Combinatorics, Probability and Computing (2019) 28, 100–129. © Cambridge University Press 2018
doi:10.1017/S0963548318000494

A Combinatorial Approach to Small Ball Inequalities for Sums and Differences

JIANGE LI and MOKSHAY MADIMAN†

Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, USA
(e-mail: lijiange@udel.edu, madiman@udel.edu)

Received 2 October 2016; revised 15 June 2018; first published online 6 Nov 2018

Small ball inequalities have been extensively studied in the setting of Gaussian processes and associated Banach or Hilbert spaces. In this paper, we focus on studying small ball probabilities for sums or differences of independent, identically distributed random elements taking values in very general sets. Depending on the setting – abelian or non-abelian groups, or vector spaces, or Banach spaces – we provide a collection of inequalities relating different small ball probabilities that are sharp in many cases of interest. We prove these distribution-free probabilistic inequalities by showing that underlying them are inequalities of extremal combinatorial nature, related among other things to classical packing problems such as the kissing number problem. Applications are given to moment inequalities.

2010 *Mathematics subject classification*: Primary 60C05
Secondary 05D05, 60B15

1. Introduction

Given the ubiquity of sums of independent, identically distributed (henceforth, i.i.d.) random variables in probability theory (as well as, indirectly, in many other parts of mathematics), it is natural to look for ways to estimate the probability that a sum lies in a given measurable set. In general, this can be a rather complex calculation, and is often intractable. The *raison d'être* of this paper is the fact that it is often much easier to estimate the probability that a symmetric random variable lies in a symmetric set; so if we can find a way to relate the desired probability to a probability of this type, then in many circumstances our task is significantly simplified.

The most general setting in which we can talk about sums (and symmetry) is that of group-valued random variables, where the group operation represents summation. Thus, to state our problem more precisely, consider i.i.d. random variables X and Y taking

† This work was supported in part by the US National Science Foundation through grants CCF-1346564 and DMS-1409504 (CAREER)

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

values in a (possibly non-abelian) topological group with group operation $+$ and its Borel σ-field; then our problem is to find good bounds on $\mathbb{P}(X+Y \in F)$ for arbitrary measurable F in terms of $\mathbb{P}(X-Y \in K)$ for symmetric (*i.e.* closed with respect to inversion in the group) measurable sets K. Since the distribution of $X-Y$ is always symmetric in that it is invariant with respect to inversion of the random variable, this would provide a reduction of the form mentioned earlier. We also study a related problem, namely that of estimating $\mathbb{P}(X-Y \in F)$ for arbitrary measurable F in terms of $\mathbb{P}(X-Y \in K)$ for symmetric measurable sets K.

It might seem that the problem stated is somewhat abstruse; however, it is closely related to a number of influential streams of recent research. To highlight these connections, we discuss the problem from various perspectives.

Symmetrization. Symmetrization is a basic and powerful meta-technique that arises in many different guises in different parts of mathematics. Instances include Steiner symmetrization in convex geometry and the study of isoperimetric phenomena [13, 14], Rademacher symmetrization in probability in Banach spaces and empirical process theory [30, 35, 61, 67], the use of rearrangements in functional analysis [3, 53], the study of partial differential equations [46] and probability theory [15, 73, 74], and others too numerous to mention. One goal of this paper is to develop a symmetrization technique for estimating small ball probabilities of sums and differences of i.i.d. random variables. We call these small ball probabilities even though there may be no 'ball' under consideration (for instance, no norm in the general group settings that we will consider), because when considered in the context of finite-dimensional vector spaces, these are related to inequalities for the probability of lying in a ball with respect to some norm.

Concentration functions. The notion of the concentration function was introduced by Paul Lévy, as a means of describing in a flexible way the spread or concentration of a real-valued random variable that may not have finite moments. For a real-valued random variable X with distribution P_X, the Lévy concentration function is given by $Q(X,s) = \sup_{x\in\mathbb{R}} P_X([x,x+s]))$ for $s>0$. While a great deal of attention had already been paid to concentration functions in classical probability theory (see *e.g.* [20, 23, 24, 34, 36, 47, 48, 52, 63, 68]), their study has received renewed attention in recent years [18, 19, 22, 27, 65, 69] because of the relevance of arithmetic structure to the concentration function of linear combinations of i.i.d. random variables, as well as applications to random matrix theory. While we do not directly address the literature on concentration functions in our paper, our results may be seen as providing bounds on multidimensional or non-Euclidean analogues of concentration functions in general spaces. Indeed, a natural way to define the concentration function in a general setting, say an abelian group G, would be to set

$$Q(X,F) = \sup_{x\in G} \mathbb{P}(X \in x+F),$$

where the set-valued parameter F plays the role of the parameter s in the definition $Q(X,s)$ of the concentration function for real-valued random variables. Since the constants that

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

appear in our results are packing numbers $N(F,K)$ that are invariant with respect to translations of F, our results directly imply concentration function bounds. For instance, Theorem 3.2 implies the following statement: *If F is an arbitrary measurable subset of an abelian topological group G and K is a symmetric measurable subset of G containing the identity in its interior, then for any pair X, Y of G-valued i.i.d. random variables, we have*

$$Q(X+Y,F) \leqslant N(F,K) \cdot Q(X-Y,K). \tag{1.1}$$

Rényi entropy comparisons. By taking balls of vanishing radius in a metrizable, locally compact abelian group G, and considering independent random variables X and Y drawn from a distribution on G that has a density with respect to the Haar measure, the inequality (1.1) yields an inequality relating the essential supremum of the density of $X+Y$ and that of the density of $X-Y$. This may be interpreted as relating the Rényi entropies of order ∞ of $X+Y$ and $X-Y$. Inequalities for these quantities have attracted interest from different communities including information theory and convex geometry (see *e.g.* [57, 58, 59, 60]), and comparing Rényi entropies of $X+Y$ and $X-Y$ in particular is closely related to the study of more-sums-than-differences sets (see *e.g.* [1]) as well as sum-difference inequalities (see *e.g.* [55, 56]) in additive combinatorics.

Packing problems/extremal combinatorics. In 1995, Alon and Yuster [2] showed that for i.i.d. real-valued random variables X and Y,

$$\mathbb{P}(|X-Y| \leqslant b) < (2\lceil b/a \rceil - 1) \cdot \mathbb{P}(|X-Y| \leqslant a), \tag{1.2}$$

thus answering a question raised by Margulis. They also observed that the optimal constants in such estimates are closely related to the kissing number problem, which is a long-standing problem in geometry; indeed, the kissing number in $\mathbb{R}^3$ was a subject of discussion between Isaac Newton and David Gregory in the 1690s. A similar probabilistic inequality proved by Katona [38] is closely related to Turán-type theorems for triangle-free graphs. It turns out that behind the main results of this paper, which among other things significantly generalize the inequality (1.2) of [2], are actually statements from extremal combinatorics, which we prove en route to proving our main results. This strengthens the link between extremal combinatorial phenomena and probabilistic inequalities, in a much more general setting than that of [2], in analogy with similar links developed by Katona in a series of papers (see *e.g.* [37, 45]).

