(k)} = a$). Note that $k-1 \geqslant 2$ and that for all $w \in W'$, the last coordinate is zero. By omitting this last coordinate zero, we can interpret W' as a subspace of $V_{k-1} \subseteq \mathbb{Q}^{k-1}$. Note that for $w \in W'$, the value of $\tilde{w}(z_\tau)$ remains the same when we omit the last coordinate of both w and z_τ. Hence, Lemma 4.2 applied to the probability distribution of $((z_\tau^{(1)}, \dots, z_\tau^{(k-1)}) \mid z_\tau^{(k)} = a)$ on $T_{m-1-a,k-1}$ gives

$$\begin{aligned}\mathrm{H}((\tilde{w}(z_\tau))_{w\in W'} \mid z_\tau^{(k)} = a) &\geqslant \frac{\dim W'}{(k-1)-1}\, \mathrm{H}((z_\tau^{(1)}, \dots, z_\tau^{(k-1)}) \mid z_\tau^{(k)} = a)\\ &= \frac{\dim W'}{k-2}\, \mathrm{H}(z_\tau \mid z_\tau^{(k)} = a)\end{aligned}$$

for every $a \in \{0, \dots, m-1\}$. Thus,

$$\begin{aligned}&\mathrm{H}((\tilde{w}(z_\tau))_{w\in W}) - \mathrm{H}(\nu) \geqslant \frac{\dim W'}{k-2} \sum_{a\in\{0,\dots,m-1\}} \mathbb{P}(z_\tau^{(k)} = a)\, \mathrm{H}(z_\tau \mid z_\tau^{(k)} = a)\\ &= \frac{\dim W'}{k-2}\, \mathrm{H}(z_\tau \mid z_\tau^{(k)}) = \frac{\dim W - 1}{k-2} (\mathrm{H}(z_\tau) - \mathrm{H}(z_\tau^{(k)})) = \frac{\dim W - 1}{k-2} (\mathrm{H}(z_\tau) - \mathrm{H}(\nu)),\end{aligned}$$

where we again used $\mathrm{H}(z_\tau^{(k)}) = \mathrm{H}(\nu)$. Now, rearranging gives the desired inequality. □

4.2. *More preparations for the proof of Proposition 3.10*

This subsection establishes the key ingredient for the proof of Proposition 3.10, namely Lemma 4.7. We will state and prove this lemma at the end of this subsection, building on the first lemma of this subsection and on the results of the previous subsection. The actual proof of Proposition 3.10 in the next subsection will only use Lemma 4.7, but the other results from this subsection and the previous subsection are needed in order to prove Lemma 4.7.

During this entire subsection, we will operate under the following assumption, which reflects the assumption of Proposition 3.10.

Assumption 4.4. We assume that $(x_1, \dots, x_k) \in X_0^k$ is fixed with $x_1 + \dots + x_k = (m-1)\cdot \mathbb{1}^n$ and such that for every $t \in T_{m-1,k}$ the number of $j \in \{1, \dots, n\}$ with $(x_1^{(j)}, \dots, x_k^{(j)}) = t$ is exactly $\tau(t)n$. Furthermore, let z_τ be a random variable on $T_{m-1,k}$ with distribution τ.

As in the last subsection, to any vector $w = (w_1, \dots, w_k) \in \mathbb{Q}^k$, we can associate a map $\tilde{w} : T_{m-1,k} \to \mathbb{Q}$ by setting $\tilde{w}(a_1, \dots, a_k) = w_1 a_1 + \dots + w_k a_k$ for every $(a_1, \dots, a_k) \in T_{m-1,k}$. Further note that for any $x_1', \dots, x_k' \in \{0, \dots, m-1\}^n$ with $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$, we have $x_1'^{(j)} + \dots + x_k'^{(j)} = m - 1$ for each coordinate $j = 1, \dots, n$. Hence, $(x_1'^{(j)}, \dots, x_k'^{(j)}) \in T_{m-1,k}$ for each $j = 1, \dots, n$.

Lemma 4.5. *Fix a subspace $W \subseteq \mathbb{Q}^k$ and a probability distribution π on $T_{m-1,k}$. Then, there are at most*

$$\exp(\mathrm{H}(\pi)n - \mathrm{H}((\tilde{w}(z_\tau))_{w \in W})n)$$

different k-tuples $(x_1', \dots, x_k')$ with $x_1', \dots, x_k' \in \{0, \dots, m-1\}^n$, $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$ and

$$w_1 x_1' + \dots + w_k x_k' = w_1 x_1 + \dots + w_k x_k$$

for all $w = (w_1, \dots, w_k) \in W$ and such that for every $t \in T_{m-1,k}$ the number of $j \in \{1, \dots, n\}$ with $(x_1'^{(j)}, \dots, x_k'^{(j)}) = t$ is exactly $\pi(t)n$.

Proof. Let M be the number of k-tuples $(x_1', \dots, x_k')$ with the properties listed in the lemma. We need to prove that $M \leqslant \exp(\mathrm{H}(\pi)n - \mathrm{H}((\tilde{w}(z_\tau))_{w \in W})n)$. If $M = 0$, this is trivially true, so we can assume that there is at least one k-tuple $(x_1', \dots, x_k')$ with the properties listed in the lemma.

Let z_π be a random variable on $T_{m-1,k}$ with distribution π.

Claim 4.6. $\mathrm{H}((\tilde{w}(z_\pi))_{w \in W}) = \mathrm{H}((\tilde{w}(z_\tau))_{w \in W})$.

Proof. We assumed that there is at least one k-tuple $(x_1', \dots, x_k')$ with the properties listed in the lemma, so let $(x_1', \dots, x_k')$ be such a k-tuple. Then, for each $j \in \{1, \dots, n\}$, we have $w_1 x_1'^{(j)} + \dots + w_k x_k'^{(j)} = w_1 x_1^{(j)} + \dots + w_k x_k^{(j)}$ for all $w = (w_1, \dots, w_k) \in W$. In other words, for every $j \in \{1, \dots, n\}$ we have $\tilde{w}(x_1'^{(j)}, \dots, x_k'^{(j)}) = \tilde{w}(x_1^{(j)}, \dots, x_k^{(j)})$ for all $w \in W$. Note that if we choose $j \in \{1, \dots, n\}$ uniformly at random, then $(x_1'^{(j)}, \dots, x_k'^{(j)}) \in T_{m-1,k}$ will be distributed according to π and $(x_1^{(j)}, \dots, x_k^{(j)}) \in T_{m-1,k}$ will be distributed according to τ. Hence, the distributions of $(\tilde{w}(z_\pi))_{w \in W}$ and of $(\tilde{w}(z_\tau))_{w \in W}$ must agree, and in particular they must have the same entropy. Thus, $\mathrm{H}((\tilde{w}(z_\pi))_{w \in W}) = \mathrm{H}((\tilde{w}(z_\tau))_{w \in W})$. □

Note that specifying a k-tuple $(x_1', \dots, x_k')$ with $x_1', \dots, x_k' \in \{0, \dots, m-1\}^n$ and $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$ is the same as specifying $t_1, \dots, t_n \in T_{m-1,k}$ and setting $(x_1'^{(j)}, \dots, x_k'^{(j)}) = t_j$ for $1 \leqslant j \leqslant n$. As we saw in the proof of the last claim, having

$$w_1 x_1' + \dots + w_k x_k' = w_1 x_1 + \dots + w_k x_k$$

for all $w = (w_1, \dots, w_k) \in W$ is equivalent to having $\tilde{w}(x_1'^{(j)}, \dots, x_k'^{(j)}) = \tilde{w}(x_1^{(j)}, \dots, x_k^{(j)})$ for all $w \in W$ and $j = 1, \dots, n$. Hence, the sequence $t_1, \dots, t_n \in T_{m-1,k}$ must satisfy $\tilde{w}(t_j) = \tilde{w}(x_1^{(j)}, \dots, x_k^{(j)})$ for all $w \in W$ and $j = 1, \dots, n$. Therefore, the number of k-tuples $(x_1', \dots, x_k')$ with the conditions in Lemma 4.5 equals the number of sequences $t_1, \dots, t_n$ of elements of $T_{m-1,k}$ in which each element $t \in T_{m-1,k}$ occurs exactly $\pi(t)n$ times and such that $\tilde{w}(t_j) = \tilde{w}(x_1^{(j)}, \dots, x_k^{(j)})$ for all $w \in W$ and $1 \leqslant j \leqslant n$. By Lemma 2.2 applied to the map $f : T_{m-1,k} \to \mathbb{Q}^{|W|}$ given by $f(t) = (\tilde{w}(t))_{w \in W}$, this number is at most

$$\exp(\mathrm{H}(z_\pi)n - \mathrm{H}((\tilde{w}(z_\pi))_{w \in W})n) = \exp(\mathrm{H}(\pi)n - \mathrm{H}((\tilde{w}(z_\tau))_{w \in W})n).$$

Here, we used Claim 4.6. This finishes the proof of Lemma 4.5. □

Now, we can finally prove Lemma 4.7, which will be the key ingredient for the proof of Proposition 3.10.

Lemma 4.7. *Let $(x_1, \dots, x_k) \in X_0^k$ be fixed as in Assumption 4.4, and furthermore let $W \subseteq V_k$ be a fixed subspace with $(1, \dots, 1, -(k-1)) \in W \subseteq \mathbb{Q}^k$. Then, there are at most*

$$\exp\left((k-1-\dim W) \cdot \frac{\mathrm{H}(\tau) - \mathrm{H}(\nu)}{k-2} \cdot n + (D_{m,k} + m^k) \log n\right)$$

different k-tuples $(x_1', \dots, x_k') \in X_0^k$ with $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$ and

$$w_1 x_1' + \dots + w_k x_k' = w_1 x_1 + \dots + w_k x_k$$

for all $w = (w_1, \dots, w_k) \in W$.

Proof. Note that by Corollary 4.3, we have

$$\mathrm{H}((\tilde{w}(z_\tau))_{w \in W}) \geqslant \mathrm{H}(\nu) + \frac{\dim W - 1}{k-2} (\mathrm{H}(\tau) - \mathrm{H}(\nu)).$$

For $x_1', \dots, x_k' \in X_0 \subseteq \{0, \dots, m-1\}^k$ with $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$, we have $(x_1'^{(j)}, \dots, x_k'^{(j)}) \in T_{m-1,k}$ for every $j \in \{1, \dots, n\}$. So for every k-tuple $(x_1', \dots, x_k') \in X_0^k$ with the conditions in the lemma, we can define a probability distribution π on $T_{m-1,k}$ by considering $(x_1'^{(j)}, \dots, x_k'^{(j)}) \in T_{m-1,k}$ where $j \in \{1, \dots, n\}$ is chosen uniformly at random. Note that then for every $t \in T_{m-1,k}$, the number of $j \in \{1, \dots, n\}$ with $(x_1'^{(j)}, \dots, x_k'^{(j)}) = t$ is exactly $\pi(t)n$.

The projection of the probability distribution π on $T_{m-1,k}$ to the first coordinate can be described by considering $x_1'^{(j)}$ where $j \in \{1, \dots, n\}$ is chosen uniformly at random. Since $x_1' \in X_0$, choosing a coordinate $x_1'^{(j)}$ of x_1' uniformly at random gives the probability distribution ν on $\{0, \dots, m-1\}$. Hence, the projection of the probability distribution π to the first coordinate is equal to ν. Analogously, we can see that all the other coordinate projections of π are equal to ν as well. Hence, by the last condition in Proposition 3.3, we must have $\mathrm{H}(\pi) \leqslant \mathrm{H}(\tau) + (D_{m,k} \log n)/n$.

Clearly, every probability in π is an integer multiple of $1/n$. Hence, the number of possibilities for the probability distribution π on $T_{m-1,k}$ is at most $n^{|T_{m-1,k}|} \leqslant n^{m^k}$. If we fix one of these probability distributions π, then by Lemma 4.5 there are at most

$$\begin{aligned}&\exp(\mathrm{H}(\pi)n - \mathrm{H}((\tilde{w}(z_\tau))_{w \in W})n)\\ &\quad\leqslant \exp\left(\mathrm{H}(\tau)n + D_{m,k} \log n - \mathrm{H}(\nu)n - \frac{\dim W - 1}{k-2}(\mathrm{H}(\tau) - \mathrm{H}(\nu))n\right).\end{aligned}$$

different k-tuples $(x_1', \dots, x_k') \in X_0^k$ with the conditions in Lemma 4.7 that give rise to this particular probability distribution π on $T_{m-1,k}$. Thus, all in all the number of k-tuples $(x_1', \dots, x_k') \in X_0^k$ with the conditions in Lemma 4.7 is at most

$$\begin{aligned}&n^{m^k} \exp\left(\mathrm{H}(\tau)n + D_{m,k} \log n - \mathrm{H}(\nu)n - \frac{\dim W - 1}{k-2}(\mathrm{H}(\tau) - \mathrm{H}(\nu))n\right)\\ &\quad= \exp\left(\frac{k-1-\dim W}{k-2}(\mathrm{H}(\tau) - \mathrm{H}(\nu))n + (D_{m,k} + m^k) \log n\right).\end{aligned}$$

This finishes the proof of Lemma 4.7. □

4.3. *Proof of Proposition* 3.10

Proof of Proposition 3.10. Recall that $(x_1,\dots,x_k)\in X_0^k$ with $x_1+\dots+x_k=(m-1)\cdot\mathbb{1}^n$ is fixed such that for every $t\in T_{m-1,k}$ the number of $j\in\{1,\dots,n\}$ with $(x_1^{(j)},\dots,x_k^{(j)})=t$ is exactly $\tau(t)n$. So, $(x_1,\dots,x_k)\in X_0^k$ satisfies Assumption 4.4. Fix d with $1\leqslant d\leqslant k-2$. We claim that there are at most

$$km^{2k^2}n^{D_{m,k}+m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)$$

different k-tuples $(x_1',x_2',\dots,x_k')\in X_0^k$ satisfying that $x_1'+\dots+x_k'=(m-1)\cdot\mathbb{1}^n$, $\dim\operatorname{span}_{\mathbb{Q}}(x_1-x_1',\dots,x_k-x_k')=d$, and $x_i=x_i'$ for some i. Since $km^{2k^2}\leqslant n^{m^k}$ for n sufficiently large, this implies the desired bound.

Let us focus on those $(x_1',x_2',\dots,x_k')\in X_0^k$ with $x_k=x_k'$. We will show that there are at most

$$m^{2k^2}n^{D_{m,k}+m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)$$

different k-tuples $(x_1',x_2',\dots,x_k')\in X_0^k$ with $x_1'+\dots+x_k'=(m-1)\cdot\mathbb{1}^n$, $\dim\operatorname{span}_{\mathbb{Q}}(x_1-x_1',\dots,x_k-x_k')=d$, and $x_k=x_k'$. Analogously, we get the same upper bound when replacing the condition $x_k=x_k'$ with $x_i=x_i'$ for any fixed $1\leqslant i\leqslant k-1$ (note that Corollary 4.3 and Lemma 4.7 have the assumption $(1,\dots,1,-(k-1))\in W$, but they can be proved analogously if one replaces $(1,\dots,1,-(k-1))$ by one of its permutations). By a union bound, this proves the claim above, and thus Proposition 3.10.

In order to simplify notation, let us call a k-tuple $(x_1',x_2',\dots,x_k')\in X_0^k$ *relevant* if $x_1'+\dots+x_k'=(m-1)\cdot\mathbb{1}^n$, $\dim\operatorname{span}_{\mathbb{Q}}(x_1-x_1',\dots,x_k-x_k')=d$, and $x_k=x_k'$. Our goal is to prove that there are at most

$$m^{2k^2}n^{D_{m,k}+m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right)$$

relevant k-tuples $(x_1',x_2',\dots,x_k')\in X_0^k$.

For every relevant k-tuple $(x_1',x_2',\dots,x_k')\in X_0^k$, let us consider the subspace $W^*\subseteq\mathbb{Q}^k$ given by

$$W^*=\{(w_1,\dots,w_k)\in\mathbb{Q}^k\mid w_1\cdot(x_1-x_1')+\dots+w_k\cdot(x_k-x_k')=0\}.$$

Claim 4.8. *$\dim_{\mathbb{Q}}W^*=k-d$.*

Proof. Note that W^* is the null-space of the $(n\times k)$-matrix with columns $x_1-x_1',\dots,x_k-x_k'$. Since $\dim\operatorname{span}_{\mathbb{Q}}(x_1-x_1',\dots,x_k-x_k')=d$, this matrix has rank d. Therefore, the dimension of its null-space is $k-d$.

Note that W^* depends on the relevant k-tuple $(x_1',x_2',\dots,x_k')\in X_0^k$ and different relevant k-tuples can give different subspaces $W^*\subseteq\mathbb{Q}^k$. However, the following claim gives an upper bound for the total number of different subspaces W^* that can occur.

