Discrete Mathematics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Majorization and Rényi entropy inequalities via Sperner theory[☆]

Mokshay Madiman [a], Liyao Wang [b], Jae Oh Woo [c,*]

[a] Department of Mathematical Sciences, University of Delaware, Newark DE 19716, United States
[b] J.P. Morgan Chase, NY, United States
[c] Department of Mathematics, and Department of Electrical and Computer Engineering, The University of Texas at Austin, United States

ARTICLE INFO

Article history:
Received 16 April 2018
Accepted 1 March 2019
Available online xxxx

Keywords:
Posets
Strong Sperner property
Majorization
Entropy power inequality
Sumset inequality

ABSTRACT

A natural link between the notions of majorization and strongly Sperner posets is elucidated. It is then used to obtain a variety of consequences, including new Rényi entropy inequalities for sums of independent, integer-valued random variables.

1. Introduction

It was observed by Erdős [12] in 1945 that the lemma of Littlewood and Offord [32] on small ball probabilities of weighted sums of Bernoulli random variables actually follows from Sperner's theorem [58] on the maximal size of antichains in the Boolean lattice. Subsequently Stanley [60] and Proctor [51] used similar ideas to attack more difficult problems; a very nice review of the key ideas can be found in [28]. The goal of this paper is to further develop the core idea of relating properties of posets to the "distributional spread" of weighted sums of independent, integer-valued random variables. We do this in two steps. First, we elucidate a natural link, which does not seem to have been explicitly observed in the literature, between the strong Sperner property of posets and its behavior for product posets on the one hand, and majorization inequalities on the other. Second, we follow a classical approach, similar to that used in our earlier papers [46,63,64], to demonstrate new Rényi entropy inequalities for sums of independent random variables using the majorization inequalities. The entropy inequalities are of interest in information theory and probability, and were our original motivation for this work—they are discussed at length in Section 2.

In order to state our main results, we need to develop some terminology. For a non-negative function $f : \mathbb{Z} \to \mathbb{R}_+$ over the integers, the support $\mathrm{Supp}(f)$ is defined by $\{x \in \mathbb{Z} : f(x) > 0\}$. We identify sets with their indicator functions; thus, for example, $f = 0.4\{0, 3\} + 0.2\{2\}$ means $f(0) = f(3) = 0.4$, $f(2) = 0.2$, and $f(x) = 0$ for $x \in \mathbb{Z} \setminus \{0, 2, 3\}$.

☆ Some portions of this paper were presented by the authors at the IEEE International Symposia on Information Theory in 2014 and 2015 (Wang et al., 2014; Woo and Madiman, 2015).

* Corresponding author.
E-mail addresses: madiman@udel.edu (M. Madiman), liyao.wang@yale.edu (L. Wang), jaeoh.woo@aya.yale.edu (J.O. Woo).

https://doi.org/10.1016/j.disc.2019.03.002

Definition 1.1. Suppose $f : \mathbb{Z} \to \mathbb{R}_+$ is finitely supported, with $|\mathrm{Supp}(f)| = n + 1$. Then we may write $\mathrm{Supp}(f) = \{x_0, \ldots, x_n\}$ with $x_0 < \cdots < x_n$, and we may represent f in the form

$$f = \sum_{r=0}^{n} a_r \{x_r\},$$

where $a_i > 0$ for each $i \in \{0, \ldots, n\}$. Given the non-negative function f, we define $f^{\#}$ by

$$f^{\#} := \sum_{r=0}^{n} a_r \{r\}.$$

Thus, $f^{\#}$ is supported on $\{0, \ldots, n\}$ and it takes the same functional values as f. If we consider the graph, we may think of $f^{\#}$ as a "squeezed rearrangement" of f, where we preserve the order of the function values but eliminate gaps in the support.

Definition 1.2. We say f is *#-log-concave* if $f^{\#}$ is log-concave, i.e., $f^{\#}(i)^2 \geq f^{\#}(i-1)f^{\#}(i+1)$ for any $i \in \mathbb{Z}$. We say that a random variable X taking values in the integers is *#-log-concave* if its probability mass function is #-log-concave. Given a random variable X with probability mass function f, we write $X^{\#}$ for a random variable with probability mass function $f^{\#}$.

In the terminology of Definition 1.1, since $a_r = 0$ for $r \in \mathbb{Z} \setminus \{0, \ldots, n\}$, f is *#-log-concave* if and only if $a_r^2 \geq a_{r-1}a_{r+1}$.

We also need the classical notion of majorization. We use $f^{[i]}$ to denote the ith largest value of f, allowing for the possibility of multiple ties. For example, $f^{[i]} = f^{[i+1]}$ when the ith largest value appears at two different locations.

Definition 1.3. Consider two finitely supported functions f and g from $\mathbb{Z}$ to $\mathbb{R}_+$, and assume $|\mathrm{Supp}(f)| = |\mathrm{Supp}(g)| = n+1$. We say f is *majorized* by g (and write $f \prec g$) if

$$\sum_{i=1}^{k} f^{[i]} \leq \sum_{i=1}^{k} g^{[i]} \quad \text{for all } k = 1, \ldots, n, \tag{1}$$

and

$$\sum_{i=1}^{n+1} f^{[i]} = \sum_{i=1}^{n+1} g^{[i]}. \tag{2}$$

For random variables X and Y with probability mass functions f and g respectively, we write $X \prec Y$ if $f \prec g$.

Our first main theorem is a majorization inequality for convolutions that holds under a log-concavity condition. Recall that, given independent random variables X, Y with probability mass functions f, g, the sum $X + Y$ has the probability mass function $f \star g$, where $\star$ denotes convolution, i.e., $f \star g(k) = \sum_{i \in \mathbb{Z}} f(i)g(k-i)$ for each $k \in \mathbb{Z}$.

Theorem 1.4. *Let N be a finite number. If $X_1, \ldots, X_N$ are independent and #-log-concave over $\mathbb{Z}$, then*

$$X_1 + \cdots + X_N \prec X_1^{\#} + \cdots + X_N^{\#}. \tag{3}$$

The proof of Theorem 1.4 is based on the strong Sperner property of the product of weighted chain posets—Section 3 summarizes the necessary background on poset theory, and the proof of the theorem is detailed in Section 4. We mention in passing that although we focus on $\mathbb{Z}$-valued random variables with finite support in this paper, Theorem 1.4 has an extension to the case where the random variables have infinite support using a similar procedure to that in [46,64,66].

Let us discuss a pleasing application of Theorem 1.4 to proving a key ingredient in rearrangement inequalities on the integers proved by Gabriel [13] (generalizing a result of Hardy and Littlewood [17]) and popularized in the book by Hardy, Littlewood and Pólya [18]. For a finite set A in $\mathbb{Z}$, note that $A^{\#} = \{0, 1, \ldots, |A| - 1\}$; here, as before, we identify the sets A and $A^{\#}$ with their indicator functions.

Corollary 1.5. *If $A_1, A_2, \ldots, A_N$ are finite sets (or indicator functions) in $\mathbb{Z}$, then*

$$A_1 \star A_2 \star \cdots \star A_N \prec A_1^{\#} \star A_2^{\#} \star \cdots \star A_N^{\#}.$$

To see how Corollary 1.5 follows from Theorem 1.4, suppose random variables $X_1, X_2, \ldots, X_N$ are uniformly distributed on the sets $A_1, A_2, \ldots, A_N$ respectively. By applying Theorem 1.4 and multiplying by the appropriate cardinalities (to convert uniform probability mass functions to indicator functions), the desired result follows.

We note that a version of Gabriel's inequality was, in fact, extended to the prime cyclic groups $\mathbb{Z}/p\mathbb{Z}$ by Lev [29]. In our companion paper [46], we develop a further generalization of such rearrangement inequalities (see [46, Theorem 6.2]) in the prime cyclic groups, with a crucial step in our proofs being the leveraging of Lev's set majorization lemma (see [29, Theorem 1] for the full statement). The results of Gabriel [13], Lev [29] and the authors [46] for general non-negative functions rather than indicator functions of sets require additional assumptions because one has to take into account the "shape" of the convolved functions. It is a pleasant feature of the statement and proof above that it does not require such assumptions.