Moment inequalities. Probability bounds are of course closely related to moment inequalities, and our results in particular can be used to develop a number of moment inequalities for functions of sums and differences of random variables under various assumptions on the distribution and/or the function. Such inequalities are of intrinsic interest since they serve as tools in a variety of areas.

Random walks. For $0 < a < 2b$, the following sharp symmetrization inequality for i.i.d. real-valued random variables X and Y is proved in [21]:

$$\mathbb{P}(|X+Y| \leqslant b) < \lceil 2b/a \rceil \cdot \mathbb{P}(|X-Y| \leqslant a). \tag{1.3}$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

For $a \geqslant 2b$, the estimate still holds with '$\leqslant$' in the middle. This generalizes the earlier work of Siegmund-Schultze and von Weizsächer [71], which considered the special case $a = b$ and used it as a key ingredient in studying the level crossing probabilities for random walks on the real line. The results of this paper contain those of [21], and although we do not investigate this direction further here, it is conceivable that our results would be useful in the study of random walks on groups.

Having briefly provided motivation from different points of view, let us say something about our methods, especially as compared with those of [21], which focused on random variables taking values in a separable Banach space and may be seen as a precursor to this paper. In exploring small ball inequalities for sums and differences on general groups, the reason we study the problem from a combinatorial point of view is twofold: firstly, it seems to be impossible to generalize the analytical technique developed in [21] to the group setting because it relies essentially on the availability of a dilation operation on the space, and secondly (and perhaps more importantly), it is reasonable to expect a certain extremal combinatorial nature underlying these probabilistic inequalities, given that they are independent of the probability distributions imposed on our random variables.

In Section 2, we justify this intuition that distribution-free probabilistic estimates should be connected to extremal combinatorial phenomena, which essentially follows from the Law of Large Numbers. Since our findings, not surprisingly, differ according to how much structure we assume of our ambient set, we deploy Sections 3, 4 and 5 respectively to explore what can be said for abelian topological groups, non-abelian topological groups, and topological vector spaces. In Section 6, we discuss the tightness of our inequalities. Section 7 is devoted to various applications, including moment inequalities in normed vector spaces.

2. A combinatorial perspective on distribution-free inequalities

In this section, we demonstrate a combinatorial approach, which enables us to prove distribution-free probabilistic inequalities by considering their combinatorial analogues. The idea was originally used by Katona [38] to prove probabilistic inequalities using results from graph theory.

Let X be a random variable taking values in a certain measurable space. Let F and K be two measurable subsets of the k-fold product space. Given a sequence $X_1, \ldots, X_m$ of independent copies of X, the random variable $T_m(X, F)$ is defined to be

$$T_m(X, F) = |\{(i_1, \ldots, i_k) : i_1 \neq \cdots \neq i_k, (X_{i_1}, \ldots, X_{i_k}) \in F\}|.$$

Similarly we can define $T_m(X, K)$. Their combinatorial counterparts can be defined for a deterministic sequence $x_1, \ldots, x_m$. More specifically, we define

$$T_m(F) = |\{(i_1, \ldots, i_k) : i_1 \neq \cdots \neq i_k, (x_{i_1}, \ldots, x_{i_k}) \in F\}|.$$

We define $T_m(K)$ in a similar manner. Clearly, $T_m(F)$ and $T_m(K)$ depend on the selection of the deterministic sequence. We will not emphasize this when it is clear from the context.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Proposition 2.1. *Suppose that there is a function $h_k(m) = o(m^k)$ and an absolute constant $C(F,K)$ such that the inequality*

$$T_m(F) \leqslant h_k(m) + C(F,K) \cdot T_m(K) \tag{2.1}$$

holds for any deterministic sequence $x_1, \ldots, x_m$. Then the inequality

$$\mathbb{P}((X_1, \ldots, X_k) \in F) \leqslant C(F,K) \cdot \mathbb{P}((X_1, \ldots, X_k) \in K) \tag{2.2}$$

holds for any i.i.d. random variables $X_1, \ldots, X_k$.

Proof. Since inequality (2.1) holds for any deterministic sequences, we have

$$T_m(X,F) \leqslant h_k(m) + C(F,K) \cdot T_m(X,K).$$

This implies in particular that

$$\mathbb{E}(T_m(X,F)) \leqslant h_k(m) + C(F,K) \cdot \mathbb{E}(T_m(X,K)). \tag{2.3}$$

Notice that $T_m(X,F)$ can be written as the summation of Bernoulli random variables with the same distribution

$$T_m(X,F) = \sum 1_{\{(X_{i_1},\ldots,X_{i_k}) \in F\}}, \tag{2.4}$$

where the summation is taken over all ordered k-tuples $(i_1, \ldots, i_k)$ with distinct coordinates. Therefore we have

$$\mathbb{E}(T_m(X,F)) = (m)_k \cdot \mathbb{P}((X_1, \ldots, X_k) \in F), \tag{2.5}$$

where the notation $(m)_k$ stands for the product $m(m-1)\cdots(m-k+1)$. Similarly we have

$$\mathbb{E}(T_m(X,K)) = (m)_k \cdot \mathbb{P}((X_1, \ldots, X_k) \in K). \tag{2.6}$$

Combining (2.3), (2.5) and (2.6), we have

$$\mathbb{P}((X_1, \ldots, X_k) \in F) \leqslant \frac{h_k(m)}{(m)_k} + C(F,K) \cdot \mathbb{P}((X_1, \ldots, X_k) \in K).$$

Since m^k and $(m)_k$ have the same order for fixed k and large m, and $h_k(m) = o(m^k)$, the proposition follows by taking the limit $m \to \infty$. □