Claim 4.9. *For fixed $(x_1,\dots,x_k)$ and fixed d, the number of possible subspaces $W^*\subseteq\mathbb{Q}^k$ is at most m^{2k^2}.*

Proof. Recall that every possible W^* occurs as the null-space of an $(n\times k)$-matrix A with columns $x_1-x_1',\dots,x_k-x_k'$ for some relevant k-tuple $(x_1',\dots,x_k')$. Each row of this matrix is of the form

$$(x_1^{(j)}-x_1'^{(j)},\dots,x_k^{(j)}-x_k'^{(j)})=(x_1^{(j)},\dots,x_k^{(j)})-(x_1'^{(j)},\dots,x_k'^{(j)})$$

for some $1 \leqslant j \leqslant n$. Note that $(x_1^{(j)}, \dots, x_k^{(j)}), (x_1'^{(j)}, \dots, x_k'^{(j)}) \in T_{m-1,k}$, so each row of A is a vector from the set $\{t - t' \mid t, t' \in T_{m-1,k}\}$. Furthermore, since

$$\operatorname{rank} A = \dim \operatorname{span}_{\mathbb{Q}}(x_1 - x_1', \dots, x_k - x_k') = d,$$

we can select d linearly independent rows of the matrix A and the matrix A' formed by these d rows has the same null-space as A. Hence, W^* occurs as the null-space of a $(d \times k)$-matrix A' such that each row of A' is from the set $\{t - t' \mid t, t' \in T_{m-1,k}\}$. Since this set has size at most $|T_{m-1,k}|^2 \leqslant (m^k)^2 = m^{2k}$, there are at most $(m^{2k})^d = m^{2kd} \leqslant m^{2k^2}$ possibilities to form such a matrix A'. Hence, there are at most m^{2k^2} possible subspaces $W^* \subseteq \mathbb{Q}^k$. □

Recall that we defined the hyperplane $V_k = \{(v_1, \dots, v_k) \in \mathbb{Q}^k \mid v_1 + \dots + v_k = 0\} \subseteq \mathbb{Q}^k$. For each relevant k-tuple $(x_1', \dots, x_k') \in X_0^k$, let us consider the subspace $W \subseteq \mathbb{Q}^k$ given by

$$W = W^* \cap V_k.$$

Clearly, $W \subseteq V_k$. By Claim 4.9, there are at most m^{2k^2} possibilities for W^*, hence there are also at most m^{2k^2} possibilities for W.

Claim 4.10. $\dim_{\mathbb{Q}} W = k - d - 1$.

Proof. Note that we have

$$\begin{aligned} 1 \cdot (x_1 - x_1') + \dots + 1 \cdot (x_k - x_k') &= (x_1 + \dots + x_k) - (x_1' + \dots + x_k') \\ &= (m-1) \cdot \mathbb{1}^n - (m-1) \cdot \mathbb{1}^n = 0, \end{aligned}$$

hence $(1, \dots, 1) \in W^*$. Clearly $(1, \dots, 1) \notin V_k$, so we have $W^* \nsubseteq V_k$. Since $V_k \subseteq \mathbb{Q}^k$ is a hyperplane (that means $\dim V_k = k - 1$), this implies $\dim W = \dim(W^* \cap V_k) = \dim W^* - 1$. By Claim 4.8, this yields $\dim W = k - d - 1$, as desired. □

For any relevant k-tuple $(x_1', \dots, x_k')$, we have $x_k - x_k' = 0$ and therefore

$$\begin{aligned} &1 \cdot (x_1 - x_1') + \dots + 1 \cdot (x_{k-1} - x_{k-1}') - (k-1) \cdot (x_k - x_k') \\ &\quad = 1 \cdot (x_1 - x_1') + \dots + 1 \cdot (x_k - x_k') = 0. \end{aligned}$$

Hence, $(1, \dots, 1, -(k-1)) \in W^*$, and by $(1, \dots, 1, -(k-1)) \in V_k$ we obtain $(1, \dots, 1, -(k-1)) \in W$ for every relevant k-tuple $(x_1', \dots, x_k')$. Furthermore, if $(x_1', \dots, x_k') \in X_0^k$ is a relevant k-tuple giving rise to the subspace $W \subseteq V_k$, then for every $(w_1, \dots, w_k) \in W$ we have $w_1 \cdot (x_1 - x_1') + \dots + w_k \cdot (x_k - x_k') = 0$ and therefore

$$w_1 x_1' + \dots + w_k x_k' = w_1 x_1 + \dots + w_k x_k.$$

Also recall that every relevant k-tuple $(x_1', \dots, x_k') \in X_0^k$ must satisfy $x_1' + \dots + x_k' = (m-1) \cdot \mathbb{1}^n$. Hence, Lemma 4.7 implies that for every possible $W \subseteq V_k$, there can be at most

$$\begin{aligned} &\exp\left((k - 1 - \dim W) \cdot \frac{\mathrm{H}(\tau) - \mathrm{H}(\nu)}{k-2} \cdot n + (D_{m,k} + m^k) \log n\right) \\ &\quad = n^{D_{m,k} + m^k} \exp\left(d \cdot \frac{\mathrm{H}(\tau) - \mathrm{H}(\nu)}{k-2} \cdot n\right) \end{aligned}$$

different relevant k-tuples $(x_1',\dots,x_k')\in X_0^k$ giving rise to this subspace W. We also saw above that in total there are at most m^{2k^2} possibilities for W. Hence all in all, the number of relevant k-tuples $(x_1',\dots,x_k')\in X_0^k$ is at most

$$m^{2k^2}n^{D_{m,k}+m^k}\exp\left(\frac{\mathrm{H}(\tau)-\mathrm{H}(\nu)}{k-2}\cdot dn\right),$$

as desired. This completes the proof of Proposition 3.10. □

5. *Proof of Proposition* 3.3 *and of Lemmas* 3.2 *and* 3.9

5.1. *Proof of Lemma* 3.2

Proof of Lemma 3.2. Recall that $0<\gamma_{m,k}<1$ was chosen to minimize

$$\frac{1+\gamma+\dots+\gamma^{m-1}}{\gamma^{(m-1)/k}}=\sum_{i=0}^{m-1}\gamma^{i-(m-1)/k}.$$

Hence, the derivative of this function must be zero at the point $\gamma=\gamma_{m,k}$, so

$$\sum_{i=0}^{m-1}\left(i-\frac{m-1}{k}\right)\gamma_{m,k}^{i-1-(m-1)/k}=0.$$

Multiplying by $\gamma_{m,k}^{1+(m-1)/k}$ and rearranging yields

$$\sum_{i=0}^{m-1}i\gamma_{m,k}^i=\frac{m-1}{k}\sum_{i=0}^{m-1}\gamma_{m,k}^i. \tag{5.1}$$

Now, recall that we defined the probability distribution $\nu_{m,k}$ on $\{0,\dots,m-1\}$ by

$$\nu_{m,k}(i)=\frac{\gamma_{m,k}^i}{1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1}}.$$

Hence, by (5.1) we obtain

$$\mathbb{E}(\nu_{m,k})=\sum_{i=0}^{m-1}i\nu_{m,k}(i)=\frac{\sum_{i=0}^{m-1}i\gamma_{m,k}^i}{1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1}}=\frac{m-1}{k}\cdot\frac{\sum_{i=0}^{m-1}\gamma_{m,k}^i}{1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1}}=\frac{m-1}{k},$$

as desired. Using this, we get

$$\begin{aligned}\mathrm{H}(\nu_{m,k})&=-\sum_{i=0}^{m-1}\nu_{m,k}(i)\log\nu_{m,k}(i)=-\sum_{i=0}^{m-1}\nu_{m,k}(i)(i\log\gamma_{m,k}-\log(1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1}))\\&=-\left(\log\gamma_{m,k}\cdot\sum_{i=0}^{m-1}i\nu_{m,k}(i)\right)+\log(1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1})\\&=-\log\gamma_{m,k}\cdot\frac{m-1}{k}+\log(1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1})\\&=\log\left(\frac{1+\gamma+\dots+\gamma^{m-1}}{\gamma^{(m-1)/k}}\right)=\log\Gamma_{m,k},\end{aligned}$$

also as desired. □

5.2. *Proof of Proposition* 3.3

Before going into the proof of Proposition 3.3, we will first prove two easy lemmas.

Lemma 5.1. *If τ is an S_k-symmetric distribution on $T_{m-1,k}$, then the marginal $\mu(\tau)$ of τ has expectation $\mathbb{E}(\mu(\tau)) = (m-1)/k$.*

Proof. Let z_τ be a random variable on $T_{m-1,k}$ with distribution τ. Note that the each of the k individual coordinates of z_τ has distribution $\mu(\tau)$. Hence, each coordinate has expectation $\mathbb{E}(\mu(\tau))$, so the sum of the k coordinates of z_τ has expectation $k\,\mathbb{E}(\mu(\tau))$. On the other hand, the sum of the coordinates of z_τ is always equal to $m-1$, hence $k\,\mathbb{E}(\mu(\tau)) = m-1$, which means that $\mathbb{E}(\mu(\tau)) = (m-1)/k$. □

Let U be the vector space formed by all functions $u : T_{m-1,k} \to \mathbb{R}$ satisfying $u(t) = u(t^\sigma)$ for all $t \in T_{m-1,k}$ and $\sigma \in S_k$ (that is, u is S_k-symmetric), $\sum_{t\in T_{m-1,k}} u(t) = 0$ as well as

$$\sum_{\substack{a_2,\dots,a_k\in\{0,\dots,m-1\}\\ a+a_2+\dots+a_k=m-1}} u(a, a_2, \dots, a_k) = 0 \tag{5.2}$$

for every $a \in \{0, \dots, m-1\}$. Note that U can be interpreted as a subspace of $\mathbb{R}^{|T_{m-1,k}|}$ and that $U \subseteq \mathbb{R}^{|T_{m-1,k}|}$ has a basis consisting of points with integer coordinates. Hence, there exists some $d_{m,k} > 0$ such that for each $u \in U$ we can find $u' \in U$ such that u' has integer coordinates and $\|u - u'\|_1 \leqslant d_{m,k}$. Basically, $d_{m,k}$ is the maximum distance of any point in U from its closest integer lattice point in U. Note that $d_{m,k} > 0$ is a constant that only depends on U and its integer lattice, hence it ultimately only depends on m and k (and not on n).

Lemma 5.2. *For every integer n and every $u \in U$, we can find $u' \in U$ such that $\|u - u'\|_1 \leqslant d_{m,k}/n$ and all coordinates of $u' \in U \subseteq \mathbb{R}^{|T_{m-1,k}|}$ are integer multiples of $1/n$.*

Proof. Let us apply the definition of $d_{m,k}$ to the point $n \cdot u \in U$. There exists a point $\tilde{u} \in U$ with integer coordinates and $\|n \cdot u - \tilde{u}\|_1 \leqslant d_{m,k}$. Now, set $u' = \tilde{u}/n \in U$. Then, all coordinates of u' are integer multiples of $1/n$ and furthermore

$$\|u - u'\|_1 = \frac{1}{n}\|n \cdot u - \tilde{u}\|_1 \leqslant \frac{d_{m,k}}{n},$$

as desired. □

Now, we are ready for the proof of Proposition 3.3.

Proof of Proposition 3.3. Recall that $\nu_{m,k}$ is a probability distribution on $\{0, \dots, m-1\}$ and by Lemma 3.2, we have $\mathbb{E}(\nu_{m,k}) = (m-1)/k$ and $\mathrm{H}(\nu_{m,k}) = \log \Gamma_{m,k}$. Furthermore, by Theorem 1.4, we can fix an S_k-symmetric probability distribution $\tau_{m,k}$ on $T_{m-1,k}$ with marginal $\mu(\tau_{m,k}) = \nu_{m,k}$ and with $\tau_{m,k}(t) > 0$ for every $t \in T_{m-1,k}$. Set $0 < c_{m,k} < 1$ to be the minimum of the finitely many values $\tau_{m,k}(t) > 0$ for $t \in T_{m-1,k}$. Note that $c_{m,k}$ only depends on m and k (but not on n).

Our goal is to find a rounded version ν of $\nu_{m,k}$ (and an appropriate τ) with the properties in Proposition 3.3. Recall that n is sufficiently large and furthermore divisible by k. First, let us form a rounded version $\tilde{\tau}$ of the probability distribution $\tau_{m,k}$ on $T_{m-1,k}$.

Claim 5.3. *There is an S_k-symmetric probability distribution $\tilde{\tau}$ on $T_{m-1,k}$ such that for every $t \in T_{m-1,k}$ the probability $\tilde{\tau}(t)$ is an integer multiple of $1/n$ and furthermore $|\tilde{\tau}(t) - \tau_{m,k}(t)| \leqslant m^k/n$ for every $t \in T_{m-1,k}$.*

Proof. For this proof only, let $T'_{m-1,k}$ denote the set of those $t \in T_{m-1,k}$ that are not permutations of $(m-1, 0, \dots, 0)$. Then, the set $T_{m-1,k} \setminus T'_{m-1,k}$ consists precisely of the k permutations of $(m-1, 0, \dots, 0)$.

We can now define $\tilde{\tau}$ as follows: For every $t \in T'_{m-1,k}$ let us round the value $\tau_{m,k}(t)$ down to the next integer multiple of k/n to obtain $\tilde{\tau}(t)$. It remains to define $\tilde{\tau}(t)$ for $t \in T_{m-1,k} \setminus T'_{m-1,k}$ (that means t is one of the k permutations of $(m-1, 0, \dots, 0)$). For those t, set

$$\tilde{\tau}(t) = \frac{1}{k}\left(1 - \sum_{t' \in T'_{m-1,k}} \tilde{\tau}(t')\right) \geqslant \frac{1}{k}\left(1 - \sum_{t' \in T'_{m-1,k}} \tau_{m,k}(t')\right) \geqslant 0,$$

Then, we have

$$\sum_{t \in T_{m-1,k}} \tilde{\tau}(t) = \sum_{t' \in T'_{m-1,k}} \tilde{\tau}(t') + k \cdot \frac{1}{k}\left(1 - \sum_{t' \in T'_{m-1,k}} \tilde{\tau}(t')\right) = 1$$

and furthermore $\tilde{\tau}(t) \geqslant 0$ for every $t \in T_{m-1,k}$. Thus, $\tilde{\tau}$ is indeed a probability distribution on $T_{m-1,k}$. It is easy to see that $\tilde{\tau}$ is S_k-symmetric. Furthermore, for every $t \in T'_{m-1,k}$, the probability $\tilde{\tau}(t)$ is an integer multiple of k/n, and so in particular of $1/n$. Since n is divisible by k, we obtain that $1 - \sum_{t' \in T'_{m-1,k}} \tilde{\tau}(t')$ is an integer multiple of k/n. Hence, $\tilde{\tau}(t)$ is also an integer multiple of $1/n$ if $t \in T_{m-1,k} \setminus T'_{m-1,k}$.

Finally, for every $t \in T'_{m-1,k}$, we have $|\tilde{\tau}(t) - \tau_{m,k}(t)| \leqslant k/n \leqslant m^k/n$. Hence, for $t \in T_{m-1,k} \setminus T'_{m-1,k}$, we have

$$\begin{aligned}|\tilde{\tau}(t) - \tau_{m,k}(t)| &= \left|\frac{1}{k}\left(1 - \sum_{t' \in T'_{m-1,k}} \tilde{\tau}(t')\right) - \frac{1}{k}\left(1 - \sum_{t' \in T'_{m-1,k}} \tau_{m,k}(t')\right)\right| \\ &\leqslant \sum_{t' \in T'_{m-1,k}} \frac{1}{k}|\tilde{\tau}(t') - \tau_{m,k}(t')| \leqslant \sum_{t' \in T'_{m-1,k}} \frac{1}{k} \cdot \frac{k}{n} = \frac{|T'_{m-1,k}|}{k} \cdot \frac{k}{n} \leqslant \frac{m^k}{n}.\end{aligned}$$

All in all, we obtain $|\tilde{\tau}(t) - \tau_{m,k}(t)| \leqslant m^k/n$ for every $t \in T_{m-1,k}$. □

Now, let ν be the marginal of $\tilde{\tau}$. Then, each probability $\nu(a)$ for $a \in \{0, \dots, m-1\}$ satisfies

$$\nu(a) = \sum_{\substack{a_2, \dots, a_k \in \{0, \dots, m-1\} \\ a + a_2 + \dots + a_k = m-1}} \tilde{\tau}(a, a_2, \dots, a_k).$$

So, $\nu(a)$ is the sum of several probabilities $\tilde{\tau}(t)$ for certain $t \in T_{m-1,k}$ and is therefore an integer multiple of $1/n$. Furthermore, by Lemma 5.1, we have $\mathbb{E}(\nu) = \mathbb{E}(\mu(\tilde{\tau})) = (m-1)/k$. Thus, ν satisfies the first two properties listed in Proposition 3.3.