Our second theorem is related to a beautiful and well known result of Sárközy and Szemerédi [56] related to what they called the Erdős–Moser problem (although the paper of Katona [25], which they cite a pre-publication version of, does not discuss it in the published paper, and the problem posed by Erdős in 1947 in the *American Mathematical Monthly* with solutions given by Moser [11] as well as several others, which is cited by several later papers on the Sárközy–Szemerédi result, seems only tangentially related). In any case, the "Erdős–Moser problem" is the following: Given N i.i.d. Bernoulli random variables $Y_1, \ldots, Y_N$, estimate the maximal probability of independent weighted sums over distinct weights:

$$\sup_{k\in\mathbb{Z}} \sup_{0<a_1\neq\cdots\neq a_N} \mathbf{P}\left(a_1Y_1 + a_2Y_2 + \cdots + a_NY_N = k\right).$$

Sárközy and Szemerédi [56] asserted that Erdős and Moser had shown that the maximal probability is of order $\left(\frac{\log N}{N}\right)^{3/2}$ and had conjectured that the logarithmic term could be removed; they proved this conjecture, thus showing that the maximal probability is of order $N^{-3/2}$. However, identification of an extremal set of weights remained open until Stanley [60] used tools from algebraic geometry to show that $(a_1, a_2, \ldots, a_N) = (1, 2, \ldots, N)$ is extremal. A more elementary algebraic proof was soon after given by Proctor [51]. Much more recently, Nguyen [49] not only observed that the maximal probability is in fact $[\sqrt{24/\pi} + o(1)]N^{-3/2}$, but he also showed a stability result around the extremal configuration.

With this background, we are ready to state our second main result.

Theorem 1.6. *Assume that $0 < a_1 < a_2 < \cdots < a_N$ and $1 \le m_N \le m_{N-1} \le \cdots \le m_1$ are all integers. If $Y_1, \ldots, Y_N$ are independent random variables with Y_i having the Binomial $\left(m_i, \frac{1}{2}\right)$ distribution for each i, then*

$$a_1Y_1 + a_2Y_2 + \cdots + a_NY_N \prec Y_1 + 2Y_2 + \cdots + NY_N. \tag{4}$$

The proof of Theorem 1.6 is based on the strong Sperner property of some products of posets, and is detailed in Section 5.

As an application of Theorem 1.6, we can go beyond the Bernoulli assumption in the prior studies of [56,60] discussed above, and identify the extremal weights for the wider class of binomially distributed random variables.

Corollary 1.7. *Let $0 < a_1 \neq \cdots \neq a_N$. Let $Y_1, \ldots, Y_N$ be i.i.d. random variables with the Binomial $\left(m, \frac{1}{2}\right)$ distribution. Then*

$$\mathbf{P}\left(a_1Y_1 + a_2Y_2 + \cdots + a_NY_N = k\right) \le \mathbf{P}\left(Y_1 + 2Y_2 + \cdots + NY_N = \left\lfloor \frac{mN(N+1)}{4} \right\rfloor\right).$$

To see how Corollary 1.7 follows from Theorem 1.6, observe that since Y_i are i.i.d. in the former, we may assume that the a_i are ordered. Then the conclusion follows from Theorem 1.6 by focusing on the largest atom of both distributions, and recognizing that the largest atom of $Y_1 + 2Y_2 + \cdots + NY_N$ is achieved at the midpoint of the range because its distribution is symmetric and unimodal (this latter fact is confirmed by Lemma 5.3, which we discuss later). Of course, Theorem 1.6 can be applied without an identically distributed assumption, but we make this assumption in Corollary 1.7 for simplicity of statement.

This paper is organized as follows. Our original motivation for pursuing this work came from a search for Rényi entropy power inequalities for integer-valued random variables, which is a problem of significant interest in information theory. We explain this motivation and describe how our main results may be applied to obtain new entropy power inequalities in Section 2. The rest of the paper focuses on the proofs of our main results—Section 3 recalls the necessary background on Sperner theory, and the proofs of the two main theorems are detailed in Sections 4 and 5.

2. Applications to Rényi entropy inequalities

2.1. Background on entropy power inequality

We first define the one-parameter family of Rényi entropies for a probability mass function on the integers (these can be defined on more general spaces by using a reference measure other than counting measure, but we do not need the more general notion here).

Definition 2.1. Let X be an integer-valued random variable with probability mass function f. The Rényi entropy of order $\alpha \in (0, 1) \cup (1, +\infty)$ is defined by

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_{i\in\mathbb{Z}} f(i)^\alpha\right).$$

For limiting cases of α, define

$$H_0(X) = \log|\mathrm{supp}(f)|,$$
$$H_1(X) = \sum_{i\in\mathbb{Z}} -f(i)\log f(i),$$
$$H_\infty(X) = -\log\sup_{i\in\mathbb{Z}} f(i).$$

The three special cases are defined in a manner consistent with taking limits of $H_\alpha(X)$ for $\alpha \in (0, 1) \cup (1, +\infty)$. Thus the Rényi entropy of order $\alpha \in [0, \infty]$ is well-defined. In particular, $H_1(\cdot)$ is simply the Shannon entropy $H(\cdot)$.

Entropy inequalities (even just for Shannon entropy) are powerful tools that have found use in virtually all parts of mathematics. For example, just within discrete mathematics, they have been used to obtain bounds for enumeration problems (see, e.g., [22,45,53]), to prove sumset inequalities in additive combinatorics (see, e.g., [40,41]), and to study probabilistic models of discrete phenomena (e.g., independent sets [24], card shuffles [48]), colorings [50]). It is also intrinsically interesting to develop an additive combinatorics of probability measures that treats measures rather than sets as the basic objects of study and uses entropy to measure their "size" [1,6,7,19,27,38,39,54,62].

Among various entropy inequalities, the so-called "entropy power inequality" in Euclidean spaces has been very successfully applied to prove coding theorems or to determine channel capacities for communication problems involving Gaussian noise in Information Theory (see, e.g., [5,65]). The entropy power inequality also plays an important role in Probability Theory (see, e.g., [21,33]) and Convex Geometry (see, e.g., [9,14,43]). The entropy power inequality can be formulated in two different ways. Firstly, the original formulation, which was suggested by Shannon [57] and proved by Stam [59], is the following. For independent random variables X and Y in $\mathbb{R}^d$,

$$e^{\frac{2}{d}h(X+Y)} \geq e^{\frac{2}{d}h(X)} + e^{\frac{2}{d}h(Y)},$$

where $h(X)$ represents the *differential entropy* of X. Formally, if X has a density function f in $\mathbb{R}^d$, then $h(X) = -\int_{\mathbb{R}^d} f(x)\log f(x)dx$. The inequality shows the superadditivity of the "entropy power" with respect to the sum of two independent random variables. Secondly, an equivalent sharp formulation (see, e.g., [63] for discussion) states that

$$h(X+Y) \geq h(Z_X^* + Z_Y^*),$$

where Z_X^* and Z_Y^* are two independent Gaussian distributions with $h(X) = h(Z_X^*)$ and $h(Y) = h(Z_Y^*)$.

The entropy power inequality stated above focuses on the continuous setting of $\mathbb{R}^d$. It has been extensively studied and many refinements exist (see, e.g., [2,35–37,44]). On the other hand, we only have a limited understanding of analogues of the entropy power inequality on discrete domains such as the integers or cyclic groups.

One of the main difficulties is that useful analytic tools in the continuous domain cannot be naturally translated into the discrete domain. For example, one can derive the entropy power inequality in $\mathbb{R}^d$ from the sharp form of Young's inequality for convolution developed by Beckner [4], as observed by Lieb [31], or using inequalities for Fisher information, which is defined using derivatives of the probability density function, as done by Stam [59]. Unfortunately, a non-trivial sharp Young's inequality cannot be achieved in the discrete setting. It is also not obvious what the right definition of Fisher information should be for the discrete setting because discrete derivatives do not satisfy the chain rule (see, e.g., [3,26,34] for possible definitions of discrete Fisher informations). Owing to the difficulty of fitting such approaches into a discrete setting, a general and sharp analogue of the entropy power inequality on the integers has not yet been established.

Nevertheless, it is a natural and interesting question to find a fully satisfactory entropy power inequality on the integers—earlier attempts in this direction include [16,23,64,67]. In our companion paper [46], we established a lower bound on the entropy of sums in prime cyclic groups (including the integers) based on rearrangement inequalities and functional ordering by majorization where the rearrangement of a function f is achieved by shuffling (permuting) the domain of f. While details may be found in [46], the goal of these rearrangement inequalities is to identify optimal permutations that maximize or minimize a sum of pairwise products.

In this paper, we focus on the integer domain or the integer lattice domain. We continue to leverage the idea of majorization used in our companion paper [46]. However, instead of establishing rearrangement inequalities, we take a different path to establish the lower bound inequality of the entropy of sums in integers. Our approach is to establish and utilize the similarity between the strong Sperner property of posets (the origin of this notion lies, of course, in Sperner's theorem [58], but the way we use this notion is inspired by Erdős [12] as described in Section 1) and functional ordering by majorization.