Although the above proof is very simple, let us demonstrate the heuristic behind this combinatorial argument. We will see that the assumption (2.1) is not artificial and it has to be true if inequality (2.2) holds for all distributions. For simplicity, we let $p = \mathbb{P}((X_1, \ldots, X_k) \in F)$. Using the representation (2.4), we have

$$\begin{aligned}\mathbb{E}\left(\frac{T_m(X,F)}{(m)_k} - p\right)^2 &= \frac{1}{(m)_k^2}\mathbb{E}\Big(\sum(1_{\{(X_{i_1},\ldots,X_{i_k}) \in F\}} - p)\Big)^2 \\ &= \frac{1}{(m)_k^2}\sum \mathbb{E}1_{\{(X_{i_1},\ldots,X_{i_k}) \in F\}} \cdot 1_{\{(X_{i'_1},\ldots,X_{i'_k} \in F\}} - p^2).\end{aligned}$$

If the k-tuples $(i_1, \ldots, i_k)$ and $(i'_1, \ldots, i'_k)$ share no common index, we have

$$\mathbb{E}(1_{\{(X_{i_1},\ldots,X_{i_k}) \in F\}} \cdot 1_{\{(X_{i'_1},\ldots,X_{i'_k}) \in F\}}) = p^2.$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Since there are $o(m^{2k})$ pairs of ordered k-tuples $(i_1,\dots,i_k)$ and $(i'_1,\dots,i'_k)$ with common indices, we have

$$\mathbb{E}\left(\frac{T_m(X,F)}{(m)_k} - \mathbb{P}((X_1,\dots,X_k)\in F)\right)^2 \longrightarrow 0, \quad \text{as } m\to\infty.$$

In particular, we have

$$\frac{T_m(X,F)}{(m)_k} \xrightarrow{\text{a.s.}} \mathbb{P}((X_1,\dots,X_k)\in F), \quad \text{as } m\to\infty.$$

A similar property holds for $T_m(X,K)$. Combining with inequality (2.2), we have

$$T_m(X,F) \leqslant o(m^k) + C(F,K)\cdot T_m(X,K), \quad \text{a.s.}$$

Hence for almost all realizations of $X_1,\dots,X_m$, *i.e.* deterministic sequences $x_1,\dots,x_m$, we will have

$$T_m(F) \leqslant o(m^k) + C(F,K)\cdot T_m(K).$$

One may notice that the sequences violating the above inequality depend on the distribution of X and the $o(m^k)$ term may depend on the sequences. However, if inequality (2.2) holds for all distributions, it should be reasonable to expect the validity of (2.1) for all deterministic sequences.

3. Small ball inequalities in abelian groups

Let G denote an abelian topological group equipped with the Borel σ-algebra (*i.e.* the σ-algebra generated by the open sets). Let X and Y be i.i.d. random variables taking values in G. A subset of G is said to be *symmetric* if it contains the group inverse of each element of this set.

For two subsets $F,K\subseteq G$, their Minkowski sum $F+K$ is defined as

$$F+K = \{x+y : x\in F, y\in K\}.$$

We define the difference set $F-K$ in a similar manner. The generalized entropy number $N(F,K)$ is defined to be the maximal number of elements we can select from F such that the difference of any two distinct elements does not belong to K. More precisely, we have

$$N(F,K) = \sup\{|S| : S\subseteq F, (S-S)\cap K\subseteq\{0\}\}. \tag{3.1}$$

Let $T=\{x_1,x_2,\dots,x_m\}$ be a multiset (or sequence) of G, that is, the elements of T are selected from G and are not necessarily distinct. For any $s\in\mathbb{R}$, the quantities $T_+(F,s)$ and $T_-(K,s)$ are defined as

$$T_+(F,s) = ms + |\{(i,j) : i\neq j, x_i+x_j\in F\}| \tag{3.2}$$

and

$$T_-(K,s) = ms + |\{(i,j) : i\neq j, x_i-x_j\in K\}|. \tag{3.3}$$

The relation between these two quantities is given in the following lemma, which is similar in spirit to Lemma 2.1 and Lemma 3.2 in [2].

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Lemma 3.1. *Suppose that K is a symmetric set with $0 \in K$. For $s \geqslant 2$ and any multiset T, we have*

$$T_+(F,s) \leqslant N(F,K) \cdot T_-(K,2s).$$

Proof. If $N(F,K) = \infty$, the above statement is obviously true. So we will assume that $N(F,K)$ is finite and prove the lemma by induction on the cardinality of T. When counting the cardinality of a multiset, every element counts even if two elements have the same value. For the base case $|T| = 1$, we have $T_+(F,s) = s$ and $T_-(K,2s) = 2s$. Since $N(F,K) \geqslant 1$, it is clear that the lemma has to be true. We assume that the lemma holds for any multiset T with cardinality $|T| \leqslant m-1$. Let t be some non-negative integer such that

$$\max_{x \in T} |(x+K) \cap T| = t+1.$$

Here we let $(x+K) \cap T$ denote the multiset consisting of elements of T which lie in $x+K$. We will use similar notations without further clarification. Let $x^* \in T$ be an element that achieves the above maximum, and set $T^* = T \backslash \{x^*\}$, where '$\backslash$' is the standard set subtraction notation. (We only throw x^* away, not other elements with the same value.) Since K is a symmetric set containing 0, we have

$$T_-(K,2s) = T^*_-(K,2s) + 2s + 2t. \tag{3.4}$$

We also have

$$T_+(F,s) \leqslant T^*_+(F,s) + s + 2|(-x^* + F) \cap T|. \tag{3.5}$$

The definition in (3.1) implies that we can select at most $N(F,K)$ elements from $(-x^* + F) \cap T$, say $\{y_1, y_2, \ldots, y_k\}$ with $k \leqslant N(F,K)$, such that their mutual differences are not in K. Therefore we have

$$(-x^* + F) \cap T \subseteq \cup_i (y_i + K) \cap T. \tag{3.6}$$

Combining with (3.5), we have

$$T_+(F,s) \leqslant T^*_+(F,s) + s + 2(t+1)N(F,K). \tag{3.7}$$

By the induction assumption, the lemma holds for T^*. Combining (3.4) and (3.7), it is not hard to check that the lemma holds when

$$s \geqslant \frac{2N(F,K)}{2N(F,K)-1},$$

which is implied by the assumption $s \geqslant 2$. □

By combining Proposition 2.1 with Lemma 3.1, we have the following result.

Theorem 3.2. *Let F and K be measurable subsets of the abelian topological group G. Suppose that K is symmetric and contains the identity of G in its interior. For any i.i.d.*

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

random variables X and Y taking values in G, we have

$$\mathbb{P}(X+Y\in F)\leqslant N(F,K)\cdot\mathbb{P}(X-Y\in K).$$

In the following, we study the comparison of $\mathbb{P}(X-Y\in F)$ and $\mathbb{P}(X-Y\in K)$. The definition of $T_-(F,s)$ is given in (3.3) with K replaced by F. Similar to Lemma 3.1, we have the following result.