Since $\nu_{m,k}$ is the marginal of $\tau_{m,k}$, we also have

$$\nu_{m,k}(a) = \sum_{\substack{a_2, \dots, a_k \in \{0, \dots, m-1\} \\ a + a_2 + \dots + a_k = m-1}} \tau_{m,k}(a, a_2, \dots, a_k).$$

Hence, for every $a \in \{0, \dots, m-1\}$,

$$|\nu(a) - \nu_{m,k}(a)| \leqslant \sum_{\substack{a_2, \dots, a_k \in \{0, \dots, m-1\} \\ a + a_2 + \dots + a_k = m-1}} |\tilde{\tau}(a, a_2, \dots, a_k) - \tau_{m,k}(a, a_2, \dots, a_k)|$$

$$\leqslant \sum_{\substack{a_2,\dots,a_k\in\{0,\dots,m-1\}\\ a+a_2+\dots+a_k=m-1}} \frac{m^k}{n} \leqslant \frac{m^{2k-1}}{n}.$$

Thus, we obtain

$$\|\nu-\nu_{m,k}\|_1=\sum_{a=0}^{m-1}|\nu(a)-\nu_{m,k}(a)|\leqslant m\cdot\frac{m^{2k-1}}{n}=\frac{m^{2k}}{n}.$$

Note that for every $a\in\{0,\dots,m-1\}$, we have (recall $0<\gamma_{m,k}<1$)

$$\nu_{m,k}(a)=\frac{\gamma_{m,k}^a}{1+\gamma_{m,k}+\dots+\gamma_{m,k}^{m-1}}\geqslant\frac{\gamma_{m,k}^{m-1}}{m}.$$

Thus, as long as n is sufficiently large, we obtain for every $a\in\{0,\dots,m-1\}$

$$\nu(a)\geqslant\nu_{m,k}(a)-|\nu(a)-\nu_{m,k}(a)|\geqslant\frac{\gamma_{m,k}^{m-1}}{m}-\frac{m^{2k}}{n}>\frac{\gamma_{m,k}^{m-1}}{2m}.$$

Now, Lemma 2.3 yields

$$|\operatorname{H}(\nu)-\operatorname{H}(\nu_{m,k})|\leqslant\frac{m^{2k}}{n}\log\left(\frac{2m}{\gamma_{m,k}^{m-1}}\right).$$

So, if we set $C_{m,k}=m^{2k}\log(2m/\gamma_{m,k}^{m-1})$, we obtain using Lemma 3.2

$$\operatorname{H}(\nu)\geqslant\operatorname{H}(\nu_{m,k})-|\operatorname{H}(\nu)-\operatorname{H}(\nu_{m,k})|\geqslant\log\Gamma_{m,k}-C_{m,k}/n.$$

Thus, ν satisfies the third property in Proposition 3.3.

We now need to find a probability distribution τ on $T_{m-1,k}$ that satisfies the last three properties listed in Proposition 3.3. We will define τ in several steps. Note that as long as n is large enough, we have $c_{m,k}-m^k/n>c_{m,k}/2$ and therefore

$$\tilde\tau(t)\geqslant\tau_{m,k}(t)-|\tilde\tau(t)-\tau_{m,k}(t)|\geqslant c_{m,k}-\frac{m^k}{n}>\frac{c_{m,k}}{2}$$

for every $t\in T_{m-1,k}$.

Recall that ν is the marginal of $\tilde\tau$, so the k coordinate projections of $\tilde\tau$ are all equal to ν. Now, let τ_0 be a probability distribution on $T_{m-1,k}$ with maximal entropy under the condition that all of the k coordinate projections of τ_0 are equal to ν. For every $\sigma\in S_k$, let τ_0^σ be the probability distribution on $T_{m-1,k}$ given by permuting τ_0 according to σ, that is $\tau_0^\sigma(t)=\tau_0(t^\sigma)$ for every $t\in T_{m-1,k}$. Then, the k coordinate projections of τ_0^σ are permutations of the k coordinate projections of τ and therefore also all equal to ν. Furthermore, by symmetry, we have $\operatorname{H}(\tau_0^\sigma)=\operatorname{H}(\tau_0)$ for each $\sigma\in S_k$. Now, set

$$\tau_1=\frac{1}{k!}\sum_{\sigma\in S_k}\tau_0^\sigma.$$

Then, τ_1 is an S_k-symmetric probability distribution on $T_{m-1,k}$ and each of its k coordinate projections equals ν, so τ_1 has marginal ν. Furthermore, by concavity of the entropy function, we have

$$\operatorname{H}(\tau_1)=\operatorname{H}\left(\frac{1}{k!}\sum_{\sigma\in S_k}\tau_0^\sigma\right)\geqslant\frac{1}{k!}\sum_{\sigma\in S_k}\operatorname{H}(\tau_0^\sigma)=\frac{1}{k!}\sum_{\sigma\in S_k}\operatorname{H}(\tau_0)=\operatorname{H}(\tau_0). \tag{5.3}$$

By the choice of τ_0, this actually implies $\mathrm{H}(\tau_1) = \mathrm{H}(\tau_0)$, although this is not relevant for our argument.

Since we assumed that n is sufficiently large, we may assume $n > 2(d_{m,k}+1)/c_{m,k}$ and set

$$\tau_2 = \frac{2(d_{m,k}+1)}{c_{m,k}n}\tilde{\tau} + \left(1 - \frac{2(d_{m,k}+1)}{c_{m,k}n}\right)\tau_1.$$

Then, τ_2 is also an S_k-symmetric probability distribution on $T_{m-1,k}$ with marginal ν (since both $\tilde{\tau}$ and τ_1 have these properties). Furthermore, by concavity of the entropy function, we have

$$\begin{aligned} \mathrm{H}(\tau_2) \geqslant \frac{2(d_{m,k}+1)}{c_{m,k}n}\,\mathrm{H}(\tilde{\tau}) + \left(1 - \frac{2(d_{m,k}+1)}{c_{m,k}n}\right)\mathrm{H}(\tau_1) \geqslant \mathrm{H}(\tau_1) - \frac{2(d_{m,k}+1)}{c_{m,k}n}\,\mathrm{H}(\tau_1) \\ \geqslant \mathrm{H}(\tau_1) - \frac{2k(d_{m,k}+1)\log m}{c_{m,k}n}, \end{aligned} \tag{5.4}$$

where we used that $\mathrm{H}(\tau_1) \leqslant \log(|T_{m-1,k}|) \leqslant \log(m^k) = k\log m$.

Recall that for sufficiently large n, we have $\tilde{\tau}(t) \geqslant c_{m,k}/2$ for all $t \in T_{m-1,k}$ and hence

$$\tau_2(t) = \frac{2(d_{m,k}+1)}{c_{m,k}n}\tilde{\tau}(t) + \left(1 - \frac{2(d_{m,k}+1)}{c_{m,k}n}\right)\tau_1(t) \geqslant \frac{2(d_{m,k}+1)}{c_{m,k}n} \cdot \frac{c_{m,k}}{2} = \frac{d_{m,k}+1}{n}$$

for each $t \in T_{m-1,k}$. Since both τ_2 and $\tilde{\tau}$ are S_k-symmetric probability distributions on $T_{m-1,k}$ with marginal ν, their difference $\tau_2 - \tilde{\tau} : T_{m-1,k} \to \mathbb{R}$ lies in the space U defined above. So by Lemma 5.2, we can find some $u \in U$ with $\|(\tau_2 - \tilde{\tau}) - u\|_1 \leqslant d_{m,k}/n$ such that $u(t)$ is an integer multiple of $1/n$ for each $t \in T_{m-1,k}$. Now, define $\tau : T_{m-1,k} \to \mathbb{R}$ by

$$\tau = \tilde{\tau} + u.$$

For each $t \in T_{m-1,k}$, both $\tilde{\tau}(t)$ and $u(t)$ are integer multiples of $1/n$, and so is $\tau(t)$. We also have that

$$\|\tau_2 - \tau\|_1 = \|\tau_2 - \tilde{\tau} - u\|_1 \leqslant d_{m,k}/n.$$

In particular, for every $t \in T_{m-1,k}$, we have $|\tau(t) - \tau_2(t)| \leqslant d_{m,k}/n$ and therefore

$$\tau(t) \geqslant \tau_2(t) - \frac{d_{m,k}}{n} \geqslant \frac{d_{m,k}+1}{n} - \frac{d_{m,k}}{n} = \frac{1}{n}.$$

In particular, all values of τ are positive. Furthermore, $\sum_{t\in T_{m-1,k}} \tilde{\tau}(t) = 1$ and $\sum_{t\in T_{m-1,k}} u(t) = 0$, hence $\sum_{t\in T_{m-1,k}} \tau(t) = 1$, so τ is a probability distribution on $T_{m-1,k}$. Since both $\tilde{\tau}$ and u are S_k-symmetric, the probability distribution τ is also S_k-symmetric. Finally, the marginal of τ is ν, because for every $a \in \{0,\dots,m-1\}$ we have, using (5.2),

$$\nu(a) = \sum_{\substack{a_2,\dots,a_\ell\in\{0,\dots,r\}\\ a+a_2+\dots+a_\ell=r}} \tilde{\tau}(a,a_2,\dots,a_\ell) = \sum_{\substack{a_2,\dots,a_\ell\in\{0,\dots,r\}\\ a+a_2+\dots+a_\ell=r}} \tau(a,a_2,\dots,a_\ell).$$

Note that Lemma 2.3 yields

$$|\,\mathrm{H}(\tau_2) - \mathrm{H}(\tau)| \leqslant \|\tau_2 - \tau\|_1 \log n \leqslant (d_{m,k}\log n)/n. \tag{5.5}$$

It remains to check the last condition in Proposition 3.3. Let τ' be a probability distribution on $T_{m-1,k}$ such that each of the k coordinate projections on τ' equals ν. By the choice of τ_0, we have $\mathrm{H}(\tau') \leqslant \mathrm{H}(\tau_0)$. Hence, from (5.3), (5.4), and (5.5), we obtain

$$\mathrm{H}(\tau') \leqslant \mathrm{H}(\tau_0) \leqslant \mathrm{H}(\tau_1) \leqslant \mathrm{H}(\tau_2) + \frac{2k(d_{m,k}+1)\log m}{c_{m,k}n}$$
$$\leqslant \mathrm{H}(\tau) + \frac{d_{m,k}\log n}{n} + \frac{2k(d_{m,k}+1)\log m}{c_{m,k}n}.$$

If n is sufficiently large, this yields $\mathrm{H}(\tau') \leqslant \mathrm{H}(\tau) + (D_{m,k}\log n)/n$ with $D_{m,k} = 2d_{m,k}$. □

5.3. *Proof of Lemma* 3.9

Proof of Lemma 3.9. Let us examine the span of $(\tilde{g}_1(x_1), \dots, \tilde{g}_k(x_k), \tilde{g}_1(x_1'), \dots, \tilde{g}_k(x_k'))$ in $\mathbb{Q}^{n+k-1}$. To simplify notation, let $u_i = \tilde{g}_i(x_i)$ and $u_i' = \tilde{g}_i(x_i')$. First, recall that $x_1 + \dots + x_k = x_1' + \dots + x_k' = (m-1)\cdot \mathbb{1}^n$ implies $u_1 + \dots + u_k = 0$ and $u_1' + \dots + u_k' = 0$. Therefore, we have

$$\mathrm{span}_{\mathbb{Q}}\left(u_1, \dots, u_k, u_1', \dots, u_k'\right) = \mathrm{span}_{\mathbb{Q}}\left(u_1, \dots, u_{k-1}, u_1', \dots, u_{k-1}'\right)$$
$$= \mathrm{span}_{\mathbb{Q}}\left(u_1, \dots, u_{k-1}, u_1 - u_1', \dots, u_{k-1} - u_{k-1}'\right).$$

Let us examine the last $k-1$ coordinates of the vectors on the right-hand side. Each $u_i = \tilde{g}_i(x_i)$ for $1 \leqslant i \leqslant k-1$ is the ith standard basis vectors when restricted to the last $k-1$ coordinates. On the other hand, each $u_i - u_i' = \tilde{g}_i(x_i) - \tilde{g}_i(x_i')$ has only zeros in the last $k-1$ coordinates. Hence, all linear relations between the vectors on the right-hand side above are between $u_1 - u_1', \dots, u_{k-1} - u_{k-1}'$. This implies that

$$\dim \mathrm{span}_{\mathbb{Q}}\left(u_1, \dots, u_k, u_1', \dots, u_k'\right) = \dim \mathrm{span}_{\mathbb{Q}}\left(u_1, \dots, u_{k-1}, u_1 - u_1', \dots, u_{k-1} - u_{k-1}'\right)$$
$$= k - 1 + \dim \mathrm{span}_{\mathbb{Q}}\left(u_1 - u_1', \dots, u_{k-1} - u_{k-1}'\right).$$

By the definition of $\tilde{g}_i$, we actually know that for $1 \leqslant i \leqslant k-1$, the vector $u_i - u_i' = \tilde{g}_i(x_i) - \tilde{g}_i(x_i') \in \mathbb{Z}^{n+k-1}$ is the same as $x_i - x_i' \in \mathbb{Z}^n$ with $k-1$ zeros attached at the end. So,

$$\dim \mathrm{span}_{\mathbb{Q}}\left(u_1 - u_1', \dots, u_{k-1} - u_{k-1}'\right) = \dim \mathrm{span}_{\mathbb{Q}}\left(x_1 - x_1', \dots, x_{k-1} - x_{k-1}'\right)$$

and therefore

$$\dim \mathrm{span}_{\mathbb{Q}}\left(\tilde{g}_1(x_1), \dots, \tilde{g}_k(x_k), \tilde{g}_1(x_1'), \dots, \tilde{g}_k(x_k')\right) = \dim \mathrm{span}_{\mathbb{Q}}\left(u_1, \dots, u_k, u_1', \dots, u_k'\right)$$
$$= k - 1 + \dim \mathrm{span}_{\mathbb{Q}}\left(x_1 - x_1', \dots, x_{k-1} - x_{k-1}'\right) = k - 1 + \dim \mathrm{span}_{\mathbb{Q}}\left(x_1 - x_1', \dots, x_k - x_k'\right),$$

where in the last step, we used that $(x_1 - x_1') + \dots + (x_k - x_k') = 0$.

It remains to show that

$$\dim \mathrm{span}_{\mathbb{F}_p}\left(g_1(x_1), \dots, g_k(x_k), g_1(x_1'), \dots, g_k(x_k')\right)$$
$$= \dim \mathrm{span}_{\mathbb{Q}}\left(\tilde{g}_1(x_1), \dots, \tilde{g}_k(x_k), \tilde{g}_1(x_1'), \dots, \tilde{g}_k(x_k')\right).$$

Recall that by definition the vectors on the left-hand side are just the projections of the vectors on the right-hand side from $\mathbb{Z}^{n+k-1}$ to $\mathbb{F}_p^{n+k-1}$. Hence, the dimension on the left-hand side is at most as large as the dimension on the right-hand side. However, if the dimension on the right-hand side is ℓ, we can take ℓ independent vectors from the set $\tilde{g}_1(x_1), \dots, \tilde{g}_k(x_k), \tilde{g}_1(x_1'), \dots, \tilde{g}_k(x_k')$. Consider the $(\ell \times (n+k-1))$-matrix (with entries in $\mathbb{Z}$) whose rows are the chosen ℓ vectors. As the rank of that matrix over $\mathbb{Q}$ is ℓ, there exists an $(\ell \times \ell)$-submatrix whose determinant is a nonzero integer. Since this determinant has absolute value at most $\ell!(m-1)^\ell \leqslant (2k)!(m-1)^{2k} < P$, this implies that the determinant must be nonzero over $\mathbb{F}_P$. So, the chosen ℓ vectors are also independent over $\mathbb{F}_P$ and the dimension on the right-hand side is at least ℓ. This proves the desired equality of the two dimensions. □

6. *Proof of Theorem* 1.4

The next three sections are devoted to proving Theorem 1.4. From now on, we consider $k \geqslant 3$ to be fixed.

6.1. *A generalization of Theorem* 1.4

Instead of proving Theorem 1.4 directly, we will prove the following more general statement that applies to a certain class of probability distributions on $\{0, \dots, n\}$ with expectation n/k. Here, $n \geqslant 0$ is any nonnegative integer. We will take $n = m-1$ and use the fact that $\nu_{m,k}$ has expectation $(m-1)/k$ to obtain Theorem 1.4.