2.2. Two entropy inequalities

A key application of Theorem 1.4 (and also Theorem 1.6) lies in establishing a lower bound on the Rényi entropy of convolutions. As the main tool in translating majorization results to entropy inequalities, we use the following basic lemma.

Lemma 2.2 ([47, *Proposition 3-C.1*]). *Assume that f and g are finitely supported non-negative functions in $\mathbb{Z}$ and $f \prec g$. For any convex function $\Phi : \mathbb{R} \to \mathbb{R}$,*

$$\sum_{i\in\mathbb{Z}} \Phi\left(f(i)\right) \le \sum_{i\in\mathbb{Z}} \Phi\left(g(i)\right).$$

We note that by choosing a convex function $\Phi(x) = -x^\alpha$ for $\alpha \in (0, 1)$, $\Phi(x) = x\log x$ for $\alpha = 1$, and $\Phi(x) = x^\alpha$ for $\alpha \in (1, +\infty)$, and by taking limits when $\alpha \in \{0, \infty\}$, inequalities for Rényi entropies of all orders follow from Lemma 2.2 whenever we have a probability mass function majorized by another. In particular, our two main theorems combined with Lemma 2.2 yield the following propositions.

Proposition 2.3. *If $X_1, \dots, X_N$ are independent and #-log-concave over $\mathbb{Z}$, then*

$$H_\alpha\left(X_1 + \cdots + X_N\right) \ge H_\alpha\left(X_1^\# + \cdots + X_N^\#\right), \tag{5}$$

for $\alpha \in [0, \infty]$.

Rényi entropy inequalities such as this (and others of similar form in our companion paper [46]) have already begun finding utility (see, e.g., [42,68]).

Proposition 2.4. *Let $0 < a_1 < \cdots < a_N$. If Y_i's are independent random variables following Binomial $\left(m_i, \frac{1}{2}\right)$ for $1 \le m_N \le m_{N-1} \le \cdots \le m_1$, then*

$$H_\alpha\left(a_1Y_1 + a_2Y_2 + \cdots + a_NY_N\right) \ge H_\alpha\left(Y_1 + 2Y_2 + \cdots + NY_N\right), \tag{6}$$

for $\alpha \in [0, \infty]$.

Nguyen [49] observed that the optimal solution of the Erdős–Moser problem (for Bernoulli Y_i) minimizes the variance of $a_1Y_1 + a_2Y_2 + \cdots + a_NY_N$ among all choices of distinct positive weights, i.e., for $0 < a_1 \ne \cdots \ne a_N$,

$$\mathrm{Var}\left(Y_1 + 2Y_2 + \cdots + NY_N\right) \le \mathrm{Var}\left(a_1Y_1 + a_2Y_2 + \cdots + a_NY_N\right).$$

Proposition 2.4 implies that the optimal solution of the Erdős–Moser problem also minimizes the Rényi entropy for any order $\alpha \in [0, +\infty]$ (and even for the more general binomial setting).

2.3. Inequalities for uniform distributions on subsets of $\mathbb{Z}^d$

In [67], we proved a discrete entropy power inequality for uniform distributions over finite subsets of the integers $\mathbb{Z}$. In the following lemma, we extend [67, Theorem II.2] from the $\alpha = 1$ case to any Rényi entropy of order $\alpha \ge 1$.

Lemma 2.5. *If X and Y are independent and uniformly distributed over finite sets $A \subset \mathbb{Z}$ and $B \subset \mathbb{Z}$ respectively,*

$$\mathcal{N}_\alpha(X + Y) + 1 \ge \mathcal{N}_\alpha(X) + \mathcal{N}_\alpha(Y), \tag{7}$$

where $\mathcal{N}_\alpha(X) = e^{(1+\alpha)H_\alpha(X)}$ for $\alpha \ge 1$.

Proof. Since the $\alpha = 1$ case is proved in [67], we assume that $\alpha > 1$. Since any uniform distribution over a finite set is #-log-concave, Proposition 2.3 implies that

$$H_\alpha(X + Y) \ge H_\alpha(X^\# + Y^\#).$$

Since $\mathcal{N}_\alpha(X) = \mathcal{N}_\alpha(X^\#)$ and $N_\alpha(Y) = N_\alpha(Y^\#)$ hold trivially, it suffices for proving the inequality (7) to only consider the case where A and B are sets of consecutive integers. Indeed, if we proved this special case, we would have

$$\begin{aligned}\mathcal{N}_\alpha(X + Y) + 1 &\ge \mathcal{N}_\alpha(X^\# + Y^\#) + 1\\ &\ge \mathcal{N}_\alpha(X^\#) + \mathcal{N}_\alpha(Y^\#)\\ &= \mathcal{N}_\alpha(X) + \mathcal{N}_\alpha(Y).\end{aligned}$$

which is the desired statement.

What remains is to prove (7) for uniform distributions on finite sets of consecutive integers. Let $|A| = n$ and $|B| = m$. Since the roles of X and Y are symmetric, we may assume that $n \ge m$. While $H_\alpha(X) = \log n$ and $H_\alpha(Y) = \log m$ due to uniformity, a direct calculation easily shows that

$$H_\alpha(X + Y) = \frac{1}{1-\alpha}\log\left[\sum_{i=1}^{m-1}\frac{i^\alpha}{m^\alpha n^\alpha} + (n - m + 1)\frac{1}{n^\alpha}\right]. \tag{8}$$

First, observe that if $m = 1$, then $H_\alpha(X+Y) = H_\alpha(X)$ and $H_\alpha(Y) = 0$. Then $\mathcal{N}_\alpha(X+Y)+1 = \mathcal{N}_\alpha(X)+1 = \mathcal{N}_\alpha(X)+\mathcal{N}_\alpha(Y)$, and hence the inequality (7) is sharp. Next, consider the expression inside the logarithm in the formula (8):

$$\begin{aligned}\sum_{i=1}^{m-1} \frac{i^\alpha}{m^\alpha n^\alpha} + (n-m+1)\frac{1}{n^\alpha} &= n^{-\alpha}m^{-\alpha}\sum_{i=1}^{m-1} i^\alpha + n^{1-\alpha} - mn^{-\alpha} + n^{-\alpha} \\ &\le n^{-\alpha}m^{-\alpha}\int_1^m x^\alpha dx + n^{1-\alpha} - mn^{-\alpha} + n^{-\alpha} \\ &= n^{1-\alpha}\left[1 + \frac{1}{n} - \frac{\alpha}{1+\alpha}\frac{m}{n} - \frac{1}{1+\alpha}\frac{m^{-\alpha}}{n}\right].\end{aligned}$$

Plugging this bound into (8), setting $k = m/n \in [0, 1]$ and writing

$$\xi(k,n) = 1 + \frac{1}{n} - \frac{\alpha}{1+\alpha}k - \frac{1}{1+\alpha}k^{-\alpha}n^{-(1+\alpha)},$$

we obtain the following lower bound for the Rényi entropy of $X+Y$:

$$H_\alpha(X+Y) \ge \frac{1}{1-\alpha}\left[\log n^{1-\alpha} + \log \xi(k,n)\right].$$

Then

$$\mathcal{N}_\alpha(X+Y) = e^{(1+\alpha)H_\alpha(X+Y)} \ge e^{(1+\alpha)\log n}e^{\frac{1+\alpha}{1-\alpha}\log\xi(k,n)} =: \nu(k,n).$$

Since the $m = 1$ case is already proved, if the following inequality is true, we are done.

$$\nu(k,n) = e^{(1+\alpha)\log n}e^{\frac{1+\alpha}{1-\alpha}\log\xi(k,n)} \ge e^{(1+\alpha)\log n} + e^{(1+\alpha)\log kn} = \mathcal{N}_\alpha(X) + \mathcal{N}_\alpha(Y).$$

By rearranging terms, the above inequality is equivalent to

$$\xi(k,n) \le \left(1 + k^{1+\alpha}\right)^{\frac{1-\alpha}{1+\alpha}}.$$

When $k = \frac{2}{n}$, we can directly show the following inequality is true by elementary calculation:

$$\xi\left(\frac{2}{n}, n\right) = 1 + \frac{1-\alpha}{1+\alpha}\frac{1}{n} - \frac{2^{-\alpha}}{1+\alpha}\frac{1}{n} \le \left(1 + 2^{1+\alpha}n^{-(1+\alpha)}\right)^{\frac{1-\alpha}{1+\alpha}}.$$

Furthermore, it is easy to show that for $k \in [0, 1]$ and $\alpha > 1$

$$\left(1 + k^{1+\alpha}\right)^{\frac{1-\alpha}{1+\alpha}} \ge 1 - \left(1 - 2^{\frac{1-\alpha}{1+\alpha}}\right)k.$$