Lemma 3.3. *Suppose that K is a symmetric set with $0\in K$. For $s\geqslant 2$ and any multiset T, we have*

$$T_-(F,s)\leqslant(1+N(F\backslash K,K))\cdot T_-(K,2s).$$

Proof. We only need to make a slight modification of the proof of Lemma 3.1. Let x^* be selected in the same way as before and we set $T^*=T\backslash\{x^*\}$. It is clear that equation (3.4) still holds. Instead of (3.5), we have

$$T_-(F,s)\leqslant T_-^*(F,s)+s+|((x^*+F)\cap T)\backslash\{x^*\}|+|((x^*-F)\cap T)\backslash\{x^*\}|.$$

Notice the following set-inclusion relations:

$$(x^*+F)\subseteq(x^*+K)\cup(x^*+F\backslash K)$$

and

$$(x^*-F)\subseteq(x^*+K)\cup(x^*-F\backslash K).$$

We use the symmetry assumption of K in the second inclusion relation. Apply the covering argument (3.6) again to $x^*+F\backslash K$ and $x^*-F\backslash K$. We have

$$T_-(F,s)\leqslant T_-^*(F,s)+s+2t+2(t+1)N(F\backslash K,K). \tag{3.8}$$

Combining (3.4) and (3.8), for $s\geqslant 2$, we will have

$$T_-(F,s)\leqslant(1+N(F\backslash K,K))\cdot T_-(K,2s).$$

Hence we complete the proof of the lemma. □

Combining Proposition 2.1 with Lemma 3.3, we have the following result.

Theorem 3.4. *Let F and K be measurable subsets of the abelian topological group G. Suppose that K is symmetric and contains the identity of G in its interior. For any i.i.d. random variables X and Y taking values in G, we have*

$$\mathbb{P}(X-Y\in F)\leqslant(1+(F\backslash K,K))\cdot\mathbb{P}(X-Y\in K).$$

We include a result of Katona [38] as another application of this combinatorial argument, which is related to Turán's theorem on triangle-free graphs. While we provide a reformulation of Katona's proof for completeness, we do not claim any novelty here.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Theorem 3.5 (Katona [38]). *Let X and Y be i.i.d. random variables taking values in a Hilbert space V with the norm $\|\cdot\|$. Then we have*

$$(\mathbb{P}(\|X\| \geqslant 1))^2 \leqslant 2\mathbb{P}(\|X+Y\| \geqslant 1).$$

Proof. Let $F, K \subseteq V \times V$ be the subsets defined by

$$F = \{(x,y) : \|x\| \geqslant 1, \|y\| \geqslant 1\}$$

and

$$K = \{(x,y) : \|x+y\| \geqslant 1\}.$$

Given a multiset $T = \{x_1, \ldots, x_m\}$ of V, we define

$$T_m(F) = |\{(i,j) : i \neq j, (x_i, x_j) \in F\}|$$

and

$$T_m(K) = |\{(i,j) : i \neq j, (x_i, x_j) \in K\}|.$$

By Proposition 2.1, the statement holds if we can show that

$$T_m(F) \leqslant 2(m + T_m(K)). \tag{3.9}$$

Suppose that there are n elements of T with norms not less than 1. Then we have

$$T_m(F) = n^2 - n. \tag{3.10}$$

Let us consider a simple graph $\mathcal{G}$ on these n elements; note that we think of any two elements as different vertices even if they have the same value. Two vertices x and y are adjacent if and only if $\|x+y\| \geqslant 1$. Then we have

$$T_m(K) \geqslant 2e(\mathcal{G}), \tag{3.11}$$

where $e(\mathcal{G})$ is the number of edges of $\mathcal{G}$. For any 3 vertices x, y, z, there exists at least a pair, say x, y, such that the angle between them is no more than $2\pi/3$, which implies that $\|x+y\| \geqslant 1$. This fact implies that the complementary graph is triangle-free. Using Turán's theorem, we have

$$e(\mathcal{G}) \geqslant \binom{n}{2} - \frac{n^2}{4}. \tag{3.12}$$

Then estimate (3.9) follows from (3.10), (3.11) and (3.12). □

Remark. Katona [38] proved the result for discrete random variables, and then extended it for continuous random variables employing continuous versions of Turán-type theorems. This discretization argument was used in a series of papers [38, 39, 40, 41, 42, 43, 44] on the optimal estimate of $\mathbb{P}(\|X+Y\| \geqslant c)$ in terms of $\mathbb{P}(\|X\| \geqslant 1)$; [45] contains a comprehensive survey. Similar results were independently obtained by Sidorenko [70].

We do not know whether Katona's result holds in general Banach spaces. This motivates us to prove the following symmetrization result.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Theorem 3.6. *Let F and K be measurable subsets of an abelian topological group G. Suppose that K is symmetric and contains the identity of G in its interior. For any G-valued i.i.d. random variables X and Y, we have*

$$(\mathbb{P}(X \in F))^2 \leqslant N(F,K) \cdot \mathbb{P}(X - Y \in K).$$

Proof. The proof is similar to that of Theorem 3.2. We can assume that $N(F,K)$ is finite. For a multiset $T = \{x_1, \ldots, x_m\}$ of G, we define

$$T(F) = |\{(i,j) : i \neq j, x_i \in F, x_j \in F\}|$$

and

$$T_-(K) = (N(F,K))^2 + 2m + |\{(i,j) : i \neq j, x_i - x_j \in K\}|.$$

Then we need to prove the following combinatorial counterpart of the statement:

$$T(F) \leqslant N(F,K) \cdot T_-(K). \tag{3.13}$$

We will prove it by induction on m. It is clear that the statement holds for $m \leqslant N(F,K)$. Without loss of generality, we can assume that $x_i \in F$ for all $x_i \in T$. Let t, x^*, T^* be defined in the same way as in Lemma 3.1. By the definition of $N(F,K)$ in (3.1), we can select at most $N(F,K)$ elements of T with mutual differences not contained by K. The pigeonhole principle implies that

$$t + 1 \geqslant \frac{m}{N(F,K)}.$$

One can check that

$$T(F) = m^2 - m = T^*(F) + 2(m-1) \tag{3.14}$$

and

$$T_-(K) = T^*_-(K) + 2t + 2 \geqslant T^*_-(K) + \frac{2m}{N(F,K)}. \tag{3.15}$$

Estimate (3.13) follows from (3.14), (3.15) and the induction assumption on T^*. □

Corollary 3.7. *Let $a, b > 0$. For any i.i.d. real-valued random variables X and Y, we have*

$$(\mathbb{P}(|X| \leqslant b))^2 \leqslant \lceil 2b/a \rceil \cdot \mathbb{P}(|X - Y| \leqslant a).$$

Moreover the estimate cannot be improved.