THEOREM 6.1. *Let $n \geqslant 0$ be an integer and let ψ be a probability distribution on $\{0, \dots, n\}$ with expectation n/k that satisfies $\psi(0) > \psi(1) > \dots > \psi(n) > 0$. If $n \geqslant k$, let us also assume $2\psi(\lfloor n/k \rfloor) < \psi(\lfloor n/k \rfloor - 1) + \psi(\lceil n/k \rceil)$. Then, ψ occurs as the marginal of an S_k-symmetric probability distribution τ on $T_{n,k}$ with $\tau(t) > 0$ for every $t \in T_{n,k}$.*

Proof of Theorem 1.4 *assuming Theorem* 6.1. Let $n = m-1$ and recall that $\nu_{m,k}$ is a probability distribution on $\{0, \dots, m-1\}$ with expectation $(m-1)/k$ (see Lemma 3.2). Since

$$\nu_{m,k}(i) = \frac{\gamma_{m,k}^i}{1 + \gamma_{m,k} + \dots + \gamma_{m,k}^{m-1}}.$$

for $i = 0, \dots, m-1$ and $0 < \gamma_{m,k} < 1$, we clearly have that $\nu_{m,k}(0) > \nu_{m,k}(1) > \dots > \nu_{m,k}(m-1) > 0$. So in order to be able to apply Theorem 6.1, we just need to check that

$$2\nu_{m,k}(\lfloor (m-1)/k \rfloor) < \nu_{m,k}(\lfloor (m-1)/k \rfloor - 1) + \nu_{m,k}(\lceil (m-1)/k \rceil)$$

if $m-1 \geqslant k$. If $m-1$ is divisible by k, this is clearly true. So, assume $m-1 \geqslant k$ and that $m-1$ is not divisible by k. Setting $\ell = \lfloor (m-1)/k \rfloor$ to simplify notation (note that $\ell \geqslant 1$), we need to check that

$$2\frac{\gamma_{m,k}^\ell}{\gamma_{m,k}^0 + \dots + \gamma_{m,k}^{m-1}} < \frac{\gamma_{m,k}^{\ell-1}}{\gamma_{m,k}^0 + \dots + \gamma_{m,k}^{m-1}} + \frac{\gamma_{m,k}^{\ell+1}}{\gamma_{m,k}^0 + \dots + \gamma_{m,k}^{m-1}}.$$

This is indeed true, since $2\gamma_{m,k}^\ell < \gamma_{m,k}^{\ell-1} + \gamma_{m,k}^{\ell+1}$. Thus, $\nu_{m,k}$ satisfies the assumptions of Theorem 6.1 with $n = m-1$ and therefore occurs as the marginal of an S_k-symmetric probability distribution $\tau_{m,k}$ on $T_{m-1,k}$ with $\tau_{m,k}(t) > 0$ for every $t \in T_{m-1,k}$ □

In the next subsection, we will introduce scaled distributions, following Pebody [**30**, Section 2]. They provide a useful framework for the proof of Theorem 6.1. In the third subsection we will state Proposition 6.5, the main proposition for the proof of Theorem 6.1, as well as the key lemmas for the proof of this proposition. All of this will be formulated in the framework of scaled distributions introduced in the next subsection. In the fourth subsection, we will finally see how Theorem 6.1 follows from Proposition 6.5.

Section 7 will be devoted to proving Proposition 6.5 and Section 8 to proving the key lemmas stated in Subsection 6.3. This will then complete the proof of Theorem 6.1 and thereby establish Theorem 1.4.

6.2. *Scaled distributions*

Here, we will introduce scaled distributions, the framework in which the proof of Theorem 6.1 will operate. Everything in this subsection follows the first half of Section 2 of Pebody's paper [30]. Throughout this subsection, n is an arbitrary nonnegative integer.

Definition 6.2 [30]. A *scaled distribution* ψ on $\{0, \dots, n\}$ is a map $\{0, \dots, n\} \to \mathbb{R}_{\geqslant 0}$. A scaled distribution ψ on $\{0, \dots, n\}$ has *mean* n/k if $\sum_{i=0}^{n} i\psi(i) = \frac{n}{k} \sum_{i=0}^{n} \psi(i)$.

If ψ is a scaled distribution that can be written as a linear combination of scaled distributions ψ_1 and ψ_2, and if ψ_1 and ψ_2 both have mean n/k, then ψ also has mean n/k.

Note that any probability distribution ψ on $\{0, \dots, n\}$ can be interpreted as a scaled distribution. Furthermore, to any nonzero scaled distribution ψ on $\{0, \dots, n\}$, we can associate an actual probability distribution $\overline{\psi}$ by setting

$$\overline{\psi}(i) = \frac{\psi(i)}{\psi(0) + \cdots + \psi(n)}.$$

Note that ψ has mean n/k (according to the definition above) if and only if $\overline{\psi}$ has expectation n/k (according to the usual definition in probability theory).

Following Pebody [30], let a scaled distribution on $\{0, \dots, n\}$ be called *n-simple* if it is of the form $\mathbf{1}_{a_1} + \mathbf{1}_{a_2} + \cdots + \mathbf{1}_{a_k}$ for some $(a_1, a_2, \dots, a_k) \in T_{n,k}$ (that means $a_1, a_2, \dots, a_k \in \{0, \dots, n\}$ with $a_1 + a_2 + \cdots + a_k = n$). Note that any n-simple scaled distribution has mean n/k.

The following lemma is a variation of [30, Lemma 4]. Although [30, Lemma 4] is only for the case $k = 3$ and has a slightly different statement, basically the same proof works here.

Lemma 6.3 (see [30, Lemma 4]). *Let ψ be a nonzero scaled distribution on $\{0, \dots, n\}$, and let $\overline{\psi}$ be the associated probability distribution as defined above. Then, the following two statements are equivalent.*

(1) *The scaled distribution ψ can be written as a positive linear combination of all the n-simple scaled distributions $\mathbf{1}_{a_1} + \mathbf{1}_{a_2} + \cdots + \mathbf{1}_{a_k}$ for all $(a_1, a_2, \dots, a_k) \in T_{n,k}$.*

(2) *The probability distribution $\overline{\psi}$ occurs as the marginal of an S_k-symmetric probability distribution τ on $T_{n,k}$ with $\tau(t) > 0$ for every $t \in T_{n,k}$.*

Proof. Note that for any S_k-symmetric symmetric probability distribution τ on $T_{n,k}$, its marginal $\mu(\tau)$ is given by

$$\mu(\tau) = \sum_{(a_1,\dots,a_k)\in T_{n,k}} \tau(a_1, \dots, a_k)\mathbf{1}_{a_1} = \cdots = \sum_{(a_1,\dots,a_k)\in T_{n,k}} \tau(a_1, \dots, a_k)\mathbf{1}_{a_k}.$$

First let us assume (1), so

$$\psi = \sum_{(a_1,\dots,a_k)\in T_{n,k}} \lambda(a_1, \dots, a_k) \cdot (\mathbf{1}_{a_1} + \cdots + \mathbf{1}_{a_k})$$

with $\lambda(a_1, \dots, a_k) > 0$ for all $(a_1, \dots, a_k) \in T_{n,k}$. Note that

$$\psi(0) + \cdots + \psi(n) = \sum_{(a_1,\dots,a_k)\in T_{n,k}} k\lambda(a_1, \dots, a_k) = k \sum_{t\in T_{n,k}} \lambda(t).$$

Now, consider the probability distribution τ on $T_{n,k}$ given by

$$\tau(a_1,\ldots,a_k)=\frac{\sum_{\sigma\in S_k}\lambda((a_1,\ldots,a_k)^\sigma)}{k!\sum_{t\in T_{n,k}}\lambda(t)}.$$

Clearly, τ is S_k-symmetric and $\tau(t)>0$ for every $t\in T_{n,k}$. Furthermore, its marginal $\mu(\tau)$ is given by

$$\begin{aligned}\mu(\tau)&=\frac{1}{k}\sum_{(a_1,\ldots,a_k)\in T_{n,k}}\tau(a_1,\ldots,a_k)\cdot(\mathbf{1}_{a_1}+\cdots+\mathbf{1}_{a_k})\\&=\frac{\sum_{(a_1,\ldots,a_k)\in T_{n,k}}\sum_{\sigma\in S_k}\lambda((a_1,\ldots,a_k)^\sigma)\cdot(\mathbf{1}_{a_1}+\cdots+\mathbf{1}_{a_k})}{k\cdot k!\sum_{t\in T_{n,k}}\lambda(t)}\\&=\frac{k!\sum_{(a_1,\ldots,a_k)\in T_{n,k}}\lambda(a_1,\ldots,a_k)\cdot(\mathbf{1}_{a_1}+\cdots+\mathbf{1}_{a_k})}{k!(\psi(0)+\cdots+\psi(n))}=\frac{\psi}{\psi(0)+\cdots+\psi(n)}=\overline{\psi}.\end{aligned}$$

Thus, $\overline{\psi}$ is the marginal of the S_k-symmetric probability distribution τ on $T_{n,k}$.

For the converse, assume that $\overline{\psi}$ is the marginal of some S_k-symmetric probability distribution τ on $T_{n,k}$ with $\tau(t)>0$ for every $t\in T_{n,k}$. Then,

$$\overline{\psi}=\sum_{(a_1,\ldots,a_k)\in T_{n,k}}\tau(a_1,\ldots,a_k)\mathbf{1}_{a_1}=\cdots=\sum_{(a_1,\ldots,a_k)\in T_{n,k}}\tau(a_1,\ldots,a_k)\mathbf{1}_{a_k},$$

so

$$\overline{\psi}=\sum_{(a_1,\ldots,a_k)\in T_{n,k}}\frac{\tau(a_1,\ldots,a_k)}{k}(\mathbf{1}_{a_1}+\cdots+\mathbf{1}_{a_k}).$$

Thus, $\overline{\psi}$ is a positive linear combination of all the n-simple scaled distributions. Now, by rescaling we can see that ψ is also a positive linear combination of all the n-simple scaled distributions. □

6.3. *Proof overview*

In this subsection, we will state the key ingredients for the proof of Theorem 6.1. First, we make the following definition that will be at the heart of the proof.

Definition 6.4. Let $n\geqslant 0$ be an integer. A scaled distribution ψ on $\{0,\ldots,n\}$ is called *n-tame* if the following four statements hold.

(i) ψ has mean n/k, that is $\sum_{i=0}^n i\psi(i)=\frac{n}{k}\sum_{i=0}^n\psi(i)$.
(ii) If $n\geqslant 1$, then $\psi(1)\geqslant\psi(2)\geqslant\cdots\geqslant\psi(n)$.
(iii) If $n\geqslant k$, then $\psi(0)\geqslant(k-1)\psi(n)+(k-2)\psi(n-1)+\cdots+\psi(n-k+2)$.
(iv) If $n\geqslant 2k$, then $2\psi(\lfloor n/k\rfloor)\leqslant\psi(\lfloor n/k\rfloor-1)+\psi(\lceil n/k\rceil)$.

In other words, if $n=0$, then ψ only needs to satisfy condition (i). If $1\leqslant n<k$, the ψ needs to satisfy (i) and the inequality in (ii). If $k\leqslant n<2k$, then in addition ψ also needs to satisfy the inequality in (iii). And if $n\geqslant 2k$, then ψ needs to satisfy (i) and all the three inequalities in (ii), (iii), and (iv). Note that if n is divisible by k, then condition (iv) is already implied by condition (ii), since $\lfloor n/k\rfloor=\lceil n/k\rceil$.

We are now ready to state the main proposition for the proof of Theorem 6.1.

Proposition 6.5. *Every n-tame scaled distribution is a nonnegative linear combination of n-simple scaled distributions.*

The proof of Proposition 6.5 will be by strong induction on n with base cases $0 \leqslant n \leqslant k$, and reducing from n to $n-k$ in each step. We will prove Proposition 6.5 in Section 7. Here, we just state the key lemmas for the proof of Proposition 6.5. These lemmas will be proved in Section 8.

Lemma 6.6. *Let $n \geqslant 1$ and let $j \in \{1, \dots, n\}$ with $j+1 \geqslant 2n/k$. Then, there exists a scaled distribution α_j on $\{0, \dots, n\}$ satisfying the following six properties.*

(a) *α_j is a nonnegative linear combination of n-simple scaled distributions.*
(b) *α_j has mean n/k.*
(c) *$\alpha_j(1) \leqslant \alpha_j(2) \leqslant \cdots \leqslant \alpha_j(j)$.*
(d) *$\alpha_j(j+1) = \alpha_j(j+2) = \cdots = \alpha_j(n) = 0$.*
(e) *$\alpha_j(j) \neq 0$.*
(f) *If $n \geqslant 2k$, then $2\alpha_j(\lfloor n/k \rfloor) \geqslant \alpha_j(\lfloor n/k \rfloor - 1) + \alpha_j(\lceil n/k \rceil)$.*

Lemma 6.7. *Let $n \geqslant k$ be an integer and let ψ be an n-tame scaled distribution on $\{0, \dots, n\}$. Then, there exists an n-tame scaled distribution φ with*

$$\varphi(0) = (k-1)\varphi(n) + (k-2)\varphi(n-1) + \cdots + \varphi(n-k+2) \tag{6.1}$$

(this means that φ has equality in condition (iii) in Definition 6.4) and such that $\psi - \varphi$ is a nonnegative linear combination of n-simple scaled distributions.

Lemma 6.8. *Let $n \geqslant 2k$ be an integer and let φ be an n-tame scaled distribution on $\{0, \dots, n\}$ with*

$$\varphi(0) = (k-1)\varphi(n) + (k-2)\varphi(n-1) + \cdots + \varphi(n-k+2).$$

Then,

$$\begin{aligned}\varphi(1) \geqslant\ & \varphi(n-1) + 2\varphi(n-2) + \cdots + (k-2)\varphi(n-k+2) \\ & + (k-1)\varphi(n-k+1) + (k-2)\varphi(n-k) + \cdots + \varphi(n-2k+3).\end{aligned}$$

Remark 6.9. Recall that Proposition 6.5 states that every n-tame scaled distribution ψ is a nonnegative linear combination of n-simple scaled distributions. It is not hard to see that conditions (i) and (iii) in Definition 6.4 are necessary in order for ψ to be a nonnegative linear combination of n-simple scaled distributions. Conditions (ii) and (iv), on the other hand, were chosen because they make the inductive proof of Proposition 6.5 work. Condition (iv) seems very unnatural and is in fact only needed to make Lemma 6.8 true. The inequality there can fail by just a slight amount if one does not have this condition (and if n is not divisible by k). One can probably replace (iv) by a different condition to ensure the inequality in Lemma 6.8 (this particular condition was chosen because it is easy to check for the probability distribution $\nu_{m,k}$). However, omitting condition (iv) entirely would make Proposition 6.5 false.

6.4. *Deriving Theorem 6.1 from Proposition 6.5*

First, let us prove the following corollary of Lemma 6.8.

Corollary 6.10. *Let $n \geqslant 0$ be an integer and let ψ be a scaled distribution on $\{0, \dots, n\}$ with mean n/k that satisfies $\psi(0) \geqslant \psi(1) \geqslant \cdots \geqslant \psi(n)$. If $n \geqslant k$, let us also assume $2\psi(\lfloor n/k \rfloor) \leqslant \psi(\lfloor n/k \rfloor - 1) + \psi(\lceil n/k \rceil)$. Then, ψ is n-tame.*

Proof. We just need to show that condition (iii) in Definition 6.4 is satisfied, all the other conditions are clear from the assumptions of the corollary. So, let us assume $n \geqslant k$ (otherwise

(iii) is vacuous). Let us define a scaled distribution φ on $\{0,\dots,n+k\}$ by taking $\varphi(0)=0$ and $\varphi(n+2)=\varphi(n+3)=\cdots=\varphi(n+k)=0$ and $\varphi(i)=\psi(i-1)$ for $i=1,\dots,n+1$. We claim that φ is $(n+k)$-tame.

(i) φ has mean $(n+k)/k$, because

$$\begin{aligned}\sum_{i=0}^{n+k} i\varphi(i) &= \sum_{i=1}^{n+1} i\varphi(i) = \sum_{i=0}^{n}(i+1)\psi(i) = \frac{n}{k}\sum_{i=0}^{n}\psi(i)+\sum_{i=0}^{n}\psi(i)\\ &= \frac{n+k}{k}\sum_{i=1}^{n+1}\varphi(i) = \frac{n+k}{k}\sum_{i=0}^{n+k}\varphi(i).\end{aligned}$$

Here, we used that ψ has mean n/k.

(ii) From $\psi(0)\geqslant\psi(1)\geqslant\cdots\geqslant\psi(n)$, we obtain $\varphi(1)\geqslant\varphi(2)\geqslant\cdots\geqslant\varphi(n+1)$, and together with $\varphi(n+2)=\varphi(n+3)=\cdots=\varphi(n+k)=0$ this gives $\varphi(1)\geqslant\varphi(2)\geqslant\cdots\geqslant\varphi(n+k)$.

(iii) We have

$$\varphi(0)=(k-1)\varphi(n+k)+(k-2)\varphi(n+k-1)+\cdots+\varphi(n+2), \tag{6.2}$$

because all these terms are zero by the definition of φ.