Hence it suffices to show

$$\xi(k,n) \le 1 - \left(1 - 2^{\frac{1-\alpha}{1+\alpha}}\right)k$$

for all $n \ge 2$ and $k \ge \frac{3}{n}$. Let $\phi(k,n) := 1 - \left(1 - 2^{\frac{1-\alpha}{1+\alpha}}\right)k - \xi(k,n)$. For a fixed $n \ge 2$,

$$\frac{\partial\phi}{\partial k} = -1 + 2^{\frac{1-\alpha}{1+\alpha}} + \frac{\alpha}{1+\alpha}\left(1 - k^{-(1+\alpha)}n^{-(1+\alpha)}\right).$$

Then $\frac{\partial\phi}{\partial k} = 0$ at

$$k^* = \frac{1}{n}\alpha^{\frac{1}{1+\alpha}}\left[-1 + 2^{\frac{1-\alpha}{1+\alpha}} + 2^{\frac{1-\alpha}{1+\alpha}}\alpha\right]^{-\frac{1}{1+\alpha}},$$

and $\frac{\partial\phi}{\partial k} > 0$ for $k > k^*$ and $\frac{\partial\phi}{\partial k} < 0$ for $k < k^*$. Finally, by elementary calculation, we can easily show that

$$\frac{1}{n} \le k^* < \frac{2}{n}.$$

Thus, $\phi(k,n)$ is minimized at $k = \frac{3}{n}$ for $k \ge \frac{3}{n}$. By elementary calculation, we can also confirm that $\phi\left(\frac{3}{n}, n\right) \ge 0$, which completes the proof.

We note that when $k = \frac{2}{n}$, $\phi\left(\frac{2}{n}, n\right) < 0$ for some n; this is why $\phi(k,n) \ge 0$ is only proved when $k \ge \frac{3}{n}$ rather than $k \ge \frac{2}{n}$. □

We remark that the $+1$ term on the left side of inequality (7) is only necessary for point masses (i.e., distributions supported on one point). In other words, if X and Y are independent and uniformly distributed over finite sets of cardinality at least 2, then we in fact have

$$\mathcal{N}_\alpha(X+Y) \ge \mathcal{N}_\alpha(X) + \mathcal{N}_\alpha(Y). \tag{9}$$

However, we highlight the formulation with $+1$ both because of the similarity with the Cauchy–Davenport Theorem [55], and because the discrete entropy power inequality over the integers in Lemma 2.5 can be extended to the integer lattice $\mathbb{Z}^d$.

Theorem 2.6. *If X and Y are uniform distributions over finite sets A and B in $\mathbb{Z}^d$,*

$$\mathcal{N}_\alpha(X+Y)+1 \geq \mathcal{N}_\alpha(X)+\mathcal{N}_\alpha(Y),$$

where $\mathcal{N}_\alpha(X)=e^{(1+\alpha)H_\alpha(X)}$ for $\alpha \geq 1$.

Proof. Consider a point $\mathbf{z}=(z_1,\ldots,z_d)$ in $\mathbb{Z}^d$ where $z_i \geq 0$ for each i. We regard $\mathbf{z}$ as a q-ary representation of an integer value, where q is large and chosen later. In other words, the point $\mathbf{z} \in \mathbb{Z}^d$ can be mapped to a unique integer value in $\mathbb{Z}$:

$$\mathbf{z}=(z_1,\ldots,z_d) \mapsto z_1 q^{d-1}+z_2 q^{d-2}+\cdots+z_d \in \mathbb{Z}. \tag{10}$$

For the set A in $\mathbb{Z}^d$, without loss of generality, we can shift A so that each point contains only non-negative components. Let A' be the set in $\mathbb{Z}$ equivalent to A via the q-ary representation (10). Similarly we can find the set B' in $\mathbb{Z}$ equivalent to B in $\mathbb{Z}^d$. We choose q large enough so that $A \star B$ in $\mathbb{Z}^d$ maps to $A' \star B'$ in $\mathbb{Z}$ via the q-ary representation.

Let X' and Y' be uniform distributions on A' and B' in $\mathbb{Z}$, respectively. This implies

$$H_\alpha(X+Y)=H_\alpha(X'+Y') \geq H_\alpha(X'^{\#}+Y'^{\#}).$$

Then the conclusion follows by applying Lemma 2.5 and from the fact that $H_\alpha(X)=H_\alpha(X')$ and $H_\alpha(Y)=H_\alpha(Y')$. $\square$

Theorem 2.6 uses the exponent $c = 1+\alpha$; Rényi entropy power inequalities with the same exponent in $\mathbb{R}$ were recently explored by Bobkov and Marsiglietti [8] (although it was shown soon after by Li [30] that this exponent can be improved). In fact, these authors proved similar inequalities in $\mathbb{R}^d$, with the exponent $\frac{1+\alpha}{d}$, mimicking the $2/d$ exponent in the original Shannon–Stam entropy power inequality.

The inequality we derived in Theorem 2.6 is independent of the dimensional factor d, and one might wonder whether an entropy power inequality in the integer lattice $\mathbb{Z}^d$ that respects the dimension exists. Even for the subclass of uniform distributions and for $\alpha = 1$, however, one can easily construct counterexamples showing that an exponent of $2/d$ fails in general in $\mathbb{Z}^d$. We remark that in order to develop a discrete Brunn–Minkowski inequality in the integer lattice, Gardner and Gronchi [15] imposed a natural and appropriate dimensional assumption, the main point of which is that at least two points should be assigned to each axis direction. However, the dimensional assumption from [15] is still not sufficient to obtain an improvement of Theorem 2.6 with exponent $\frac{1+\alpha}{d}$ (as can be checked by counterexamples). Hence we leave the discovery of appropriate dimensional entropy inequalities in the integer lattice as an open question for future works.

3. Background on Sperner theory

In this section, we summarize the basic elements of Sperner Theory as needed for our proofs. A comprehensive summary can be found in books by Stanley [61] and Engel [10].

3.1. Partially ordered set (poset)

A set S with a binary relation $\preccurlyeq$ is said to be *partially ordered* if the relation $\preccurlyeq$ satisfies the reflexive, anti-symmetric, and transitive properties, i.e., for any $a, b, c \in S$,

- $a \preccurlyeq a$ (reflexive),
- if $a \preccurlyeq b$ and $b \preccurlyeq a$, then $a = a$ (antisymmetric),
- if $a \preccurlyeq b$ and $b \preccurlyeq c$, then $a \preccurlyeq c$ (transitive).

We emphasize that we use the symbol $\prec$ to represent majorization, and this has no relation to the partial order $\preceq$. If S is partially ordered, we call S a *partially ordered set*, or a *poset*. For $a, b \in S$, a and b are *comparable* if $a \preccurlyeq b$ or $b \preccurlyeq a$. Otherwise, a and b are *incomparable*. A *chain poset* is a poset in which any two elements are comparable. A subset C of S is called a *chain* of S if C is a chain poset as a sub-poset of S. We define *the length* of a chain C to be the number of elements in C.

A subset A of S is called an *antichain* if any two distinct elements of A are incomparable. A subset K of S is called a *k-family* of S if it is a union of at most k antichains. We say a poset S is *weighted* if each element has a positive weight. The *weight function* $w : S \to \mathbb{R}_+$ defines the weight of each element in S. We use a triple $(S, w, \preccurlyeq)$ to represent the weighted poset, but we sometimes omit to mention the weight function w explicitly when we describe a weighted poset. If the poset has no weight function (or *unweighted*), we implicitly assume that each weight of an element is 1.

A chain C of S is called *maximal* if there is no larger chain C' such that $C \subseteq C'$. An element s in S is called *minimal* if $t \preceq s$ implies $s = t$ in S. The element s of S is said to *cover* the element t of S if $t \preceq s$ and if $t \preceq s' \preceq s$ implies $s = s'$ when $s' \neq t$. If every maximal chain of the poset S has length $n+1$, we call S a *graded poset* with rank n. In such a case, we can define a unique *rank function* $\rho : S \to \{0, 1, \ldots, n\}$ of S such that $\rho(a) = 0$ if a is a minimal element of S, and $\rho(b) = \rho(a) + 1$ if b covers a. Then the *rank of* a is assigned to be $\rho(a)$.

Given a weighted and ranked poset $(S, \preceq, w)$, the sum of all weights at the same rank $r \in \{0, 1, \ldots, n\}$ is called the *weighted Whitney number of rank* r. Similarly, if the poset is unweighted, the *Whitney number of rank* r is the total number of elements at the rank r. We say that weighted Whitney numbers are *log-concave* if the sequence of weighted Whitney numbers is log-concave in an increasing order of the rank. We say that weighted Whitney numbers are *rank-symmetric* if the sequence of weighted Whitney numbers is symmetric in an increasing order of the rank. Similarly, we say that weighted Whitney numbers are *rank-unimodal* if the sequence of weighted Whitney numbers is unimodal in an increasing order of the rank.