The result follows from Theorem 3.6 with $F = [-b, b]$ and $K = [-a, a]$. The estimate is sharp for the uniform distribution over $\{-b, -b + (1+\epsilon)a, \ldots, -b + (\lceil 2b/a \rceil - 1)(1+\epsilon)a\}$ for some small $\epsilon > 0$.

To conclude this section, we ask the question of whether there is an analogue of Proposition 2.1 when the probabilities in comparison are of different magnitudes. For

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

example, we do not know a combinatorial counterpart of the distribution-free inequality

$$\mathbb{P}((X_1,\dots,X_k)\in F)\leqslant f(\mathbb{P}(X_1,\dots,X_l)\in K),$$

where f is a polynomial.

4. Small ball inequalities in non-abelian groups

In this section, we let G denote an arbitrary topological group with the identity e. We will show that Theorem 3.2 and Theorem 3.4 still hold for certain measurable sets F and K in this general setting. Similar to the sumset $F+K$ in the abelian case, we define the product set $F\cdot K$ in this non-abelian setting as

$$F\cdot K=\{xy : x\in F, y\in K\}.$$

The entropy number $N(F,K)$ is redefined as

$$N(F,K)=\sup\{|S| : S\subseteq F, (S\cdot S^{-1})\cap K\subseteq\{e\}\}, \tag{4.1}$$

where S^{-1} is the set of all inverses of the elements of S. For $s\in\mathbb{R}$ and a multiset $T=\{x_1,\dots,x_m\}$, the quantities $T_+(F,s)$ and $T_-(K,s)$ are redefined as

$$T_+(F,s)=ms+|\{(i,j) : i\neq j, x_ix_j\in F\}| \tag{4.2}$$

and

$$T_-(K,s)=ms+|\{(i,j) : i\neq j, x_ix_j^{-1}\in K\}|. \tag{4.3}$$

Similar to Lemma 3.1, we have the following result.

Lemma 4.1. *Suppose K is a normal subgroup of G. For $s\geqslant 2$ and any multiset T, we have*

$$T_+(F,s)\leqslant N(F,K)\cdot T_-(K,2s).$$

Proof. The lemma can be proved with a slight modification of the proof of Lemma 3.1. In order to see how the assumption of K is used, we write the proof again. Let t be some non-negative integer such that

$$\max_{x\in T}|(xK)\cap T|=t+1,$$

where xK is the set of the products of x and the elements of K. Let x^* be an element such that the maximum can achieved and $T^*=T\backslash\{x^*\}$. By the definition of $T_-(K,2s)$, we have

$$T_-(K,2s)=T_-^*(K,2s)+2s+|((Kx^*)\cap T)\backslash\{x^*\}|+|((x^*K^{-1})\cap T)\backslash\{x^*\}|.$$

Since K is a normal subgroup, the estimate of $T_-(K,2s)$ in (3.4) still holds. Similar to (3.5), we have

$$T_+(F,s)\leqslant T_+^*(F,s)+s+|((x^*)^{-1}F)\cap T|+|(F(x^*)^{-1})\cap T|.$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Let $\alpha_1, \alpha_2 \in F$ be any two elements, and $u_1 = (x^*)^{-1}\alpha_1$, $u_2 = (x^*)^{-1}\alpha_2$. Since K is a normal subgroup, we can see that $u_1 u_2^{-1} \in K$ if only if $\alpha_1 \alpha_2^{-1} \in K$. (The assumption that K is a normal subgroup is important here.) By the definition of $N(F,K)$ in (4.1), we can select at most $N(F,K)$ elements from $((x^*)^{-1}F)) \cap T$, say $\{y_1, \ldots, y_k\}$, such that $y_i y_j^{-1} \notin K$ for any $y_i \neq y_j$. Then we have the covering relation

$$((x^*)^{-1}F) \cap T \subseteq \cup_i (y_i K) \cap T,$$

which implies that

$$|((x^*)^{-1}F) \cap T| \leqslant (t+1)N(F,K).$$

Similarly we have the same estimate for $|(F(x^*)^{-1}) \cap T|$. Then the estimate of $T_+(F,s)$ is exactly the same as (3.7). So we have proved the lemma. □

Using Proposition 2.1 together with Lemma 4.1, we have the following result.

Theorem 4.2. *Suppose that K is a normal subgroup of the topological group G and the interior of K contains the identity. For any i.i.d. random variables X and Y taking values in G, and any measurable subset F of G, we have*

$$\mathbb{P}(XY \in F) \leqslant N(F,K) \cdot \mathbb{P}(XY^{-1} \in K).$$

We omit the proof of the following lemma, since it is essentially the same as that of Lemma 3.3.

Lemma 4.3. *Suppose that $K \subseteq G$ is a normal subgroup. For $s \geqslant 2$ and any multiset T, we have*

$$T_-(F,s) \leqslant (1 + N(F \backslash K, K)) \cdot T_-(K, 2s).$$

Combining with Proposition 2.1, we have the following result.

Theorem 4.4. *Suppose that K is a normal subgroup of the topological group G and the interior of K contains the identity. For any i.i.d. random variables X and Y taking values in G, and any measurable subset F of G, we have*

$$\mathbb{P}(XY^{-1} \in F) \leqslant (1 + N(F \backslash K, K)) \cdot \mathbb{P}(XY^{-1} \in K).$$

5. Small ball inequalities in vector spaces

Let V be a topological vector space over a field $\mathbb{F}$ with the Borel σ-algebra generated by all open sets. Let $F, K \subseteq V$ be measurable subsets and let $a, b \in \mathbb{F}$. We consider the comparison of $\mathbb{P}(aX + bY \in F)$ and $\mathbb{P}(X - Y \in K)$ for i.i.d. random variables X and Y taking values in V.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Theorem 5.1. *Let F and K be measurable subsets of a topological vector space V over a field $\mathbb{F}$. Suppose that K is symmetric and contains the zero vector in its interior. Let a and b be non-zero elements of $\mathbb{F}$. For any i.i.d. random variables X and Y taking values in V, we have*

$$\mathbb{P}(aX + bY \in F) \leqslant N(a, b, F, K) \cdot \mathbb{P}(X - Y \in K),$$

where the constant $N(a, b, F, K)$ is defined as

$$N(a, b, F, K) = \frac{1}{2}(N(a^{-1}F, K) + N(b^{-1}F, K)).$$

Proof. The proof is essentially the same as that of Theorem 3.2. Let $T = \{x_1, \ldots, x_m\}$ be a multiset of V. For $s \in \mathbb{R}$, we define