(iv) Recall that we assumed $n\geqslant k$. So by the assumptions on ψ, we have $2\psi(\lfloor n/k\rfloor)\leqslant\psi(\lfloor n/k\rfloor-1)+\psi(\lceil n/k\rceil)$ and therefore

$$\begin{aligned}2\varphi(\lfloor(n+k)/k\rfloor) &= 2\varphi(\lfloor n/k\rfloor+1)=2\psi(\lfloor n/k\rfloor)\\ &\leqslant \psi(\lfloor n/k\rfloor-1)+\psi(\lceil n/k\rceil)=\varphi(\lfloor(n+k)/k\rfloor-1)+\varphi(\lceil(n+k)/k\rceil).\end{aligned}$$

So, φ is indeed $(n+k)$-tame. Note that $n+k\geqslant 2k$, since we assumed $n\geqslant k$. So by (6.2), we can apply Lemma 6.8 and obtain

$$\begin{aligned}\varphi(1)\geqslant{}&\varphi(n+k-1)+2\varphi(n+k-2)+\cdots+(k-2)\varphi(n+2)\\ &+(k-1)\varphi(n+1)+(k-2)\varphi(n)+\cdots+\varphi(n-k+3).\end{aligned}$$

When plugging in the definition of φ, we get $\psi(0)\geqslant(k-1)\psi(n)+(k-2)\psi(n-1)+\cdots+\psi(n-k+2)$. Thus, ψ indeed satisfies condition (iii) in Definition 6.4. □

Now, we are ready to derive Theorem 6.1 from Proposition 6.5.

Proof of Theorem 6.1. Let ψ be a probability distribution on $\{0,\dots,n\}$ with expectation n/k and $\psi(0)>\psi(1)>\cdots>\psi(n)>0$. If $n\geqslant k$, we also assume $2\psi(\lfloor n/k\rfloor)<\psi(\lfloor n/k\rfloor-1)+\psi(\lceil n/k\rceil)$. We need to prove that ψ occurs as the marginal of an S_k-symmetric probability distribution τ on $T_{n,k}$ with $\tau(t)>0$ for every $t\in T_{n,k}$. By Lemma 6.3, it is enough to show that ψ, interpreted as a scaled distribution, can be written as a positive linear combination of all the n-simple scaled distributions $\mathbf{1}_{a_1}+\mathbf{1}_{a_2}+\cdots+\mathbf{1}_{a_k}$ for all $(a_1,a_2,\dots,a_k)\in T_{n,k}$.

So, let us interpret ψ as a scaled distribution and let us take some small $x>0$ such that

$$\psi'=\psi-x\cdot\sum_{(a,\dots,a_k)\in T_{n,k}}(\mathbf{1}_{a_1}+\cdots+\mathbf{1}_{a_k})$$

is a scaled distribution with $\psi'(0)\geqslant\psi'(1)\geqslant\cdots\geqslant\psi'(n)\geqslant 0$ and with $2\psi'(\lfloor n/k\rfloor)\leqslant\psi'(\lfloor n/k\rfloor-1)+\psi'(\lceil n/k\rceil)$ if $n\geqslant k$. Note that ψ' has mean n/k, since both ψ and all the n-simple scaled distributions $\mathbf{1}_{a_1}+\cdots+\mathbf{1}_{a_k}$ have mean n/k. So by Corollary 6.10, the scaled distribution ψ' is n-tame. Hence, by Proposition 6.5, ψ' is a nonnegative linear combination on n-simple scaled distributions. Now, using $x > 0$ and

$$\psi = \psi' + x \cdot \sum_{(a,\dots,a_k)\in T_{n,k}} (\mathbf{1}_{a_1} + \dots + \mathbf{1}_{a_k}),$$

this gives a way to express ψ as a positive linear combination of all the n-simple scaled distributions $\mathbf{1}_{a_1} + \mathbf{1}_{a_2} + \dots + \mathbf{1}_{a_k}$ for all $(a_1, a_2, \dots, a_k) \in T_{n,k}$. This finishes the proof of Theorem 6.1. □

The next section will be devoted to proving Proposition 6.5, assuming the lemmas stated in Subsection 6.3. These lemmas will be proved in Section 8.

7. *Proof of Proposition* 6.5

We will prove Proposition 6.5 by strong induction on n, in each step reducing any instance of the statement for n to an instance of the statement for $n - k$. The base cases $n = 0, \dots, k$ will be treated in the first subsection. The induction step will be performed in the second subsection.

7.1. *The base cases of the induction*

Here, we will treat the cases $n = 0, \dots, k$ of Proposition 6.5.

If $n = 0$, then we have as desired

$$\psi = (\psi(0)/k)(\underbrace{\mathbf{1}_0 + \dots + \mathbf{1}_0}_{k}).$$

Now, assume that $1 \leqslant n \leqslant k$ and keep n fixed. Then, Proposition 6.5 is true by the following claim.

Claim 7.1. *Let $1 \leqslant n \leqslant k$. Any scaled distribution ψ on $\{0, \dots, n\}$ satisfying both*

(i) ψ has mean n/k and
(ii) $\psi(1) \geqslant \psi(2) \geqslant \dots \geqslant \psi(n)$.

can be written as a nonnegative linear combination of n-simple scaled distributions.

Proof. Suppose that there are counterexamples to the claim for some fixed $1 \leqslant n \leqslant k$. Then, we can find a scaled distribution ψ satisfying (i) and (ii) such that ψ is not a nonnegative linear combination of n-simple scaled distributions. Clearly, such a ψ is not identically zero. For any such ψ, let $j \in \{0, \dots, n\}$ be chosen maximal with $\psi(j) > 0$. Among all possible scaled distributions ψ that contradict the claim, let us choose one where this j is minimal. Thus, ψ is a scaled distribution on $\{0, \dots, n\}$ satisfying (i) and (ii), but ψ cannot be written as a nonnegative linear combination of n-simple scaled distributions. Furthermore, we have $\psi(j+1) = \dots = \psi(n) = 0$, but $\psi(j) > 0$. Finally, by the choice of ψ, any scaled distribution φ satisfying (i) and (ii) and $\varphi(j) = \varphi(j+1) = \dots = \varphi(n) = 0$ can be written as a nonnegative linear combination of n-simple scaled distributions.

Note that $j \neq 0$, because otherwise ψ is supported just on zero and can therefore not have mean n/k. Thus, $j \in \{1, \dots, n\}$ and in particular $j + 1 \geqslant 2 \geqslant 2n/k$. Thus, Lemma 6.6 gives a scaled distribution α_j satisfying the conditions (a) to (f). Now, choose $x \in \mathbb{R}_{\geqslant 0}$ maximal such that both $\psi(0) - x\alpha_j(0) \geqslant 0$ and $\psi(j) - x\alpha_j(j) \geqslant 0$ (such a maximal x exists since $x = 0$ satisfies the conditions, but due to $\alpha_j(j) > 0$ by (e) there is an upper bound for x). Note that for this maximal x, we have $\psi(0) - x\alpha_j(0) = 0$ or $\psi(j) - x\alpha_j(j) = 0$.

Now, set $\varphi = \psi - x\alpha_j$. We claim that φ is a scaled distribution satisfying (i) and (ii). Note that we have

$$\varphi(0) = \psi(0) - x\alpha_j(0) \geqslant 0$$

and

$$\varphi(j) = \psi(j) - x\alpha_j(j) \geqslant 0.$$

Furthermore, by (ii) and (c), we have $\psi(1) \geqslant \psi(2) \geqslant \cdots \geqslant \psi(j)$ and $\alpha_j(1) \leqslant \alpha_j(2) \leqslant \cdots \leqslant \alpha_j(j)$, hence

$$\varphi(1) \geqslant \varphi(2) \geqslant \cdots \geqslant \varphi(j) \geqslant 0.$$

From $\psi(j+1) = \cdots = \psi(n) = 0$ and (d), we can deduce that $\varphi(j+1) = \cdots = \varphi(n) = 0$. In particular, we have established that φ has nonnegative values, so it is a scaled distribution. We also have

$$\varphi(1) \geqslant \varphi(2) \geqslant \cdots \geqslant \varphi(j) \geqslant 0 = \varphi(j+1) = \cdots = \varphi(n),$$

so φ satisfies (ii). Finally $\varphi = \psi - x\alpha_j$ has mean n/k because both ψ and α_j have mean n/k (see (i) and (b)). Thus, φ also satisfies (i).

We claim that φ is a nonnegative linear combination of n-simple scaled distributions. Recall that by the choice of x, we have $\psi(0) - x\alpha_j(0) = 0$ or $\psi(j) - x\alpha_j(j) = 0$. This means $\varphi(0) = 0$ or $\varphi(j) = 0$. If $\varphi(j) = 0$, then φ is a scaled distribution satisfying (i) and (ii) and $\varphi(j) = \varphi(j+1) = \cdots = \varphi(n) = 0$. We saw above that by the choice of ψ this indeed implies that φ is a nonnegative linear combination of n-simple scaled distributions.

So, now suppose that $\varphi(0) = 0$. Then, φ is supported on a subset of $\{1, \dots n\}$, but has mean $n/k \leqslant 1$. This is only possible if $n = k$ and φ is supported on $\{1\}$. But then

$$\varphi = (\varphi(1)/k)(\underbrace{\mathbf{1}_1 + \cdots + \mathbf{1}_1}_{k})$$

and therefore φ is a nonnegative linear combination of n-simple scaled distributions.

So, we have shown in any case that φ is a nonnegative linear combination of n-simple scaled distributions. But now, using (a), we see that $\psi = \varphi + x\alpha_j$ is also a nonnegative linear combination of n-simple scaled distributions. This contradicts our choice of ψ. Hence, there cannot be any counterexamples to the claim, so the claim is true. □

7.2. *The induction step*

We will now perform the induction step for proving Proposition 6.5. Let us assume that $n \geqslant k+1$ and that we have already proved Proposition 6.5 for all smaller values of n. Let ψ be an n-tame scaled distribution on $\{0, \dots, n\}$. We need to show that ψ can be written as a nonnegative linear combination of n-simple scaled distributions.

First, we apply Lemma 6.7 to obtain an n-tame scaled distribution φ satisfying (6.1) and such that $\psi - \varphi$ is a nonnegative linear combination of n-simple scaled distributions. If we can write φ as a nonnegative linear combination of n-simple scaled distributions, then $\psi = \varphi + (\psi - \varphi)$ will also be a nonnegative linear combination of n-simple scaled distributions. Hence, it suffices to show that φ is a nonnegative linear combination of n-simple scaled distributions.

If $n \geqslant 2k$, then we can apply Lemma 6.8 and obtain

$$\begin{aligned}\varphi(1) - \varphi(n-1) - 2\varphi(n-2) - \cdots - (k-2)\varphi(n-k+2)\\ \geqslant (k-1)\varphi(n-k+1) + (k-2)\varphi(n-k) + \cdots + \varphi(n-2k+3).\end{aligned} \tag{7.1}$$

Let us now define a new scaled distribution ϑ on $\{0,\dots,n\}$ by

$$\vartheta = \varphi - \sum_{i=0}^{k-2} \varphi(n-i)(\mathbf{1}_{n-i} + \underbrace{\mathbf{1}_1 + \dots + \mathbf{1}_1}_{i} + \underbrace{\mathbf{1}_0 + \dots + \mathbf{1}_0}_{k-1-i}).$$

Note that ϑ satisfies

$$\sum_{i=0}^{n} i\vartheta(i) = \frac{n}{k}\sum_{i=0}^{n}\vartheta(i), \tag{7.2}$$

since ϑ is a linear combination of φ and certain n-simple scaled distributions, all of which have mean n/k. Let us show that ϑ indeed has nonnegative values. For $\vartheta(0)$, we have

$$\vartheta(0) = \varphi(0) - (k-1)\varphi(n) - (k-2)\varphi(n-1) - \dots - \varphi(n-k+2) = 0,$$

where for the second equality sign we used (6.1). We also have that

$$\vartheta(n) = \vartheta(n-1) = \dots = \vartheta(n-k+2) = 0.$$

Finally, $\vartheta(i) = \varphi(i) \geqslant 0$ for $2 \leqslant i \leqslant n-k+1$. Therefore, $\vartheta(i) \geqslant 0$ for $i = 0$ and for $2 \leqslant i \leqslant n$. Let us now check that $\vartheta(1) \geqslant 0$ as well. If $k+1 \leqslant n < 2k$, using (7.2) and $\vartheta(0) = 0$, we have

$$\sum_{i=2}^{n}\left(i - \frac{n}{k}\right)\vartheta(i) = \left(\frac{n}{k} - 1\right)\vartheta(1) + \frac{n}{k}\vartheta(0) = \left(\frac{n}{k} - 1\right)\vartheta(1).$$

Since the left-hand side is nonnegative, the right-hand side must also be nonnegative and therefore $\vartheta(1) \geqslant 0$. If $n \geqslant 2k$, then (7.1) implies

$$\begin{aligned}\vartheta(1) &= \varphi(1) - \varphi(n-1) - 2\varphi(n-2) - \dots - (k-2)\varphi(n-k+2)\\ &\geqslant (k-1)\varphi(n-k+1) + (k-2)\varphi(n-k) + \dots + \varphi(n-2k+3)\\ &= (k-1)\vartheta(n-k+1) + (k-2)\vartheta(n-k) + \dots + \vartheta(n-2k+3), \end{aligned} \tag{7.3}$$

and in particular $\vartheta(1) \geqslant 0$. This proves that ϑ has nonnegative values and is therefore a scaled distribution.

We have already seen that it suffices to prove that φ is a nonnegative linear combination of n-simple scaled distributions. Given that

$$\varphi = \vartheta + \sum_{i=0}^{k-2} \varphi(n-i)(\mathbf{1}_{n-i} + \underbrace{\mathbf{1}_1 + \dots + \mathbf{1}_1}_{i} + \underbrace{\mathbf{1}_0 + \dots + \mathbf{1}_0}_{k-1-i}),$$

it is therefore sufficient to show that ϑ is a nonnegative linear combination of n-simple scaled distributions.

Using property (ii) of φ together with $\vartheta(i) = \varphi(i)$ for $2 \leqslant i \leqslant n-k+1$, we obtain

$$\vartheta(2) \geqslant \vartheta(3) \geqslant \dots \geqslant \vartheta(n-k+1).$$

Also, if $n \geqslant 3k$, then $2 \leqslant \lfloor n/k \rfloor - 1 < \lfloor n/k \rfloor \leqslant \lceil n/k \rceil \leqslant n-k+1$ and therefore property (iv) of φ implies

$$2\vartheta(\lfloor n/k \rfloor) \leqslant \vartheta(\lfloor n/k \rfloor - 1) + \vartheta(\lceil n/k \rceil).$$

Now, let us define a scaled distribution η on $\{0,\dots,n-k\}$ by $\eta(i) = \vartheta(i+1)$ for $i = 0,\dots,n-k$. We will prove that η is an $(n-k)$-tame scaled distribution and then use the induction assumption. So, let us check that η satisfies conditions (i)–(iv) for an $(n-k)$-tame scaled distribution (see Definition 6.4).

(i) Recall that $\vartheta(0) = 0$ and $\vartheta(n) = \vartheta(n-1) = \cdots = \vartheta(n-k+2) = 0$. Now, using (7.2), we obtain

$$\sum_{i=0}^{n-k} i\eta(i) = \sum_{i=1}^{n-k+1} (i-1)\vartheta(i) = \sum_{i=0}^{n}(i-1)\vartheta(i) = \sum_{i=0}^{n} i\vartheta(i) - \sum_{i=0}^{n}\vartheta(i)$$
$$= \left(\frac{n}{k} - 1\right)\sum_{i=0}^{n}\vartheta(i) = \frac{n-k}{k}\sum_{i=1}^{n-k+1}\vartheta(i) = \frac{n-k}{k}\sum_{i=0}^{n-k}\eta(i).$$

(ii) Recall that $\vartheta(2) \geqslant \vartheta(3) \geqslant \cdots \geqslant \vartheta(n-k+1)$. This directly implies $\eta(1) \geqslant \eta(2) \geqslant \cdots \geqslant \eta(n-k)$.