The weighted and ranked poset $(S, \preceq, w)$ has the *k-Sperner property* if the maximum total weight among all k-families in S equals the largest sum of k weighted Whitney numbers. The weighted and ranked poset $(S, \preceq, w)$ is *strongly Sperner* (or has the *strong Sperner property*) if it is k-Sperner for all $k = 1, 2, \ldots$.

The *product* of the posets S and T is defined to be the Cartesian product $S \times T$, equipped with the partial order defined by the requirement that $(s, t) \preceq (s', t')$ in $S \times T$ if and only if $s \preceq s'$ in S and $t \preceq t'$ in T. If S and T are weighted with weight functions w_S and w_T, then the weight function $w_{S\times T}$ of $S \times T$ is defined to be $w_{S\times T}(s, t) = w_S(s)w_T(t)$.

We will also need later the notion of isomorphism between two posets. We say that two posets $(Q, \preccurlyeq)$ and $(R, \preccurlyeq)$ are *isomorphic* if there exists a bijective map $\phi : Q \to R$ such that $q_1 \preccurlyeq q_2$ iff $\phi(q_1) \preccurlyeq \phi(q_2)$ for $q_1, q_2 \in Q$ and $\phi(q_1), \phi(q_2) \in R$.

3.2. Normalized matching property

Consider a ranked and weighted poset $(S, w, \preccurlyeq)$ with the rank function ρ. For any subset A of S, we define the *upper shade of* A, denoted $\nabla(A)$, as the set of all elements covering A. If $a' \in \nabla(A)$, then there exists an element $a \in A$ such that $a \preceq a'$ and $\rho(a') = \rho(a) + 1$.

Let N_r be the collection of all elements at rank r. A ranked and weighted poset $(S, w, \preccurlyeq)$ is called *normal* if for any antichain A subject to a subset of elements of rank r, the weight sum ratio of A with respect to the weighted Whitney number of rank r is less than or equal to the weight sum ratio of the shade of A at rank $r+1$ with respect to the weighted Whitney number of rank $r + 1$, i.e.,

$$\frac{w(A)}{w(N_r)} \leq \frac{w(\nabla(A))}{w(N_{r+1})} \tag{11}$$

where $A \subseteq N_r$ is an antichain, $w(A)$ is the sum of all weights of elements in A, and $w(N_r)$ is the weighted Whitney number of rank r. Hsieh and Kleitman [20] proved that the normalized matching property is preserved under the product of normal posets under the assumption of log-concave weighted Whitney numbers.

Proposition 3.1 (*See [10, Theorem 4.6.2] or [20]*)**.** *A product of two normal posets with log-concave weighted Whitney numbers is again a normal poset with log-concave weighted Whitney numbers.*

Proposition 3.2 (*See [10, Corollary 4.5.3]*)**.** *A normal poset is strongly Sperner.*

Applying Propositions 3.1 and 3.2 to chain posets, we have the following corollary.

Corollary 3.3. *For each $i \in \{1, \ldots, N\}$, assume that $S(m_i)$ has log-concave weighted Whitney numbers. Then the product of chain posets $S(m_1, \ldots, m_N)$ is strongly Sperner with log-concave weighted Whitney numbers.*

Proof. Since the weight sum ratio in (11) of any antichain for a chain poset is always 1, any weighted chain is normal. Thus, the conclusion follows from Propositions 3.1 and 3.2. □

3.3. Strongly Sperner posets

Let $N, n_1, \ldots, n_N$ be fixed positive integers. A product of chain posets $S(n_1, \ldots, n_N)$ can be defined to be a collection of N-tuples of integers $(a_1, \ldots, a_N)$ such that $0 \leq a_i \leq n_i$ for each $i \in \{1, \ldots, N\}$. The relation $a = (a_1, \ldots, a_N) \preccurlyeq b = (b_1, \ldots, b_N)$ iff $a_i \leq b_i$ for each $i \in \{1, \ldots, N\}$. Fig. 1 shows an example of the product poset $S(2, 3)$.

Next, we introduce the poset $M(m)$, a collection of m-tuples $(a_1, \ldots, a_m)$ such that $0 = a_1 = \cdots = a_i < a_{i+1} < \cdots < a_m \leq m$ with $i \in \{0, \ldots, m\}$. As noted, we allow one exceptional case $i = 0$. If $i = 0$, we mean $a_1 > 0$, and so $0 < a_1 < a_2 < \cdots < a_m \leq m$. The relation $a = (a_1, \ldots, a_n) \preccurlyeq b = (b_1, \ldots, b_m)$ holds iff $a_i \leq b_i$ for all i. The poset $M(m)$ that we have thus defined is graded, with the rank function given by $\rho(a) = \sum_{i=1}^{m} a_i$.

Stanley [60] originally proved that $M(m)$ is rank-symmetric, rank-unimodal, and strongly Sperner, by leveraging ideas from algebraic geometry. Subsequently, Proctor [51] gave a more accessible proof just based on basic linear algebra. Following conventional terminology, we say that a ranked poset is *Peck* if the poset is rank-symmetric, rank-unimodal, and strongly Sperner. Fig. 2 shows an example of the poset $M(5)$.

Lemma 3.4 (*See [51,60]*)**.** *The poset $M(m)$ is a Peck poset.*

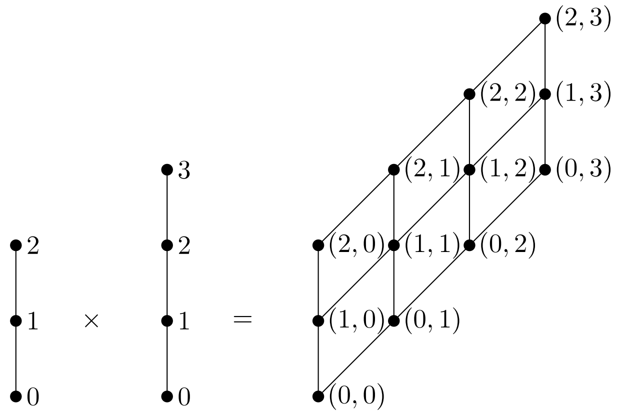

Fig. 1. Poset of $S(2,3)$.

Proctor, Saks, and Sturtevant [52] proved that the Peck property is invariant under the product of posets.

Lemma 3.5 (*See [52, Theorem 3.2]*)**.** *A product of Peck posets is again Peck, and hence strongly Sperner.*

4. Proof of Theorem 1.4

We establish the link between a non-negative function and a weighted chain poset. Consider a #-log-concave function $f=\sum_{r=0}^{n} a_r\{x_r\}$, where $x_0<\cdots<x_n$ and $a_r^2\ge a_{r-1}a_{r+1}$. Setting $S_f:=\mathrm{Supp}(f)=\{x_0,\ldots,x_n\}$, we observe that $(S_f,f,\le)$ forms a ranked and weighted chain poset with the weight function $f(x_r)=a_r$. Thus, we regard a non-negative function with finite support as a weighted chain poset:

$$f\equiv(S_f,f,\preccurlyeq),$$

where the relation $\preccurlyeq$ is the same as the usual $\le$. Since f is #-log-concave, weighted Whitney numbers of S_f are log-concave. Similarly $f^{\#}=\sum_{r=0}^{n} a_r\{r\}$ forms a weighted chain poset $S_{f^{\#}}=\{0,\ldots,n\}$ with log-concave Whitney numbers. Based on the construction, S_f is isomorphic to $S_{f^{\#}}$ by mapping $\phi(x_r)=r$ so that $f(x_r)=f^{\#}(\phi(x_r))$, i.e., the isomorphism map ϕ can be chosen to be the rank function of S_f.