$$T_+(F, s, a, b) = ms + |\{(i, j) : i \neq j, ax_i + bx_j \in F\}|.$$

By Proposition 2.1, it suffices to prove the combinatorial analogue

$$T_+(F, s, a, b) \leqslant N(a, b, F, K) \cdot T_-(K, 2s) \tag{5.1}$$

for $s \geqslant 2$, where $T_-(K, 2s)$ is defined to be the same as (3.3). We select x^* in the same way as in Lemma 3.1 and set $T^* = T \setminus \{x^*\}$. The estimate of $T_-(K, 2s)$ in (3.4) still holds. Similar to (3.5), we have

$$T_+(F, s, a, b) \leqslant T_+^*(F, s, a, b) + s + |(-b^{-1}ax^* + b^{-1}F) \cap T| + |(-a^{-1}bx^* + a^{-1}F) \cap T|.$$

Applying the covering argument (3.6) to $-b^{-1}ax^* + b^{-1}F$ and using the definition of $N(b^{-1}F, K)$, we will have

$$|(-b^{-1}ax^* + b^{-1}F) \cap T| \leqslant (t + 1)N(b^{-1}F, K).$$

Similarly we can see that

$$|(-a^{-1}bx^* + a^{-1}F) \cap T| \leqslant (t + 1)N(a^{-1}F, K).$$

Hence we have

$$T_+(F, s, a, b) \leqslant T_+^*(F, s, a, b) + s + 2(t + 1)N(a, b, F, K). \tag{5.2}$$

Then estimate (5.1) follows from (3.4) and (5.2). So we complete the proof. □

As an easy consequence of Theorem 5.1, we have the following extension of (1.2) and (1.3).

Corollary 5.2. *Let a, b, c, d be non-zero reals and $c, d > 0$. For any real-valued i.i.d. random variables X and Y, we have*

$$\mathbb{P}(|aX + bY| \leqslant c) \leqslant \frac{1}{2}\left(\left\lceil \frac{2c}{|a|d} \right\rceil + \left\lceil \frac{2c}{|b|d} \right\rceil\right)\mathbb{P}(|X - Y| \leqslant d).$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Proof. We can take $F = [-c, c]$ and $K = [-d, d]$. Elementary geometric argument will yield

$$N(a^{-1}F, K) = \left\lceil \frac{2c}{|a|d} \right\rceil, \quad N(b^{-1}F, K) = \left\lceil \frac{2c}{|b|d} \right\rceil.$$

Then the statement follows from Theorem 5.1. □

Remark. For $a = b$, Theorem 5.1 and Corollary 5.2 match Theorem 3.2 and (1.3), respectively. For $a = -b$, Theorem 5.1 and Corollary 5.2 are weaker than Theorem 3.4 and (1.2), respectively. This is due to the subtle difference between the covering arguments used in the proofs. The case when $a = 0$ or $b = 0$ is covered by Corollary 3.7.

6. Discussion of tightness

In this section, we study the near-extremal distributions for the probabilistic estimates developed in Section 3. The discussion will mainly focus on Theorem 3.2 and Theorem 3.4 for random variables taking values in the Euclidean space $\mathbb{R}^d$. We will see their close connections with the sphere packing problem in combinatorial geometry.

In general it is hard to compute the ratios of $\mathbb{P}(X \pm Y) \in F$ to $\mathbb{P}(X - Y) \in K$. If X and Y are assumed to be uniform over a finite set $T = \{x_1, x_2, \ldots, x_n\}$, then we have

$$\mathbb{P}(X \pm Y \in F) = \frac{1}{n^2} \sum_{i=1}^{n} |(\mp x_i + F) \cap T|$$

and

$$\mathbb{P}(X - Y \in K) = \frac{1}{n^2} \sum_{i=1}^{n} |(x_i + K) \cap T|.$$

If the set T is K-separated, *i.e.* $x_i - x_j \notin K$ for $i \neq j$, we will have $|(x_i + K) \cap T| = 1$ for all $x_i \in T$. We can even make a further assumption that, except for $o(n)$ of them, all the sets $\mp x_i + F$ contain the same number of elements of T. This is possible if T is selected to consist of certain lattice points. (So we need X and Y to be in a topological vector space V.) Under these assumptions, we have

$$\frac{\mathbb{P}(X \pm Y \in F)}{\mathbb{P}(X - Y \in K)} \to \max_{x \in T} |(\mp x + F) \cap T|, \quad \text{as } n \to \infty.$$

Theorem 3.2 will be tight if there exists a K-separated lattice $\mathcal{L}$ and a point $x \in V$ (not necessarily a lattice point) such that $x + F$ contains $N(F, K)$ points of $\mathcal{L}$. The set T can be taken as the union of a subset of $\mathcal{L}$ and the reflection of this subset after a certain translation. Similarly, Theorem 3.4 is tight if for every lattice point $x \in \mathcal{L}$ the set $x + (F \backslash K)$ contains $N(F \backslash K, K)$ points of $\mathcal{L}$. In this case, we only need to take T to be a certain subset of the lattice $\mathcal{L}$. This idea can be used to produce near-optimal distributions for estimates (1.3) and (1.2). For estimate (1.3), we can take X to be uniform over $\{-(n-1)\delta - r, \ldots, -\delta - r, \delta, 2\delta, \ldots, n\delta\}$, where $r > 0$ and $\delta > a$. For any $a, b > 0$, we can always select appropriate r and δ such that the ratio of $\mathbb{P}(|X + Y| \leqslant b)$ to $\mathbb{P}(|X - Y| \leqslant a)$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

approaches $\lceil 2b/a \rceil$ as $n \to \infty$. This example is essentially the same as the one given in [21]. To see the sharpness of (1.2), we can take X to be uniform over $\{\delta, 2\delta, \ldots, n\delta\}$ for a certain $\delta > a$. This example was given in [2].

In the Euclidean space $\mathbb{R}^d$, let us take F and K to be closed balls centred at the origin of radius r and 1, respectively. For simplicity, we let $N_+(r)$ and $N_-(r)$ denote $N(F,K)$ and $N(F\backslash K, K) + 1$, respectively. Then $N_+(r)$ represents the maximal number of points in a Euclidean ball of radius r with all mutual distances greater than 1. We have an extra restriction on $N_-(r)$ that one of these points should be at the centre of the ball. These are the so-called sphere packing problems. The dual problem of $N_+(r)$ asks for the smallest radius $r_+(n)$ of the ball to contain n points with mutual distances at least 1. We define $r_-(n)$ in a similar way with the restriction that one of the points should be at the centre of the ball.