(iii) Assume that $n-k \geqslant k$, which means $n \geqslant 2k$. Then, (7.3) implies

$$\eta(0) \geqslant (k-1)\eta(n-k) + (k-2)\eta(n-k-1) + \cdots + \eta(n-2k+2).$$

(iv) Assume that $n-k \geqslant 2k$, which means $n \geqslant 3k$. Recall that we have $2\vartheta(\lfloor n/k \rfloor) \leqslant \vartheta(\lfloor n/k \rfloor - 1) + \vartheta(\lceil n/k \rceil)$ in this case. Thus,

$$2\eta(\lfloor (n-k)/k \rfloor) = 2\vartheta(\lfloor n/k \rfloor) \leqslant \vartheta(\lfloor n/k \rfloor - 1) + \vartheta(\lceil n/k \rceil)$$
$$= \eta(\lfloor (n-k)/k \rfloor - 1) + \eta(\lceil (n-k)/k \rceil).$$

Hence, η is indeed an $(n-k)$-tame scaled distribution. By the induction assumption for $n-k$, we can conclude that η can be written as a nonnegative linear combination of $(n-k)$-simple scaled distributions. So, let

$$\eta = \sum_{(a_1,\ldots,a_k)\in T_{n-k,k}} \lambda(a_1,\ldots,a_k)\cdot(\mathbf{1}_{a_1} + \cdots + \mathbf{1}_{a_k})$$

with $\lambda(a_1,\ldots,a_k) \geqslant 0$ for all $(a_1,\ldots,a_k) \in T_{n-k,k}$. Using that $\eta(i) = \vartheta(i+1)$ for $i = 0,\ldots,n-k$ as well as $\vartheta(0) = 0$ and $\vartheta(n) = \vartheta(n-1) = \cdots = \vartheta(n-k+2) = 0$, we obtain

$$\vartheta = \sum_{(a_1,\ldots,a_k)\in T_{n-k,k}} \lambda(a_1,\ldots,a_k)\cdot(\mathbf{1}_{a_1+1} + \cdots + \mathbf{1}_{a_k+1}).$$

Note that for every $(a_1,\ldots,a_k) \in T_{n-k,k}$, we have $(a_1+1,\ldots,a_k+1) \in T_{n,k}$. Thus, for every $(a_1,\ldots,a_k) \in T_{n-k,k}$ the scaled distribution $\mathbf{1}_{a_1+1} + \cdots + \mathbf{1}_{a_k+1}$ is n-simple. So, the above equation establishes that ϑ is a nonnegative linear combination of n-simple scaled distributions. This finishes the proof of Proposition 6.5.

8. *Proof of Lemmas* 6.6, 6.7, *and* 6.8

8.1. *Proof of Lemma* 6.6

We will prove Lemma 6.6 by constructing the scaled distributions α_j explicitly (distinguishing several cases).

Proof of Lemma 6.6. Recall that $n \geqslant 1$ and $j \in \{1,\ldots,n\}$ with $j+1 \geqslant 2n/k$. First, let us check that $j \geqslant n/k$. If $n \geqslant k$, this follows from

$$j \geqslant \frac{2n}{k} - 1 \geqslant \frac{n}{k} + 1 - 1 = \frac{n}{k}.$$

If $n < k$, then $j \geqslant 1 \geqslant n/k$. So, we indeed have $j \geqslant n/k$ in either case.

Now, let $n = \ell j + r$ for integers ℓ and r with $0 \leqslant r \leqslant j-1$ (so r is the remainder of n upon division by j). As $n/k \leqslant j \leqslant n$, we have $1 \leqslant \ell \leqslant k$. If $\ell = k$, then from $j \geqslant n/k$ we get that actually $j = n/k$ and we can just take

$$\alpha_j = \underbrace{\mathbf{1}_{n/k} + \cdots + \mathbf{1}_{n/k}}_{k} = k\mathbf{1}_{n/k}.$$

It is easy to check that this α_j satisfies the conditions (a)–(f) in Lemma 6.6. So from now on, let us assume that $1 \leqslant \ell \leqslant k-1$. We will distinguish several cases. In each case, we will give an explicit definition of α_j and then check the conditions (a)–(f) in Lemma 6.6.

Case 1: $r \neq \lceil n/k \rceil$ *or* $n < 2k$. In this case, set

$$\alpha_j = \sum_{i=r}^{j} (\underbrace{\mathbf{1}_j + \cdots + \mathbf{1}_j}_{\ell-1} + \mathbf{1}_i + \mathbf{1}_{j+r-i} + \underbrace{\mathbf{1}_0 + \cdots + \mathbf{1}_0}_{k-\ell-1}).$$

Evaluating this gives

$$\alpha_j(i) = \begin{cases} (k-\ell-1)(j+1-r) & \text{if } i = 0 \\ 0 & \text{if } 1 \leqslant i \leqslant r-1 \\ 2 & \text{if } r \leqslant i \leqslant j-1 \\ (\ell-1)(j+1-r)+2 & \text{if } i = j \\ 0 & \text{if } j+1 \leqslant i \end{cases}$$

if $r \geqslant 1$ and

$$\alpha_j(i) = \begin{cases} (k-\ell-1)(j+1-r)+2 & \text{if } i = 0 \\ 2 & \text{if } 1 \leqslant i \leqslant j-1 \\ (\ell-1)(j+1-r)+2 & \text{if } i = j \\ 0 & \text{if } j+1 \leqslant i \end{cases}$$

if $r = 0$. By definition, α_j satisfies (a) and therefore automatically also (b). By looking at the evaluations above, it is easy to see that α_j satisfies (c), (d), and (e). It remains to check (f). If $n < 2k$, then (f) is vacuously true, so let us assume $n \geqslant 2k$. Then,

$$j \geqslant \frac{2n}{k} - 1 \geqslant \frac{n}{k} + 2 - 1 = \frac{n}{k} + 1 > \lceil n/k \rceil.$$

Thus, $1 \leqslant \lfloor n/k \rfloor - 1 \leqslant \lfloor n/k \rfloor \leqslant \lceil n/k \rceil \leqslant j-1$. Hence, each $i \in \{\lfloor n/k \rfloor - 1, \lfloor n/k \rfloor, \lceil n/k \rceil\}$ satisfies $\alpha_j(i) \in \{0, 2\}$, and $\alpha_j(i) = 2$ if and only if $i \geqslant r$. We need to show

$$2\alpha_j(\lfloor n/k \rfloor) \geqslant \alpha_j(\lfloor n/k \rfloor - 1) + \alpha_j(\lceil n/k \rceil). \tag{8.1}$$

Recall that we assumed $r \neq \lceil n/k \rceil$ for this case. If $r > \lceil n/k \rceil$, then each of the values $\alpha_j(\lfloor n/k \rfloor)$, $\alpha_j(\lfloor n/k \rfloor - 1)$ and $\alpha_j(\lceil n/k \rceil)$ is zero, hence (8.1) is satisfied. If $r < \lceil n/k \rceil$, then $r \leqslant \lfloor n/k \rfloor$ and so $\alpha_j(\lfloor n/k \rfloor) = 2$. Thus,

$$2\alpha_j(\lfloor n/k \rfloor) = 4 = 2 + 2 \geqslant \alpha_j(\lfloor n/k \rfloor - 1) + \alpha_j(\lceil n/k \rceil)$$

and (8.1) is satisfied as well.

Case 2: $r = \lceil n/k \rceil$, $n \geqslant 2k$, *and* $\ell \leqslant k-2$. In this case, set

$$\alpha_j = \sum_{i=r+1}^{j-1} (\underbrace{\mathbf{1}_j + \cdots + \mathbf{1}_j}_{\ell-1} + \mathbf{1}_i + \mathbf{1}_{j+r-i} + \underbrace{\mathbf{1}_0 + \cdots + \mathbf{1}_0}_{k-\ell-1})$$

$$+ \frac{2}{r+1} \sum_{i=0}^{r} (\underbrace{\mathbf{1}_j + \cdots + \mathbf{1}_j}_{\ell} + \mathbf{1}_i + \mathbf{1}_{r-i} + \underbrace{\mathbf{1}_0 + \cdots + \mathbf{1}_0}_{k-\ell-2}).$$

Evaluating this gives

$$\alpha_j(i) = \begin{cases} (k-\ell-1)(j-1-r) + 2(k-\ell-2) + \frac{4}{r+1} & \text{if } i = 0 \\ \frac{4}{r+1} & \text{if } 1 \leqslant i \leqslant r \\ 2 & \text{if } r+1 \leqslant i \leqslant j-1 \\ (\ell-1)(j-1-r) + 2\ell & \text{if } i = j \\ 0 & \text{if } j+1 \leqslant i. \end{cases}$$

Recall that $j-1-r \geqslant 0$ and $\ell \geqslant 1$. Also, recall that $r = \lceil n/k \rceil \geqslant 2$. By definition, α_j satisfies (a) and therefore also (b). By looking at the evaluation above, it is easy to see that α_j satisfies (c)–(e). In order to check (f), note that $1 \leqslant \lfloor n/k \rfloor - 1 \leqslant \lfloor n/k \rfloor \leqslant \lceil n/k \rceil = r$. Hence, we have $\alpha_j(\lfloor n/k \rfloor) = \alpha_j(\lfloor n/k \rfloor - 1) = \alpha_j(\lceil n/k \rceil) = \frac{4}{r+1}$ and so (f) is satisfied as well.

Case 3: *$r = \lceil n/k \rceil$, $n \geqslant 2k$, and $\ell = k-1$.* In this case, we have $(k-1)j + r = n$, thus $kj = n + (j-r)$. Hence,

$$n + (j-r) = kj \geqslant k\left(\frac{2n}{k} - 1\right) \geqslant k\left(\frac{n}{k} + 2 - 1\right) = k\left(\frac{n}{k} + 1\right) = n + k.$$

So, $j - r \geqslant k \geqslant 3$ and therefore $j - 1 - r \geqslant 2$. Now, set

$$\alpha_j = \sum_{i=r+1}^{j-1} (\underbrace{\mathbf{1}_j + \cdots + \mathbf{1}_j}_{k-2} + \mathbf{1}_i + \mathbf{1}_{j+r-i})$$

Evaluating this gives

$$\alpha_j(i) = \begin{cases} 0 & \text{if } 0 \leqslant i \leqslant r \\ 2 & \text{if } r+1 \leqslant i \leqslant j-1 \\ (k-2)(j-1-r) & \text{if } i = j \\ 0 & \text{if } j+1 \leqslant i. \end{cases}$$

Recall that $j-1-r \geqslant 2$ and $k \geqslant 3$. Hence, $(k-2)(j-1-r) \geqslant 2$. Also recall that $r = \lceil n/k \rceil$. By definition, α_j satisfies (a) and therefore automatically also (b). By looking at the evaluation above, it is easy to see that α_j satisfies (c)–(e). In order to check (f), note that $\lfloor n/k \rfloor - 1 \leqslant \lfloor n/k \rfloor \leqslant \lceil n/k \rceil = r$. Hence, we have $\alpha_j(\lfloor n/k \rfloor) = \alpha_j(\lfloor n/k \rfloor - 1) = \alpha_j(\lceil n/k \rceil) = 0$ and so (f) is satisfied as well. This finishes the proof of Lemma 6.6. □

8.2. *Proof of Lemma* 6.7

Here, we will derive Lemma 6.7 from Lemma 6.6. However, we first need another lemma:

Lemma 8.1. *Let $n \geqslant 1$ and let ψ be an n-tame scaled distribution on $\{0, \ldots, n\}$ that is not identically zero. Let $j \in \{0, \ldots, n\}$ be chosen maximal with $\psi(j) > 0$. Then, $j + 1 \geqslant 2n/k$.*

Proof. By property (ii) in Definition 6.4, we have $\psi(i) \geqslant \psi(j+1-i)$ for $1 \leqslant i \leqslant (j+1)/2$. Hence,

$$\begin{aligned} i\psi(i) + (j+1-i)\psi(j+1-i) &= \frac{j+1}{2}(\psi(i) + \psi(j+1-i)) - \left(\frac{j+1}{2} - i\right)(\psi(i) - \psi(j+1-i)) \\ &\leqslant \frac{j+1}{2}(\psi(i) + \psi(j+1-i)) \end{aligned}$$

for $1 \leqslant i \leqslant (j+1)/2$. Thus, by symmetry between i and $j+1-i$, we actually obtain

$$i\psi(i) + (j+1-i)\psi(j+1-i) \leqslant \frac{j+1}{2}(\psi(i) + \psi(j+1-i)) \tag{8.2}$$

for $1 \leqslant i \leqslant j$. Note that by the choice of j, we have $\psi(i) = 0$ for all $i > j$. So using (8.2) and property (i) in Definition 6.4, we obtain

$$\begin{aligned}\frac{2n}{k}\sum_{i=0}^{n}\psi(i) &= 2\sum_{i=0}^{n} i\psi(i) = 2\sum_{i=1}^{j} i\psi(i) = \sum_{i=1}^{j} i\psi(i) + \sum_{i=1}^{j}(j+1-i)\psi(j+1-i)\\ &\leqslant \sum_{i=1}^{j}\frac{j+1}{2}(\psi(i)+\psi(j+1-i)) = \frac{j+1}{2}\sum_{i=1}^{j}\psi(i) + \frac{j+1}{2}\sum_{i=1}^{j}\psi(j+1-i)\\ &= (j+1)\sum_{i=1}^{j}\psi(i) \leqslant (j+1)\sum_{i=0}^{n}\psi(i).\end{aligned}$$

As $\sum_{i=0}^{n}\psi(i) > 0$, this yields $j+1 \geqslant 2n/k$ as desired. □

Now, we are ready for the proof of Lemma 6.7.

Proof of Lemma 6.7. Recall that $n \geqslant k$. Suppose there is a counterexample to Lemma 6.7. That means, one can find an n-tame scaled distribution ψ for which there exists no n-tame scaled distribution φ with the desired conditions. Clearly, such a ψ is not identically zero (because then we could just take φ to be identically zero as well). So for any such ψ, let $j \in \{0,\dots,n\}$ be chosen maximal with $\psi(j) > 0$. Among all possible scaled distributions ψ that contradict Lemma 6.7, let us choose one where this j is minimal. Thus, ψ is an n-tame scaled distribution on $\{0,\dots,n\}$, but there exists no φ with the desired properties. And $j \in \{0,\dots,n\}$ is maximal with $\psi(j) > 0$. This means $\psi(j+1) = \dots = \psi(n) = 0$, but $\psi(j) > 0$. Now, by the choice of ψ, for any n-tame scaled distribution ψ' with $\psi'(j) = \psi'(j+1) = \dots = \psi'(n) = 0$, we can find an n-tame scaled distribution φ' satisfying (6.1) such that $\psi' - \varphi'$ is a nonnegative linear combination of n-simple scaled distributions.

Note that by Lemma 8.1, we have $j+1 \geqslant 2n/k$ and in particular $j \geqslant (2n/k) - 1 \geqslant 1$. Thus, Lemma 6.6 gives a scaled distribution α_j satisfying conditions (a)–(f). For each $x \in \mathbb{R}_{\geqslant 0}$, define a map $\psi_x : \{0,\dots,n\} \to \mathbb{R}$ by

$$\psi_x = \psi - x\alpha_j.$$

Now, choose the maximum $x \in \mathbb{R}_{\geqslant 0}$ such that

$$\psi_x(j) = \psi(j) - x\alpha_j(j) \geqslant 0$$

and

$$\psi_x(0) \geqslant (k-1)\psi_x(n) + (k-2)\psi_x(n-1) + \dots + \psi_x(n-k+2).$$

Such a maximum x exists since $x = 0$ satisfies the conditions (for the second condition, see property (iii) of ψ), but due to $\alpha_j(j) > 0$ by (e), there is an upper bound for x. Note that for this maximum x, we have $\psi_x(j) = 0$ or $\psi_x(0) = (k-1)\psi_x(n) + (k-2)\psi_x(n-1) + \dots + \psi_x(n-k+2)$.

We claim that ψ_x is an n-tame scaled distribution. Recall that $\psi_x(j) \geqslant 0$. By (ii) for ψ and (c), we have $\psi(1) \geqslant \psi(2) \geqslant \dots \geqslant \psi(j)$ and $\alpha_j(1) \leqslant \alpha_j(2) \leqslant \dots \leqslant \alpha_j(j)$, hence

$$\psi_x(1) \geqslant \psi_x(2) \geqslant \dots \geqslant \psi_x(j) \geqslant 0.$$

From $\psi(j+1) = \dots = \psi(n) = 0$ and (d), we can deduce that

$$\psi_x(j+1) = \dots = \psi_x(n) = 0. \tag{8.3}$$

In particular, $\psi_x(i) \geqslant 0$ for $1 \leqslant i \leqslant n$. Now, recall that $n \geqslant k$ and

$$\psi_x(0) \geqslant (k-1)\psi_x(n) + (k-2)\psi_x(n-1) + \dots + \psi_x(n-k+2),$$

hence $\psi_x(0) \geqslant 0$ as well. So, we have established that ψ_x has nonnegative values, therefore it is a scaled distribution. We also have

$$\psi_x(1) \geqslant \psi_x(2) \geqslant \cdots \geqslant \psi_x(j) \geqslant 0 = \psi_x(j+1) = \cdots = \psi_x(n),$$

so ψ_x satisfies (ii). Furthermore, $\psi_x = \psi - x\alpha_j$ has mean n/k because both ψ and α_j have mean n/k (see (i) for ψ and (b)). Thus, ψ_x also satisfies (i). Note that (iii) is satisfied since we chose x such that

$$\psi_x(0) \geqslant (k-1)\psi_x(n) + (k-2)\psi_x(n-1) + \cdots + \psi_x(n-k+2).$$

It remains to check (iv). If $n < 2k$, then (iv) is vacuous, so assume $n \geqslant 2k$. Then, by (iv) for ψ and (f) we have $2\psi(\lfloor n/k \rfloor) \leqslant \psi(\lfloor n/k \rfloor - 1) + \psi(\lceil n/k \rceil)$ and $2\alpha_j(\lfloor n/k \rfloor) \geqslant \alpha_j(\lfloor n/k \rfloor - 1) + \alpha_j(\lceil n/k \rceil)$, hence

$$2\psi_x(\lfloor n/k \rfloor) \leqslant \psi_x(\lfloor n/k \rfloor - 1) + \psi_x(\lceil n/k \rceil).$$

So, ψ_x is indeed an n-tame scaled distribution. Recall that by the choice of x, we have $\psi_x(j) = 0$ or $\psi_x(0) = (k-1)\psi_x(n) + (k-2)\psi_x(n-1) + \cdots + \psi_x(n-k+2)$. Suppose the latter, then ψ_x would be an n-tame scaled distribution satisfying (6.1) and $\psi - \psi_x = x\alpha_j$ would be a nonnegative linear combination of n-tame scaled distributions (see (a)). This contradicts our assumption of ψ being a counterexample to Lemma 6.7.