Next, consider N non-negative functions $f_1,\ldots,f_N$, all of which are #-log-concave. Define $F(x^{(1)},\ldots,x^{(N)}):=f_1(x^{(1)})\cdots f_N(x^{(N)})$. Similarly define $F^{\#}(x^{(1)},\ldots,x^{(N)}):=f_1^{\#}(x^{(1)})\cdots f_N^{\#}(x^{(N)})$. As shown above, there exists an isomorphic map ϕ_i between S_{f_i} and $S_{f_i^{\#}}$ for each $i=1,\ldots,N$. Thus, we choose $\phi:S_{f_1}\times\cdots\times S_{f_N}\to S_{f_1^{\#}}\times\cdots\times S_{f_N^{\#}}$ by

$$\phi\left(x_{r_1}^{(1)},\ldots,x_{r_N}^{(N)}\right)=\left(\phi_1(x_{r_1}^{(1)}),\ldots,\phi_N(x_{r_N}^{(N)})\right). \tag{12}$$

Then $S_{f_1}\times\cdots\times S_{f_N}$ is isomorphic to $S_{f_1^{\#}}\times\cdots\times S_{f_N^{\#}}$ by ϕ. i.e.

$$\left(S_{f_1}\times\cdots\times S_{f_N},F,\preccurlyeq\right)\equiv\left(S_{f_1^{\#}}\times\cdots\times S_{f_N^{\#}},F^{\#},\preccurlyeq\right),$$

where $F\left(x_{r_1}^{(1)},\ldots,x_{r_N}^{(N)}\right)=F^{\#}\left(\phi_1(x_{r_1}^{(1)}),\ldots,\phi_N(x_{r_N}^{(N)})\right)$.

Lemma 4.1. *$S_{f_1^{\#}}\times\cdots\times S_{f_N^{\#}}$ forms a normal poset with log-concave weighted Whitney numbers.*

Proof. Each $S_{f_i}^{\#}$ is a chain, thus it is a normal poset with log-concave weights. Corollary 3.3 confirms that the product of normal posets is again normal with log-concave weighted Whitney numbers. □

Next, we establish a link between a product of posets and a convolution of non-negative functions through an antichain. We define a level set

$$L[x]:=\left\{\left(x^{(1)},\ldots,x^{(N)}\right):x^{(1)}+\cdots+x^{(N)}=x,\,x^{(i)}\in S_{f_i}\text{ for }i=1,\ldots,N\right\}.$$

Lemma 4.2. *$\phi\,(L[x])$ forms an antichain in $S_{f_1^{\#}}\times\cdots\times S_{f_N^{\#}}$.*

Please cite this article as: M. Madiman, L. Wang and J.O. Woo, Majorization and Rényi entropy inequalities via Sperner theory, Discrete Mathematics (2019), https://doi.org/10.1016/j.disc.2019.03.002.

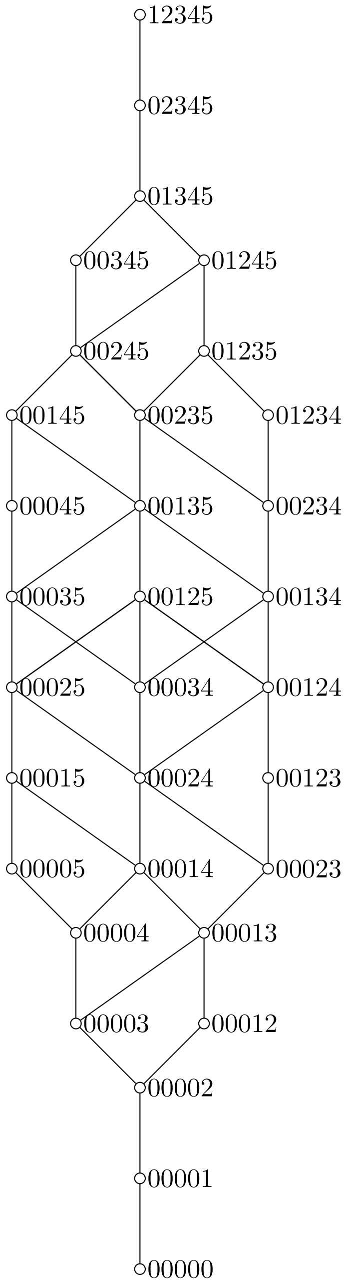

Fig. 2. Poset of $M(5)$.

Proof. Note that ϕ in (12) is a bijective and order-preserving map. Thus, it suffices to consider elements in $L[x]$. Suppose that there exist two distinct comparable elements $\mathbf{x} := \left(x^{(1)}, \ldots, x^{(N)}\right)$ and $\mathbf{y} := \left(y^{(1)}, \ldots, y^{(N)}\right)$ such that $\mathbf{x}, \mathbf{y} \in L[x]$ and $\mathbf{x} \preccurlyeq \mathbf{y}$. This implies $x^{(i)} \leq y^{(i)}$ for each $i = 1, \ldots, N$. Since $\mathbf{x}$ and $\mathbf{y}$ are distinct, there exists some j such that $x^{(j)} < y^{(j)}$. Hence

$$x^{(1)} + \cdots + x^{(N)} < y^{(1)} + \cdots + y^{(N)}.$$

This contradicts the fact that both $\mathbf{x}$ and $\mathbf{y}$ are in $L[x]$. $\square$

Since $S_{f_1^\#} \times \cdots \times S_{f_N^\#}$ is strongly Sperner, majorization follows.

Proposition 4.3. *If $f_1, \ldots, f_N$ are #-log-concave probability mass functions,*

$$f_1 \star \cdots \star f_N \prec f_1^{\#} \star \cdots \star f_N^{\#}.$$

Proof. Let $f_{\text{sum}} := f_1 \star \cdots \star f_N$ and $f_{\text{opt}} := f_1^{\#} \star \cdots \star f_N^{\#}$. We may write out the convolution f_{sum} as

$$f_{\text{sum}}(x) = \sum_{(x^{(1)},\ldots,x^{(N)}) \in L[x]} f_1(x^{(1)}) \cdots f_N(x^{(N)}).$$

The isomorphism ρ in (12) and Lemma 4.2 imply that $f_{\text{sum}}(x)$ can be regarded as a sum of weights of an antichain in $S_{f_1^{\#}} \times \cdots \times S_{f_N^{\#}}$. Since $S_{f_1^{\#}} \times \cdots \times S_{f_N^{\#}}$ is a normal poset by Lemma 4.1, $S_{f_1^{\#}} \times \cdots \times S_{f_N^{\#}}$ is strongly Sperner. Therefore

$$\sum_{i=1}^{k} f_{\text{sum}}^{[i]} \le \sum_{i=1}^{k} f_{\text{opt}}^{[i]},$$

where the left-hand side corresponds to the sum of k antichains and the right-hand side corresponds to the sum of k largest Whitney numbers in $S_{f_1^{\#}} \times \cdots \times S_{f_N^{\#}}$. Since f_{sum} and f_{opt} are still probability mass functions,

$$\sum_{i=1}^{M} f_{\text{sum}}^{[i]} = \sum_{i=1}^{M} f_{\text{opt}}^{[i]} = 1,$$

for some sufficiently large $M > 0$. Thus we have the desired majorization. $\square$

To conclude this section, we note that Theorem 1.4 is simply a restatement of Proposition 4.3 using random variables rather than their probability mass functions.

5. Proof of Theorem 1.6

Assume that $0 < a_1 < \cdots < a_N$. Let $X_{i,j}$ for $1 \le i \le N$ and $1 \le j \le m_i \le N$ be independent random variables following Bernoulli $\left(\frac{1}{2}\right)$. From the assumption, $1 \le m_N \le m_{N-1} \le \cdots \le m_1$. Then, we can decompose each Y_i as follows:

$$\begin{aligned} Y_1 &:= X_{1,1} + \cdots + X_{1,m_1}, \\ &\vdots \qquad\quad \vdots \\ Y_N &:= X_{N,1} + \cdots + X_{N,m_N}. \end{aligned}$$

By the construction, $Y_1, \ldots, Y_N$ are independent random variables, with $Y_i \sim$ Binomial $\left(m_i, \frac{1}{2}\right)$ and $1 \le m_N \le m_{N-1} \le \cdots \le m_1$. We denote by n_j the number of defined $X_{i,j}$'s for each $1 \le j \le N$. Let $\mathbf{Z}_j := \left(X_{1,j}, \ldots, X_{m_j,j}\right)$.

For each j, consider an element $\mathbf{i}_j = (b_1, \ldots, b_{m_j})$ in $M(m_j)$. We encode $b_k = i > 0$ for some k in $\mathbf{i}_j$ iff $X_{i,j} = 1$. Otherwise, $b_k = 0$. For example, when $m_j = 5$,

$$\mathbf{i}_j = (0, 0, 0, 2, 4) \quad \text{iff} \quad \mathbf{Z}_j = \left(X_{1,j} = 0, X_{2,j} = 1, X_{3,j} = 0, X_{4,j} = 1, X_{5,j} = 0\right).$$

Thus as described above, there exists a bijective link between an element $\mathbf{i}_j$ in $M(m_j)$ and each realization of the random vector $\mathbf{Z}_j$. Furthermore, we are able to construct another bijective map between an element $(\mathbf{i}_1, \ldots, \mathbf{i}_N)$ in $M(m_1) \times \cdots \times M(m_N)$ and each realization of the random array $(\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$.