Bateman and Erdős [4] studied the diameters of the extremal configurations of points on the plane. Their results imply the values of $r_+(n)$ and, by the duality, $N_+(r)$ in $\mathbb{R}^2$. We have the following list of $N_+(r)$ for r in a certain range:

$$N_+(r) = \begin{cases} 1 & \text{if } 0 < r \leqslant 1/2, \\ 2 & \text{if } 1/2 < r \leqslant \sqrt{3}/3, \\ 3 & \text{if } \sqrt{3}/3 < r \leqslant \sqrt{2}/2, \\ 4 & \text{if } \sqrt{2}/2 < r \leqslant \frac{1}{2}\csc(\pi/5), \\ 5 & \text{if } \frac{1}{2}\csc(\pi/5) < r \leqslant 1, \\ 7 & \text{if } 1 < r \leqslant 1+\epsilon, \text{ small } \epsilon > 0. \end{cases}$$

The extremal configurations given by Bateman and Erdős are lattice points. Hence the values of $N_+(r)$ in the list are tight. It is easy to see that $r_-(2) = \cdots = r_-(7) = 1$ with all the points on the unit circle except one point at the centre. Bateman and Erdős also gave the values of $r_-(n)$ for $n = 8, 9, 10, 11$. This yields the following list of $N_-(r)$:

$$N_-(r) = \begin{cases} 7 & \text{if } 1 < r \leqslant \frac{1}{2}\csc(\pi/7), \\ 8 & \text{if } \frac{1}{2}\csc(\pi/7) < r \leqslant \frac{1}{2}\csc(\pi/8), \\ 9 & \text{if } \frac{1}{2}\csc(\pi/8) < r \leqslant \frac{1}{2}\csc(\pi/9), \\ 10 & \text{if } \frac{1}{2}\csc(\pi/9) < r \leqslant \frac{1}{2}\csc(\pi/10). \end{cases}$$

For the sphere packing problems in $\mathbb{R}^d$, people are generally interested in the optimal packing density ρ_d. When r is large, it is not hard to see the following asymptotic estimate:

$$N_+(r) \approx N_-(r) = (1+o(1))\frac{\mathrm{vol}(B(r))}{\mathrm{vol}(B(1/2))}\rho_d = (1+o(1))(2r)^d\rho_d.$$

In $\mathbb{R}^2$, it is known that hexagonal lattice packing is optimal among all packings (not necessarily lattice packings) with packing density $\rho_2 = \sqrt{3}\pi/6 \approx 0.9069$. The sphere packing problem in three-dimensional Euclidean space has a long history. Kepler conjectured in 1611 that no arrangement of equally sized spheres can fill the space with a greater average density than that of the face-centred cubic and hexagonal close packing arrangements.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

The density of these arrangements is $\sqrt{2}\pi/6 \approx 0.7404$. Gauss proved in 1831 that Kepler's conjecture is true if the spheres have to be arranged in a regular lattice. The complete proof of Kepler's conjecture was announced only around 20 years ago by Hales (see *e.g.* [33]), and a formal verification using automated proof checking was only completed in 2014. In the very recent breakthrough work [72], Viazovska proved that the E_8 lattice packing gives the optimal packing density in dimension 8, and the density is $\rho_8 = \pi^4/384 \approx 0.025367$. Building on Viazovska's work, it is shown in [16] that Leech lattice packing is optimal in 24 dimensions, and the packing density is $\rho_{24} = \pi^{12}/12! \approx 0.00193$.

Another interesting problem related to our study is the kissing number problem. In three dimensions it asks how many billiard balls can be arranged so that they all just touch another billiard ball of the same size. This question was a subject of a famous debate between Isaac Newton and David Gregory in the 1690s. Newton believed the answer was 12, while Gregory thought that 13 might be possible. Generally we can define the d-dimensional kissing number τ_d as the maximal number of points on the unit sphere with Euclidean distances at least 1. For $1 < r < 1 + \epsilon_d$ with small $\epsilon_d > 0$, we have the equation

$$N_-(r) = \tau_d + 1. \tag{6.1}$$

The number $\tau_3 = 12$ was studied by various researchers in the nineteenth century. The best proof now available is due to Leech [50]. The answers $\tau_8 = 240$ and $\tau_{24} = 196\,560$ are given by [66] and [51], respectively. It is somewhat surprising that they are technically easier to establish than τ_3. The correct answer $\tau_4 = 24$ was obtained much later by Musin [64]. For all these results, the extremal configurations follows from lattice packings. Using the relation (6.1), Theorem 3.4 can give explicit tight estimates for r slightly greater than 1 in corresponding dimensions. These are all the known values of the kissing number so far. In high dimensions, τ_d grows exponentially with unknown base. We refer to the monograph [17] for more discussions of sphere packing problems and their relations with number theory and coding theory.

7. Applications

7.1. Concentration functions

Let G be an abelian topological group and let $F \subseteq G$ be a measurable subset. For a random variable X taking values in G, the generalized Lévy's concentration function $Q(X, F)$ is defined to be

$$Q(X, F) = \sup_{x \in G} \mathbb{P}(X \in x + F). \tag{7.1}$$

The main study of concentration functions is devoted to the sum of independent random variables in Banach spaces with F taken to be norm balls (see *e.g.* [34, 36, 63]). In the i.i.d. case, Theorem 3.2 provides a symmetrization technique to treat different sets and also general groups where no norm may exist.

For a random variable X, we let $\widetilde{X}$ denote the symmetrized random variable $X - Y$, where Y is an independent copy of X.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Theorem 7.1. *Let G be an abelian topological group, and let K be a symmetric measurable subset of G such that the interior of K contains the identity of G. For any independent (not necessarily identical) random variables $X_1, \dots, X_n$ taking values in G, and any measurable subset F of G, we have*

$$(Q(X_1 + \cdots + X_n, F))^2 \leqslant N(F + F, K) \cdot Q(\widetilde{X}_1 + \cdots + \widetilde{X}_n, K). \tag{7.2}$$

If $X_1, \dots, X_n$ also have identical distribution, then we have

$$Q(X_1 + \cdots + X_n, F) \leqslant N(F, K) \cdot Q(\widetilde{X}_1 + \cdots + \widetilde{X}_{\lfloor n/2 \rfloor}, K). \tag{7.3}$$