Hence, we must have $\psi_x(j) = 0$. Thus, together with (8.3), we obtain $\psi_x(j) = \psi_x(j+1) = \cdots = \psi_x(n) = 0$. So, ψ_x is an n-tame scaled distribution with $\psi_x(j) = \psi_x(j+1) = \cdots = \psi_x(n) = 0$. We saw above that by the choice of ψ, this implies that for ψ_x, we can find an n-tame scaled distribution φ satisfying (6.1) such that $\psi_x - \varphi$ is a nonnegative linear combination of n-simple scaled distributions. But then, using (a), we obtain that $\psi - \varphi = (\psi_x - \varphi) + x\alpha_j$ is also a nonnegative linear combination of n-simple scaled distributions. This contradicts our choice of ψ. Hence, there cannot be any counterexamples to Lemma 6.7, so Lemma 6.7 is true. □

8.3. *Proof of Lemma* 6.8

The goal of this subsection is to prove Lemma 6.8. So, assume $n \geqslant 2k$ and let φ be an n-tame scaled distribution on $\{0, \ldots, n\}$ satisfying

$$\varphi(0) = (k-1)\varphi(n) + (k-2)\varphi(n-1) + \cdots + \varphi(n-k+2). \tag{8.4}$$

Let $s = \lfloor n/k \rfloor$ and $r = \frac{n}{k} - s$, so $0 \leqslant r < 1$ and $n = k(r+s)$. Also note that $s \geqslant 2$.

Set

$$\begin{aligned}\lambda(0) &= \varphi(0) - \sum_{i=n-(k-1)+1}^{n} (i - n + (k-1))\varphi(i)\\ &= \varphi(0) - \varphi(n-k+2) - 2\varphi(n-k+3) - \cdots - (k-1)\varphi(n)\end{aligned}$$

and note that by (8.4), we have $\lambda(0) = 0$. Furthermore, for each $1 \leqslant \ell \leqslant s-1$, set

$$\begin{aligned}\lambda(\ell) = \varphi(\ell) - &\sum_{i=n-(\ell+1)(k-1)+1}^{n-\ell(k-1)} (i - n + (\ell+1)(k-1))\varphi(i)\\ &- \sum_{i=n-\ell(k-1)+1}^{n-(\ell-1)(k-1)} (n - (\ell-1)(k-1) - i)\varphi(i).\end{aligned}$$

For each index i occurring in the sums, we have $i \geqslant n-(\ell+1)(k-1)+1 \geqslant n-s(k-1)+1 \geqslant s+1$. Note that we can rewrite $\lambda(\ell)$ for $1 \leqslant \ell \leqslant s-1$ as

$$\lambda(\ell)=\varphi(\ell)-\sum_{j=1}^{k-1} j\cdot\varphi(n-(\ell+1)(k-1)+j)-\sum_{j=1}^{k-1}(k-1-j)\cdot\varphi(n-\ell(k-1)+j).$$

Recalling that $\varphi(1)\geqslant\cdots\geqslant\varphi(n)$ by property (ii) in Definition 6.4, we obtain

$$\lambda(1)\geqslant\cdots\geqslant\lambda(s-1).$$

The claim of Lemma 6.8 is equivalent to $\lambda(1)\geqslant 0$. Let us assume for contradiction that the claim is false, that is, $\lambda(1)<0$. Then, $0>\lambda(1)\geqslant\cdots\geqslant\lambda(s-1)$. Let us consider the term

$$r\varphi(s)+\sum_{\ell=0}^{s-1}\left(\frac{n}{k}-\ell\right)\lambda(\ell). \tag{8.5}$$

We now plug in the definition of $\lambda(\ell)$ for $\ell=0,\dots,s-1$. For every $1\leqslant\ell\leqslant s-1$ and every i with $n-\ell(k-1)+1\leqslant i\leqslant n-(\ell-1)(k-1)$, the coefficient of $\varphi(i)$ in $\lambda(\ell)$ is given by $-(n-(\ell-1)(k-1)-i)$ and its coefficient in $\lambda(\ell-1)$ is $-(i-n+\ell(k-1))$. Hence, the total coefficient of $\varphi(i)$ in (8.5) is

$$\begin{aligned}-&\left(\frac{n}{k}-\ell\right)(n-(\ell-1)(k-1)-i)-\left(\frac{n}{k}-\ell+1\right)(i-n+\ell(k-1))\\&=-\left(\frac{n}{k}-\ell\right)(k-1)-(i-n+\ell(k-1))=-\frac{n}{k}(k-1)+n-i=\frac{n}{k}-i=-\left(i-\frac{n}{k}\right).\end{aligned}$$

Furthermore, for all i with $n-s(k-1)+1\leqslant i\leqslant n-(s-1)(k-1)$, the coefficient of $\varphi(i)$ in $\lambda(s-1)$ is $-(i-n+s(k-1))$, and so the total coefficient of $\varphi(i)$ in (8.5) is

$$-\left(\frac{n}{k}-s+1\right)(i-n+s(k-1))=-(r+1)(i-n+s(k-1)).$$

Hence, when plugging in the definition of $\lambda(\ell)$ for $\ell=0,\dots,s-1$ into the term (8.5), we obtain

$$\begin{aligned}r\varphi(s)+\sum_{\ell=0}^{s-1}\left(\frac{n}{k}-\ell\right)\lambda(\ell)&=\left(\frac{n}{k}-s\right)\varphi(s)+\sum_{\ell=0}^{s-1}\left(\frac{n}{k}-\ell\right)\lambda(\ell)\\&=\sum_{\ell=0}^{s}\left(\frac{n}{k}-\ell\right)\varphi(\ell)-\sum_{i=n-s(k-1)+1}^{n-(s-1)(k-1)}(r+1)(i-n+s(k-1))\varphi(i)\\&\quad-\sum_{i=n-(s-1)(k-1)+1}^{n}\left(i-\frac{n}{k}\right)\varphi(i)\end{aligned}$$

Since φ has mean n/k by property (i) in Definition 6.4, we have $\sum_{i=0}^{n} i\varphi(i)=\frac{n}{k}\sum_{i=0}^{n}\varphi(i)$ and therefore

$$\sum_{\ell=0}^{s}\left(\frac{n}{k}-\ell\right)\varphi(\ell)=\sum_{i=s+1}^{n}\left(i-\frac{n}{k}\right)\varphi(i).$$

Thus, recalling $n-s(k-1)+1 \geqslant s+1$, we obtain

$$\begin{aligned}
&r\varphi(s)+\sum_{\ell=0}^{s-1}\left(\frac{n}{k}-\ell\right)\lambda(\ell)\\
&\quad=\sum_{i=s+1}^{n}\Big(i-\frac{n}{k}\Big)\,\varphi(i)-\sum_{i=n-s(k-1)+1}^{n-(s-1)(k-1)}(r+1)(i-n+s(k-1))\varphi(i)-\sum_{i=n-(s-1)(k-1)+1}^{n}\Big(i-\frac{n}{k}\Big)\,\varphi(i)\\
&\quad=\sum_{i=s+1}^{n-s(k-1)}\Big(i-\frac{n}{k}\Big)\,\varphi(i)+\sum_{i=n-s(k-1)+1}^{n-(s-1)(k-1)}\Big(i-\frac{n}{k}-(r+1)(i-n+s(k-1))\Big)\,\varphi(i).
\end{aligned}$$

Note that

$$\begin{aligned}
i-\frac{n}{k}-(r+1)(i-n+s(k-1)) &= i-\frac{n}{k}-i+n-s(k-1)-r(i-n+s(k-1))\\
&=\frac{n}{k}(k-1)-s(k-1)-r(i-n+s(k-1))\\
&=r(k-1)-r(i-n+s(k-1))=r(n-(s-1)(k-1)-i).
\end{aligned}$$

Hence,

$$r\varphi(s)+\sum_{\ell=0}^{s-1}\left(\frac{n}{k}-\ell\right)\lambda(\ell)=\sum_{i=s+1}^{n-s(k-1)}\Big(i-\frac{n}{k}\Big)\,\varphi(i)+\sum_{i=n-s(k-1)+1}^{n-(s-1)(k-1)}r(n-(s-1)(k-1)-i)\varphi(i).$$

Recall that $\lambda(0)=0$ and $0>\lambda(1)\geqslant\cdots\geqslant\lambda(s-1)$. Noting that the coefficient of each $\lambda(\ell)$ on the left-hand side of the last equation is strictly positive, this implies (recalling $s \geqslant 2$)

$$r\varphi(s)>\sum_{i=s+1}^{n-s(k-1)}\Big(i-\frac{n}{k}\Big)\,\varphi(i)+\sum_{i=n-s(k-1)+1}^{n-(s-1)(k-1)}r(n-(s-1)(k-1)-i)\varphi(i). \tag{8.6}$$

As $s+1>\frac{n}{k}$, all terms on the right-hand side are nonnegative. Hence, the left-hand side must be positive. In particular, we must have $r>0$. So, $s<\frac{n}{k}$ and therefore $n-s(k-1)\geqslant s+1$. Furthermore, we obtain $\lceil n/k\rceil=\lfloor n/k\rfloor+1=s+1$. Thus, property (iv) in Definition 6.4 gives $2\varphi(s)\leqslant\varphi(s-1)+\varphi(s+1)$. Hence,

$$r\varphi(s)=r^2\varphi(s)+\frac{1}{2}r(1-r)\cdot 2\varphi(s)\leqslant r^2\varphi(s)+\frac{1}{2}r(1-r)\cdot(\varphi(s-1)+\varphi(s+1)).$$

Together with (8.6), this gives

$$\begin{aligned}
&r^2\varphi(s)+\frac{1}{2}r(1-r)\varphi(s-1)+\frac{1}{2}r(1-r)\varphi(s+1)\\
&\quad>\sum_{i=s+1}^{n-s(k-1)}\Big(i-\frac{n}{k}\Big)\,\varphi(i)+\sum_{i=n-s(k-1)+1}^{n-(s-1)(k-1)}r(n-(s-1)(k-1)-i)\varphi(i).
\end{aligned}$$

Note that the coefficient of $\varphi(s+1)$ on the right-hand side is $s+1-\frac{n}{k}=1-r$. So, subtracting $\frac{1}{2}r(1-r)\varphi(s+1)$ from both sides gives

$$r^2\varphi(s)+\frac{1}{2}r(1-r)\varphi(s-1)>(1-r)\left(1-\frac{1}{2}r\right)\varphi(s+1)+\sum_{i=s+2}^{n-s(k-1)}\left(i-\frac{n}{k}\right)\varphi(i)$$
$$+\sum_{j=1}^{k-1}r(k-1-j)\varphi(n-s(k-1)+j).$$

Using again $\varphi(1)\geqslant\cdots\geqslant\varphi(n)$ and $s\geqslant 2$ as well as $n-s(k-1)\geqslant s+1$, this yields

$$r^2\varphi(1)+\frac{1}{2}(r-r^2)\varphi(1)>\frac{1}{2}(2-3r+r^2)\varphi(n-2(k-1))+\sum_{i=s+2}^{n-s(k-1)}\left(i-\frac{n}{k}\right)\varphi(n-2(k-1))$$
$$+\sum_{j=1}^{k-2}r(k-1-j)\varphi(n-2(k-1)+j). \tag{8.7}$$

Note that

$$\sum_{i=s+2}^{n-s(k-1)}\left(i-\frac{n}{k}\right)=\sum_{i=s+2}^{n-s(k-1)}(i-s-1+(1-r))$$
$$=(n-s(k-1)-s-1)(1-r)+\sum_{j=1}^{n-s(k-1)-s-1}j.$$

Using $n-s(k-1)-s-1=n-sk-1=kr-1$, this yields

$$\sum_{i=s+2}^{n-s(k-1)}\left(i-\frac{n}{k}\right)=(kr-1)(1-r)+\frac{(kr-1)kr}{2}=(k+1)r-kr^2-1+\frac{k^2r^2-kr}{2}$$
$$=\frac{1}{2}(-2+(k+2)r+(k^2-2k)r^2).$$

Plugging this into (8.7) and simplifying, we obtain

$$\frac{r}{2}(1+r)\varphi(1)>\frac{1}{2}((k-1)r+(k-1)^2r^2)\varphi(n-2(k-1))+\sum_{j=1}^{k-2}r(k-1-j)\varphi(n-2(k-1)+j).$$

As $\varphi(n-2(k-1)+1)\geqslant\cdots\geqslant\varphi(n-2(k-1)+(k-2))$ and $r(k-2)\geqslant\cdots\geqslant r\cdot 1$, we have by Chebyshev's sum inequality (or alternatively by the rearrangement inequality)

$$\sum_{j=1}^{k-2}r(k-1-j)\varphi(n-2(k-1)+j)\geqslant\sum_{j=1}^{k-2}\frac{r(k-2)+\cdots+r\cdot 1}{k-2}\varphi(n-2(k-1)+j)$$
$$=\sum_{j=1}^{k-2}\frac{r(k-1)}{2}\varphi(n-2(k-1)+j).$$

Thus, using $\varphi(1) \geqslant \cdots \geqslant \varphi(n)$ again,

$$\frac{r}{2}(1+r)\varphi(1) > \frac{r}{2}((k-1)+(k-1)^2r)\varphi(n-2(k-1)) + \sum_{j=1}^{k-2}\frac{r}{2}(k-1)\varphi(n-2(k-1)+j)$$

$$\geqslant \frac{r}{2}(1+r)(k-1)\varphi(n-(k-1)) + \frac{r^2}{2}(k-1)(k-2)\varphi(n-2(k-1))$$

$$+\sum_{j=1}^{k-2}\frac{r}{2}(k-1)\varphi(n-2(k-1)+j)$$

$$\geqslant \frac{r}{2}(1+r)(k-1)\varphi(n-(k-1)) + \sum_{j=1}^{k-2}\left(\frac{r}{2}+\frac{r^2}{2}\right)(k-1)\varphi(n-2(k-1)+j).$$

Dividing by $\frac{r}{2}(1+r) = \frac{r}{2}+\frac{r^2}{2}$ yields

$$\varphi(1) > (k-1)\varphi(n-(k-1)) + \sum_{j=1}^{k-2}(k-1)\varphi(n-2(k-1)+j)$$

$$\geqslant (k-1)\varphi(n-(k-1)) + \sum_{j=1}^{k-2}[j\cdot\varphi(n-2(k-1)+j) + (k-1-j)\cdot\varphi(n-(k-1)+j)].$$

In other words,

$$\varphi(1) > \varphi(n-1) + 2\varphi(n-2) + \cdots + (k-2)\varphi(n-k+2)$$
$$+(k-1)\varphi(n-k+1) + (k-2)\varphi(n-k) + \cdots + \varphi(n-2k+3),$$

so the claim of Lemma 6.8 is true. This is a contradiction to our assumption (recall that we assumed that the lemma is false), which finishes the proof of the lemma.

9. *Upper bound*

Here, we give a proof of Theorem 1.2. The proof is very similar to that of Theorem 4 in [**26**], which in turn was inspired by the proof of Theorem 4.14 in [**8**]. We repeat these arguments here for the reader's convenience.

Let $m = p^\ell$ for a prime number p and an integer $\ell \geqslant 1$. For every integer $0 \leqslant a \leqslant m-1$, we have

$$\binom{z}{a} \equiv \binom{z'}{a} \pmod{p}$$

if z and z' are nonnegative integers with $z \equiv z' \pmod{m}$. This can be derived from Lucas' theorem. Hence, there is a well-defined map $\mathbb{Z}_m \to \mathbb{F}_p$ given by $z \mapsto \binom{z}{a}$.