Let $L_j(\mathbf{Z}_j) := a_1 X_{1,j} + \cdots + a_j X_{m_j,j}$ for each $1 \le j \le N$. Based on the construction, we see that

$$Y_{\text{sum}} := a_1 Y_1 + \cdots + a_N Y_N = L_1(\mathbf{Z}_1) + \cdots + L_N(\mathbf{Z}_N).$$

As in Section 4, we define the level set $L_j[x_j]$ in $M(m_j)$ as follows:

$$L_j[x_j] := \left\{\mathbf{i}_j \in M(m_j) : \mathbf{i}_j \text{ bijectively corresponds to } \mathbf{Z}_j \text{ such that } L_j(\mathbf{Z}_j) = x_j\right\}.$$

Lemma 5.1. *$L_j[x_j]$ forms an antichain in $M(m_j)$.*

Proof. Suppose that there exist two distinct elements $\mathbf{i}_j$ and $\mathbf{i}'_j$ in $L_j[x_j]$ such that $\mathbf{i}_j \preccurlyeq \mathbf{i}'_j$. Assume that $\mathbf{i}_j$ and $\mathbf{i}'_j$ correspond to $\mathbf{Z}_j$ and $\mathbf{Z}'_j$, respectively. Since $\mathbf{i}_j = (b_1, \ldots, b_{m_j})$ and $\mathbf{i}'_j = (b'_1, \ldots, b'_{m_j})$ are distinct, there exists some $k > 0$ such that $b_k < b'_k$. Since a_i are positive and ordered, we must have $0 < a_{b_k} < a_{b'_k}$, which implies that

$$L_j(\mathbf{Z}_j) < L_j(\mathbf{Z}'_j).$$

This contradicts the assumption that both $\mathbf{i}_j$ and $\mathbf{i}'_j$ are in $L_j[x_j]$, thus proving the desired statement. $\square$

Please cite this article as: M. Madiman, L. Wang and J.O. Woo, Majorization and Rényi entropy inequalities via Sperner theory, Discrete Mathematics (2019), https://doi.org/10.1016/j.disc.2019.03.002.

More generally, we define a level set $L[x]$ in $M(m_1) \times \cdots \times M(m_N)$ as

$$L[x] := \{(\mathbf{i}_1, \ldots, \mathbf{i}_N) \in M(m_1) \times \cdots \times M(m_N) : (\mathbf{i}_1, \ldots, \mathbf{i}_N) \\ \text{bijectively corresponds to } (\mathbf{Z}_1, \ldots, \mathbf{Z}_N) \text{ such that } L_1(\mathbf{Z}_1) + \cdots + L_N(\mathbf{Z}_N) = x\}.$$

Following the same argument in Lemma 5.1, we see that $L[x]$ forms an antichain. We omit the proof for simplicity.

Lemma 5.2. *$L[x]$ forms an antichain in $M(m_1) \times \cdots \times M(m_N)$.*

Next, we invoke Lemma 3.5, which says that a product of Peck posets is again Peck, together with Lemma 3.4, which asserts that $M(m)$ is Peck.

Lemma 5.3. *$M(m_1) \times \cdots \times M(m_N)$ is Peck, and thus strongly Sperner.*

We note that $|M(m_j)| = 2^{m_j}$, so $|M(m_1) \times \cdots \times M(m_N)| = 2^{m_1+\cdots+m_N}$. Then,

$$\mathbf{P}(Y_{\mathrm{sum}} = x) = \frac{|L[x]|}{2^{m_1+\cdots+m_N}}.$$

Before explaining the link between the strong Sperner property and majorization, it is necessary to identify the saturated or extremal situation. Let

$$Y_{\mathrm{opt}} := Y_1 + \cdots + NY_N,$$

which corresponds to the coefficients $a_1 = 1, \ldots, a_N = N$. Clearly we may write $Y_{\mathrm{opt}} = R_1(\mathbf{Z}_1) + \cdots + R_N(\mathbf{Z}_N)$, where $R_j(\mathbf{Z}_j) := X_{1,j} + \cdots + m_j X_{m_j,j}$ for each $1 \le j \le N$. Then, as Stanley and Proctor explained in [51,60], the size of each level set of $R_j(\mathbf{Z}_j)$ has a bijective correspondence to a Whitney number of $M(m_j)$ by matching the rank to the value $R_j(\mathbf{Z}_j)$. Hence each level set of Y_{opt} has a bijective correspondence to a Whitney number of $M(m_1) \times \cdots \times M(m_N)$ by applying the property of the product of posets. Therefore we confirm that Y_{opt} is the extremal case.

Now it remains to establish majorization through the strong Sperner property of $M(m_1) \times \cdots \times M(m_N)$. Let $\mathbb{Z}_+^0$ be the set of non-negative integers. Observe that

$$\left(2^{m_1+\cdots+m_N}\right) \sup_{C \subset \mathbb{Z}_+^0} \sum_{|C|=k} \mathbf{P}(Y_{\mathrm{sum}} \in C) \le \left(2^{m_1+\cdots+m_N}\right) \sup_{C \subset \mathbb{Z}_+^0} \sum_{|C|=k} \mathbf{P}\left(Y_{\mathrm{opt}} \in C\right), \tag{13}$$

where the left-hand side corresponds to the sum of weights from k antichains in $M(m_1) \times \cdots \times M(m_N)$ and the right-hand side exactly corresponds to the sum of k-largest Whitney numbers from $M(m_1) \times \cdots \times M(m_N)$. We see that Eq. (13) is confirming the condition (1). The condition (2) follows using the fact that the total sum of a probability mass function equals 1. Thus we conclude that Y_{sum} is majorized by Y_{opt}, which completes the proof of Theorem 1.6.

Acknowledgments

The authors are indebted to Manjunath Krishnapur for very useful discussions, and to two anonymous reviewers for detailed and constructive comments that greatly improved the exposition. This work was supported in part by the U.S. National Science Foundation through grants DMS-1409504 (CAREER) and CCF-1346564.

References

[1] E. Abbe, J. Li, M. Madiman, Entropies of weighted sums in cyclic groups and an application to polar codes, Entropy, 19 (9) (2017), Special issue on Entropy and Information Inequalities (edited by V. Jog and J. Melbourne).
[2] S. Artstein, K.M. Ball, F. Barthe, A. Naor, Solution of Shannon's problem on the monotonicity of entropy, J. Amer. Math. Soc. 17 (4) (2004) 975–982 (electronic).
[3] A.D. Barbour, O. Johnson, I. Kontoyiannis, M. Madiman, Compound Poisson approximation via information functionals, Electron. J. Probab. 15 (42) (2010) 1344–1368.
[4] W. Beckner, Inequalities in Fourier analysis, Ann. of Math. (2) 102 (1) (1975) 159–182.
[5] P. Bergmans, A simple converse for broadcast channels with additive white Gaussian noise, IEEE Trans. Inform. Theory 20 (2) (1974) 279–280.
[6] S. Bobkov, M. Madiman, Dimensional behaviour of entropy and information, C. R. Acad. Sci. Paris I 349 (2011) 201–204.
[7] S.G. Bobkov, M.M. Madiman, On the problem of reversibility of the entropy power inequality, in: Limit Theorems in Probability, Statistics and Number Theory, in: Springer Proc. Math. Stat., vol. 42, Springer, Heidelberg, 2013, pp. 61–74.
[8] S.G. Bobkov, A. Marsiglietti, Variants of the entropy power inequality, IEEE Trans. Inform. Theory 63 (12) (2017) 7747–7752.
[9] A. Dembo, T. Cover, J. Thomas, Information-theoretic inequalities, IEEE Trans. Inform. Theory 37 (6) (1991) 1501–1518.
[10] K. Engel, Sperner theory, in: Encyclopedia of Mathematics and its Applications, vol. 65, Cambridge University Press, Cambridge, 1997.
[11] P. Erdős, L. Moser, Elementary problems and solutions: Solutions: E736, Amer. Math. Monthly 54 (4) (1947) 229–230.
[12] P. Erdös, On a lemma of Littlewood and Offord, Bull. Amer. Math. Soc. 51 (1945) 898–902.
[13] R.M. Gabriel, The rearrangement of positive Fourier coefficients, Proc. Lond. Math. Soc. 2 (1) (1932) 32–51.
[14] R.J. Gardner, The Brunn-Minkowski inequality, Bull. Amer. Math. Soc. (N.S.) 39 (3) (2002) 355–405 (electronic).
[15] R.J. Gardner, P. Gronchi, A Brunn-Minkowski inequality for the integer lattice, Trans. Amer. Math. Soc. 353 (10) (2001) 3995–4024 (electronic).
[16] S. Haghighatshoar, E. Abbe, E. Telatar, A new entropy power inequality for integer-valued random variables, IEEE Trans. Inform. Theory 60 (7) (2014) 3787–3796.
[17] G.H. Hardy, J.E. Littlewood, Notes on the theory of series (VIII): An inequality, J. Lond. Math. Soc. 3 (2) (1928) 105–110.