Lemma 9.1 [**26**, Lemma 9]. *Let p be a prime and $m = p^\ell$ be a prime power, and let $z_1, \dots, z_k \in \mathbb{Z}_m$. Then, over $\mathbb{F}_p$, we have the identity*

$$\sum_{\substack{a_1,\dots,a_k\in\{0,\dots,m-1\}\\ a_1+\cdots+a_k\leqslant m-1}} (-1)^{a_1+\cdots+a_k}\binom{z_1}{a_1}\cdots\binom{z_k}{a_k} = \begin{cases} 1 & \text{if } z_1+\cdots+z_k = 0 \text{ in } \mathbb{Z}_m\\ 0 & \text{otherwise.}\end{cases}$$

Proof. Note that for every integer $0 \leqslant z \leqslant m-1$, we have

$$\sum_{0 \leqslant a \leqslant m-1} (-1)^a \binom{z}{a} = (1-1)^z = \begin{cases} 1 & \text{if } z = 0 \\ 0 & \text{if } 1 \leqslant z \leqslant m-1. \end{cases}$$

Hence, for every $z \in \mathbb{Z}_m$ we have

$$\sum_{0 \leqslant a \leqslant m-1} (-1)^a \binom{z}{a} = \begin{cases} 1 & \text{if } z = 0 \text{ in } \mathbb{Z}_m \\ 0 & \text{if } z \neq 0 \text{ in } \mathbb{Z}_m \end{cases}$$

over $\mathbb{F}_p$. Furthermore, note that for $0 \leqslant a \leqslant m-1$ and for nonnegative integers $z_1, \dots, z_k$, we have

$$\binom{z_1 + \dots + z_k}{a} = \sum_{\substack{a_1,\dots,a_k \in \{0,\dots,m-1\} \\ a_1 + \dots + a_k = a}} \binom{z_1}{a_1} \dots \binom{z_k}{a_k},$$

hence the same identity holds over $\mathbb{F}_p$ for $z_1, \dots, z_k \in \mathbb{Z}_m$. Now, for all $z_1, \dots, z_k \in \mathbb{Z}_m$, we obtain

$$\sum_{\substack{a_1,\dots,a_k \in \{0,\dots,m-1\} \\ a_1 + \dots + a_k \leqslant m-1}} (-1)^{a_1+\dots+a_k} \binom{z_1}{a_1} \dots \binom{z_k}{a_k} = \sum_{0 \leqslant a \leqslant m-1} (-1)^a \binom{z_1 + \dots + z_k}{a}$$
$$= \begin{cases} 1 & \text{if } z_1 + \dots + z_k = 0 \text{ in } \mathbb{Z}_m \\ 0 & \text{otherwise} \end{cases}$$

over $\mathbb{F}_p$. □

Let us now prove Theorem 1.2 using Tao's slice rank method [35].

Proof of Theorem 1.2. Let $(x_{1,j}, x_{2,j}, \dots, x_{k,j})_{j=1}^L$ be a k-colored sum-free set in $\mathbb{Z}_m^n$. We need to prove that $L \leqslant (\Gamma_{m,k})^n$. We will first prove that $L \leqslant k \cdot (\Gamma_{m,k})^n$, and the additional factor k will then be removed using a power trick.

Let us define a tensor $G : \{1, \dots, L\}^k \to \mathbb{F}_p$ by setting

$$G(j_1, \dots, j_k) = \prod_{i=1}^n \left(\sum_{\substack{a_1,\dots,a_k \in \{0,\dots,m-1\} \\ a_1 + \dots + a_k \leqslant m-1}} (-1)^{a_1+\dots+a_k} \binom{x_{1,j_1}^{(i)}}{a_1} \dots \binom{x_{k,j_k}^{(i)}}{a_k} \right) \tag{9.1}$$

for all $j_1, \dots, j_k \in \{1, \dots, L\}$. It follows from Lemma 9.1 that for each $i = 1, \dots, n$ the sum on the right-hand side is 1 if and only if $x_{1,j_1}^{(i)} + \dots + x_{k,j_k}^{(i)} = 0$ and zero otherwise. Thus, we obtain

$$G(j_1, \dots, j_k) = \begin{cases} 1 & \text{if } x_{1,j_1} + \dots + x_{k,j_k} = 0 \\ 0 & \text{otherwise} \end{cases}$$

for all $j_1, \dots, j_k \in \{1, \dots, L\}$. Hence, G is a diagonal tensor and by Tao's slice rank Lemma [35, Lemma 1] the tensor G has slice rank L.

On the other hand, by multiplying (9.1) out, we can write $G(j_1, \dots, j_k)$ as a linear combination of terms of the form

$$\left(\binom{x_{1,j_1}^{(1)}}{a_{1,1}} \binom{x_{1,j_1}^{(2)}}{a_{1,2}} \dots \binom{x_{1,j_1}^{(n)}}{a_{1,n}} \right) \dots \left(\binom{x_{k,j_k}^{(1)}}{a_{k,1}} \binom{x_{k,j_k}^{(2)}}{a_{k,2}} \dots \binom{x_{k,j_k}^{(n)}}{a_{k,n}} \right)$$

with $a_{1,i}+\dots+a_{k,i}\leqslant m-1$ for each $i=1,\dots,n$. Thus, $\sum_{s=1}^k(a_{s,1}+\dots+a_{s,n})\leqslant n(m-1)$, so for each of these terms in the linear combination we can choose some $s\in\{1,\dots,k\}$ with $a_{s,1}+\dots+a_{s,n}\leqslant n(m-1)/k$. Let us now sort the terms into groups depending on the chosen index $s\in\{1,\dots,k\}$ and on the n-tuple $(a_{s,1},\dots,a_{s,n})$. Then, each group gives a term of the form

$$\left(\binom{x^{(1)}_{s,j_s}}{a_{s,1}}\binom{x^{(2)}_{s,j_s}}{a_{s,2}}\dots\binom{x^{(n)}_{s,j_s}}{a_{s,n}}\right)\cdot G'(j_1,\dots,j_{s-1},j_{s+1},\dots,j_n)$$

for some function G' depending only on $j_1,\dots,j_{s-1},j_{s+1},\dots,j_n$ and not on j_s. In other words, we obtain a slice rank decomposition of G with one slice for each group given by $s\in\{1,\dots,k\}$ and an n-tuple $(a_{s,1},\dots,a_{s,n})$. Since G has slice rank L, the number of groups must be at least L, so

$$L\leqslant k\cdot|\{(a_1,\dots,a_n)\in\{0,\dots,m-1\}^n\mid a_1+\dots+a_n\leqslant n(m-1)/k\}|.$$

The following lemma gives an upper bound for the quantity on the right-hand side.

Lemma 9.2. $|\{(a_1,\dots,a_n)\in\{0,\dots,m-1\}^n\mid a_1+\dots+a_n\leqslant n(m-1)/k\}|\leqslant(\Gamma_{m,k})^n$.

We postpone the proof of this lemma for a moment, in order to first finish the proof of 1.2. Applying Lemma 9.2, we obtain $L\leqslant k\cdot(\Gamma_{m,k})^n$ for every k-colored sum-free set $(x_{1,j},x_{2,j},\dots,x_{k,j})_{j=1}^L$ in $\mathbb{Z}_m^n$ (for all n). Note that if $(x_{1,j},x_{2,j},\dots,x_{k,j})_{j=1}^L$ is a k-colored sum-free set in $\mathbb{Z}_m^n$, then for every integer $\ell\geqslant1$, we can construct a k-colored sum-free set of size L^ℓ in $\mathbb{Z}_m^{n\ell}=\mathbb{Z}_m^n\times\dots\times\mathbb{Z}_m^n$ by taking the collection of k-tuples

$$((x_{1,j_1},x_{1,j_2},\dots,x_{1,j_\ell}),\dots,(x_{k,j_1},x_{k,j_2},\dots,x_{k,j_\ell}))_{(j_1,\dots,j_\ell)\in\{1,\dots,L\}^\ell}.$$

Thus, $L^\ell\leqslant k\cdot(\Gamma_{m,k})^{n\ell}$, and we can conclude $L\leqslant(\Gamma_{m,k})^n$ by taking $\ell\to\infty$. □

Lemma 9.2 has a standard proof, it was given for example in [**8**, Proposition 4.12], see also [**26**, Lemma 5]. For the reader's convenience, we repeat the proof here:

Proof of Lemma 9.2. Let $Z_1,\dots,Z_n$ be independent random variables, uniformly distributed on $\{0,\dots,m-1\}$. Then, the desired number of n-tuples $(a_1,\dots,a_n)\in\{0,\dots,m-1\}^n$ with $a_1+\dots+a_n\leqslant n(m-1)/k$ equals $m^n\,\mathbb{P}(Z_1+\dots+Z_n\leqslant n(m-1)/k)$. So, we need to prove $\mathbb{P}(Z_1+\dots+Z_n\leqslant n(m-1)/k)\leqslant m^{-n}(\Gamma_{m,k})^n$.

For every $0<\gamma<1$, we have by Markov's inequality

$$\begin{aligned}\mathbb{P}\left(Z_1+\dots+Z_n\leqslant n(m-1)/k\right)&=\mathbb{P}\left(\gamma^{Z_1+\dots+Z_n}\geqslant\gamma^{n(m-1)/k}\right)\\&\leqslant\gamma^{-n(m-1)/k}\,\mathbb{E}\left(\gamma^{Z_1+\dots+Z_n}\right)\\&=\gamma^{-n(m-1)/k}\left(\mathbb{E}\left(\gamma^{Z_1}\right)\right)^n\\&=\gamma^{-n(m-1)/k}\left(\frac{1+\gamma+\dots+\gamma^{m-1}}{m}\right)^n\\&=m^{-n}\left(\frac{1+\gamma+\dots+\gamma^{m-1}}{\gamma^{(m-1)/k}}\right)^n.\end{aligned}$$

Taking $\gamma=\gamma_{m,k}$, this gives $\mathbb{P}(Z_1+\dots+Z_n\leqslant n(m-1)/k)\leqslant m^{-n}(\Gamma_{m,k})^n$, as desired. □

Acknowledgements. We would like to thank Jacob Fox and Terence Tao for helpful discussions.

References

1. N. Alon, 'Testing subgraphs in large graphs', 42nd IEEE Symposium on Foundations of Computer Science (IEEE Computer Society, Los Alamitos, CA, 2001) 434–441.

2. N. Alon, A. Shpilka and C. Umans, 'On sunflowers and matrix multiplication', *Comput. Complexity* 22 (2013) 219–243.

3. N. Alon and J. Spencer, *The probabilistic method*, 4th edn (Wiley, Hoboken, NJ, 2016).

4. F. A. Behrend, 'On sets of integers which contain no three terms in arithmetical progression', *Proc. Natl. Acad. Sci. USA* 32 (1946) 331–332.

5. M. Bennett, 'Bounds on sizes of caps in $AG(n, q)$ via the croot-lev-pach polynomial method', Preprint, 2018, arXiv:1806.05303.

6. A. Bhattacharyya, E. Grigorescu, P. Raghavendra and A. Shapira, 'Testing odd-cycle-freeness in Boolean functions', *Combin. Probab. Comput.* 21 (2012) 835–855.

7. A. Bhattacharyya and N. Xie, 'Lower bounds for testing triangle-freeness in Boolean functions', *Comput. Complexity* 24 (2015) 65–101. (A preliminary version appeared in SODA 2010, pp. 87–98.)

8. J. Blasiak, T. Church, H. Cohn, J. A. Grochow, E. Naslund, W. F. Sawin and C. Umans, 'On cap sets and the group-theoretic approach to matrix multiplication', *Discrete Analysis* 2017:3 (2017) 27pp.

9. D. Coppersmith and S. Winograd, 'Matrix multiplication via arithmetic progressions', *J. Symbolic Comput.* 9 (1990) 251–280.

10. E. Croot, V. F. Lev and P. P. Pach, 'Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small', *Ann. of Math.* (2) 185 (2017) 331–337.

11. Z. Dvir and S. Moran, 'A Sauer-Shelah-Perles lemma for sumsets', *Electron. J. Combin.* 25 (2018) Paper 4.38, 7pp.

12. J. S. Ellenberg, 'Sumsets as unions of sumsets of subsets', *Discrete Analysis* 2017:14 (2017) 5pp.

13. J. S. Ellenberg and D. Gijswijt, 'On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression', *Ann. of Math.* (2) 185 (2017) 339–343.

14. J. Fox, 'A new proof of the graph removal lemma', *Ann. of Math.* (2) 174 (2011) 561–579.

15. J. Fox and L. M. Lovász, 'A tight bound for Green's arithmetic triangle removal lemma in vector spaces', *Adv. Math.* 321 (2017) 287–297.

16. J. Fox, L. M. Lovász and L. Sauermann, 'A polynomial bound for the arithmetic k-cycle removal lemma in vector spaces', *J. Combin. Theory Ser. A* 160 (2018) 186–201.

17. J. Fox and L. Sauermann, 'Erdős-Ginzburg-Ziv constants by avoiding three-term arithmetic progressions', *Electron. J. Combin.* 25 (2018) Paper 2.14, 9pp.

18. H. Fu and R. Kleinberg, 'Improved lower bounds for testing triangle-freeness in Boolean functions via fast matrix multiplication', *Approximation, randomization, and combinatorial optimization*, Leibniz International Proceedings in Informatics 28 (eds K. Jansen, J. D. P. Rolim, N. R. Devanur and C. Moore; Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Wadern, 2014) 669–676.

19. G. Ge and C. Shangguan, 'Rank counting and maximum subsets of $\mathbb{F}_q^n$ containing no right angles', Preprint, 2016, arXiv:1612.08255.

20. B. Green, 'A Szemerédi-type regularity lemma in abelian groups, with applications', *Geom. Funct. Anal.* 15 (2005) 340–376.

21. B. Green, 'Sárközy's theorem in function fields', *Q. J. Math.* 68 (2017) 237–242.

22. I. Haviv and N. Xie, 'Sunflowers and testing triangle-freeness of functions', ITCS'15—Proceedings of the 6th Innovations in Theoretical Computer Science (Association for Computing Machinery, New York, 2015) 357–366.

23. G. Hegedűs, 'The Erdős-Ginzburg-Ziv constant and progression-free subsets', *J. Number Theory* 186 (2018) 238–247.

24. R. Kleinberg, W. F. Sawin and D. E. Speyer, 'The growth rate of tri-colored sum-free sets', *Discrete Analysis* 2018:12 (2018) 10pp.

25. S. Lovett, 'The analytic rank of tensors and its applications', Preprint, 2018, arXiv:1806.09179.

26. E. Naslund, 'Exponential bounds for the Erdős-Ginzburg-Ziv constant', Preprint, 2017, arXiv:1701.04942.

27. E. Naslund, 'The partition rank of a tensor and k-right corners in $\mathbb{F}_q^n$', Preprint, 2017, arXiv:1701.04475.

28. E. Naslund and W. Sawin, 'Upper bounds for sunflower-free sets', *Forum Math. Sigma* 5 (2017) e15, 10pp.

29. S. Norin, 'A distribution on triples with maximum entropy marginal', Preprint, 2016, arXiv:1608.00243.

30. L. Pebody, 'Proof of a conjecture of Kleinberg–Sawin–Speyer', Preprint, 2016, arXiv:1608.05740v1.

31. L. Pebody, 'Proof of a conjecture of Kleinberg–Sawin–Speyer', *Discrete Analysis* 2018:13 (2018) 7pp.

32. F. Petrov, 'Many zero divisors in a group ring imply bounds on progression–free subsets', Preprint, 2016, arXiv:1606.03256.

33. F. Petrov and C. Pohoata, 'Improved Bounds for Progression-Free Sets in C_8^n', Preprint, 2018, arXiv:1805.05549.

34. W. Sawin, 'Bounds for matchings in nonabelian groups', *Electron. J. Combin.* 25 (2018) Paper 4.23, 21pp.

35. T. Tao, 'A symmetric formulation of the Croot-Lev-Pach-Ellenberg-Gijswijt capset bound', 2016 blog post, https://terrytao.wordpress.com/2016/05/18/a-symmetric-formulation-of-the-croot-lev-pach-ellenberg-gijswijt-capset-bound.

László Miklós Lovász
Department of Mathematics
Massachusetts Institute of Technology
Headquarters Office
Simons Building (Building 2)
Room 106, 77 Massachusetts
Avenue Cambridge, MA 02139-4307
USA

lmlovasz@mit.edu

Lisa Sauermann
Department of Mathematics
Stanford University
450 Serra Mall, Building 380
Stanford, CA 94305-2125
USA

lsauerma@stanford.edu

The *Proceedings of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.