[18] G.H. Hardy, J.E. Littlewood, G. Pólya, Inequalities, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1988, reprint of the 1952 edition.
[19] M. Hochman, On self-similar sets with overlaps and inverse theorems for entropy, Ann. of Math. (2) 180 (2) (2014) 773–822.
[20] W.N. Hsieh, D.J. Kleitman, Normalized matching in direct products of partial orders, Stud. Appl. Math. 52 (1973) 285–289.
[21] O. Johnson, Information Theory and the Central Limit Theorem, Imperial College Press, London, 2004.
[22] O. Johnson, I. Kontoyiannis, M. Madiman, Log-concavity, ultra-log-concavity, and a maximum entropy property of discrete compound Poisson measures, Discrete Appl. Math. 161 (2013) 1232–1250, http://dx.doi.org/10.1016/j.dam.2011.08.025.
[23] O. Johnson, Y. Yu, Monotonicity, thinning, and discrete versions of the entropy power inequality, IEEE Trans. Inform. Theory 56 (11) (2010) 5387–5395.
[24] J. Kahn, An entropy approach to the hard-core model on bipartite graphs, Combin. Probab. Comput. 10 (2001) 219–237.
[25] G. Katona, On a conjecture of Erdős and a stronger form of Sperner's theorem, Stud. Sci. Math. Hungar. 1 (1966) 59–63.
[26] I. Kontoyiannis, P. Harremoës, O. Johnson, Entropy and the law of small numbers, IEEE Trans. Inform. Theory 51 (2) (2005) 466–472.
[27] I. Kontoyiannis, M. Madiman, Sumset and inverse sumset inequalities for differential entropy and mutual information, IEEE Trans. Inform. Theory 60 (8) (2014) 4503–4514.
[28] M. Krishnapur, Anti-concentration inequalities, Lecture notes.
[29] V.F. Lev, Linear equations over $\mathbb{F}_p$ and moments of exponential sums, Duke Math. J. 107 (2) (2001) 239–263.
[30] J. Li, Rényi entropy power inequality and a reverse, Studia Math. 242 (2018) 303–319.
[31] E.H. Lieb, Proof of an entropy conjecture of Wehrl, Comm. Math. Phys. 62 (1) (1978) 35–41.
[32] J.E. Littlewood, A.C. Offord, On the number of real roots of a random algebraic equation. III, Rec. Math. [Mat. Sbornik] N.S. 12 (54) (1943) 277–286.
[33] M. Madiman, A primer on entropic limit theorems, Preprint.
[34] M. Madiman, Topics in Information Theory, Probability and Statistics (Ph.D. thesis), Brown University, Providence RI, 2005.
[35] M. Madiman, A. Barron, The monotonicity of information in the central limit theorem and entropy power inequalities, in: Proc. IEEE Intl. Symp. Inform. Theory, Seattle, 2006, pp. 1021–1025.
[36] M. Madiman, A. Barron, Generalized entropy power inequalities and monotonicity properties of information, IEEE Trans. Inform. Theory 53 (7) (2007) 2317–2329.
[37] M. Madiman, F. Ghassemi, Combinatorial entropy power inequalities: A preliminary study of the Stam region, IEEE Trans. Inform. Theory 65 (3) (2019) 1375–1386, available online at arXiv:1704.01177.
[38] M. Madiman, I. Kontoyiannis, The entropies of the sum and the difference of two IID random variables are not too different, in: Proc. IEEE Intl. Symp. Inform. Theory, Austin, Texas, 2010.
[39] M. Madiman, I. Kontoyiannis, Entropy bounds on abelian groups and the Ruzsa divergence, IEEE Trans. Inform. Theory 64 (1) (2018) 77–92.
[40] M. Madiman, A. Marcus, P. Tetali, Information-theoretic inequalities in additive combinatorics, in: Proc. IEEE Inform. Theory Workshop, Cairo, Egypt, 2010.
[41] M. Madiman, A. Marcus, P. Tetali, Entropy and set cardinality inequalities for partition-determined functions, Random Structures Algorithm 40 (2012) 399–424.
[42] M. Madiman, J. Melbourne, P. Xu, Rogozin's convolution inequality for locally compact groups, Preprint, arXiv:1705.00642.
[43] M. Madiman, J. Melbourne, P. Xu, Forward and reverse entropy power inequalities in convex geometry, in: E. Carlen, M. Madiman, E.M. Werner (Eds.), Convexity and Concentration, in: IMA Volumes in Mathematics and its Applications, vol. 161, Springer, 2017, pp. 427–485.
[44] M. Madiman, P. Nayar, T. Tkocz, Two remarks on generalized entropy power inequalities, Preprint.
[45] M. Madiman, P. Tetali, Information inequalities for joint distributions, with interpretations and applications, IEEE Trans. Inform. Theory 56 (6) (2010) 2699–2713.
[46] M. Madiman, L. Wang, J.O. Woo, Rényi entropy inequalities for sums in prime cyclic groups, Preprint, arXiv:1710.00812.
[47] A.W. Marshall, I. Olkin, B.C. Arnold, Inequalities: Theory of Majorization and Its Applications, second ed., in: Springer Series in Statistics, Springer, New York, 2011.
[48] B. Morris, Improved mixing time bounds for the Thorp shuffle, Combin. Probab. Comput. 22 (1) (2013) 118–132.
[49] H.H. Nguyen, A new approach to an old problem of Erdős and Moser, J. Combin. Theory Ser. A 119 (5) (2012) 977–993.
[50] R. Peled, Y. Spinka, Rigidity of proper colorings of $\mathbb{Z}^d$, Preprint, arXiv:1808.03597.
[51] R.A. Proctor, Solution of two difficult combinatorial problems with linear algebra, Amer. Math. Monthly 89 (10) (1982) 721–734.
[52] R.A. Proctor, M.E. Saks, D.G. Sturtevant, Product partial orders with the Sperner property, Discrete Math. 30 (2) (1980) 173–180.
[53] J. Radhakrishnan, Entropy and counting, in: IIT Kharagpur Golden Jubilee Volume, 2001.
[54] I.Z. Ruzsa, Sumsets and entropy, Random Structures Algorithms 34 (1) (2009) 1–10.
[55] I.Z. Ruzsa, Sumsets and structure, in: Combinatorial Number Theory and Additive Group Theory, in: Adv. Courses Math. CRM Barcelona, Birkhäuser Verlag, Basel, 2009, pp. 87–210.
[56] A. Sárközi, E. Szemerédi, Über ein problem von Erdős und Moser, Acta Arith. 11 (1965) 205–208.
[57] C. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.
[58] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, Math. Z. 27 (1) (1928) 544–548.
[59] A. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, Inf. Control 2 (1959) 101–112.
[60] R.P. Stanley, Weyl groups, the hard Lefschetz theorem, and the Sperner property, SIAM J. Algebr. Discrete Methods 1 (2) (1980) 168–184.
[61] R.P. Stanley, Enumerative Combinatorics, vol. 1, second ed., in: Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge University Press, Cambridge, 2012.
[62] T. Tao, Sumset and inverse sumset theory for Shannon entropy, Combin. Probab. Comput. 19 (4) (2010) 603–639.
[63] L. Wang, M. Madiman, Beyond the entropy power inequality, via rearrangements, IEEE Trans. Inform. Theory 60 (9) (2014) 5116–5137.
[64] L. Wang, J.O. Woo, M. Madiman, A lower bound on the Rényi entropy of convolutions in the integers, in: Proc. IEEE Intl. Symp. Inform. Theory, Honolulu, Hawaii, 2014, pp. 2829–2833.
[65] H. Weingarten, Y. Steinberg, S. Shamai, The capacity region of the Gaussian multiple-input multiple-output broadcast channel, IEEE Trans. Inform. Theory 52 (9) (2006) 3936–3964.
[66] J.O. Woo, Information Theoretic Inequalities, Limit Theorems, and Universal Compression over Unknown Alphabets (Ph.D. thesis), Yale University, New Haven, 2015.
[67] J.O. Woo, M. Madiman, A discrete entropy power inequality for uniform distributions, in: Proc. IEEE Intl. Symp. Inform. Theory, Hong Kong, China, 2015.
[68] P. Xu, J. Melbourne, M. Madiman, A new bound on the L^p Lebesgue constant, Preprint, arXiv:1808.07732.