Proof. By the definition of generalized Lévy's concentration function (7.1), for any $0 < \epsilon < 1$, there exists $x \in G$ such that

$$\begin{aligned}
(Q(X_1 + \cdots + X_n, F))^2 &\leqslant (\mathbb{P}(X_1 + \cdots + X_n \in x + F) + \epsilon)^2 \\
&\leqslant \mathbb{P}(X_1 + \cdots + X_n \in x + F) \cdot \mathbb{P}(X_1' + \cdots + X_n' \in x + F) + 3\epsilon \\
&\leqslant \mathbb{P}(X_1 + \cdots + X_n + X_1' \cdots + X_n' \in 2x + F + F) + 3\epsilon \\
&\leqslant N(F + F, K) \cdot \mathbb{P}(X_1 - X_1' + \cdots + X_n - X_n', K) + 3\epsilon \\
&\leqslant N(F + F, K) \cdot Q(\widetilde{X}_1 + \cdots + \widetilde{X}_n, K) + 3\epsilon,
\end{aligned}$$

where $(X_1', \dots, X_n')$ is an independent copy of $(X_1, \dots, X_n)$, and the second-to-last inequality follows from Theorem 3.2. Then (7.2) follows by letting $\epsilon \to 0$. To prove (7.3), it is not hard to see that, for independent random variables X and Y,

$$Q(X + Y, F) \leqslant Q(X, F).$$

Therefore we have

$$Q(X_1 + \cdots + X_n, F) \leqslant Q(X_1 + \cdots + X_{2\lfloor n/2 \rfloor}, F).$$

Similarly we can select $x \in G$ such that

$$\begin{aligned}
Q(X_1 + \cdots + X_{2\lfloor n/2 \rfloor}, F) &\leqslant \mathbb{P}(X_1 + \cdots + X_{2\lfloor n/2 \rfloor} \in x + F) + \epsilon \\
&\leqslant N(F, K) \cdot \mathbb{P}(X_1 - X_2 + \cdots + X_{2\lfloor n/2 \rfloor - 1} - X_{2\lfloor n/2 \rfloor} \in K) + \epsilon \\
&\leqslant N(F, K) \cdot Q(\widetilde{X}_1 + \cdots + \widetilde{X}_{\lfloor n/2 \rfloor}, K) + \epsilon.
\end{aligned}$$

In the second inequality we use Theorem 3.2, and (7.3) follows by letting $\epsilon \to 0$. □

If F and K are taken to be balls of vanishing radius in a metrizable, locally compact abelian group, then Theorem 7.1 yields an inequality relating the essential suprema of the densities of $X + Y$ and $X - Y$. This may be interpreted as relating the Rényi entropies of order ∞ of $X + Y$ and $X - Y$.

Let X be a random variable taking values in a locally compact abelian group G. Suppose it has density f with respect to the invariant Haar measure μ. For $p \in (0, 1) \cup (1, \infty)$, the Rényi entropy of order p (*i.e.* p-Rényi entropy) is defined as

$$h_p(X) = \frac{1}{1 - p} \log \int_G f(x)^p \mu(dx).$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Defining by continuity, $h_1(X)$ corresponds to the classical Shannon entropy. By taking limits, we have

$$h_0(X) = \log \mu(\mathrm{supp}(f))$$

and

$$h_\infty(X) = -\log \|f\|_\infty,$$

where $\mathrm{supp}(f)$ is the support of f, and $\|f\|_\infty$ is the essential supremum of f.

We need the following definition before introducing our statement. Let f be a real- or complex-valued locally integrable function over a metric space with a certain reference measure μ. We say that x is a Lebesgue point of f if

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f(y)\mu(dy) = f(x),$$

where $B(x,r)$ is the ball of radius r centred at x.

Theorem 7.2. *Let G be a metrizable, locally compact abelian group with the Haar measure μ. Suppose that the following two conditions hold:*

(1) *For any locally integrable function, almost every point of G is a Lebesgue point.*
(2) *We have $\limsup_{r\to 0} N(B_r, B_r) < \infty$, where B_r is the ball of radius r centred at the identity of G.*

For any i.i.d. absolute continuous random variables X and Y in G, we have

$$h_\infty(X+Y) - h_\infty(X-Y) \geqslant -\log\left(\limsup_{r\to 0} N(B_r, B_r)\right).$$

Proof. It is clear that $X+Y$ and $X-Y$ are also absolute continuous random variables. We let f_{X+Y} and f_{X-Y} denote the densities of $X+Y$ and $X-Y$, respectively, with respect to the Haar measure μ. We let $\|f_{X+Y}\|_\infty$ and $\|f_{X-Y}\|_\infty$ denote the essential suprema of f_{X+Y} and f_{X-Y}, respectively. We consider the following two possible cases.

Case 1: $\|f_{X+Y}\|_\infty < \infty$**.** For any $\epsilon > 0$, we can find a Lebesgue point of f_{X+Y}, say $z \in G$, such that

$$f_{X+Y}(z) \geqslant \|f_{X+Y}\|_\infty - \epsilon.$$

The existence of such a point is due to the first assumption. Since z is a Lebesgue point, when $r > 0$ is small, we have

$$\begin{aligned}\int_{B(z,r)} f_{X+Y}(w)\mu(dw) &\geqslant (f_{X+Y}(z) - \epsilon)\mu(B(z,r)) \\ &\geqslant (\|f_{X+Y}\|_\infty - 2\epsilon)\mu(B_r).\end{aligned}$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

In the second inequality we used the invariant property of Haar measure: $\mu(B(z,r)) = \mu(B_r)$. Notice that

$$\begin{aligned}\int_{B(z,r)} f_{X+Y}(w)d\mu(w) &\leqslant Q(X+Y, B_r)\\ &\leqslant N(B_r, B_r)Q(X-Y, B_r)\\ &\leqslant N(B_r, B_r)\|f_{X-Y}\|_\infty \mu(B_r).\end{aligned}$$

The second inequality follows from Theorem 7.1. So we have

$$\|f_{X+Y}\|_\infty - 2\epsilon \leqslant N(B_r, B_r)\|f_{X-Y}\|_\infty.$$

As $\epsilon \to 0$, we will have

$$\|f_{X+Y}\|_\infty \leqslant \limsup_{r\to 0} N(B_r, B_r) \cdot \|f_{X-Y}\|_\infty,$$

which is equivalent to the desired result.

Case 2: $\|f_{X+Y}\|_\infty = \infty$**.** Similarly, for any $M > 0$, we can find a Lebesgue point of f_{X+Y}, say $z \in G$, such that

$$f_{X+Y}(z) \geqslant M.$$

For any $\epsilon > 0$, the same argument implies that for small $r > 0$ we have

$$M - \epsilon \leqslant N(B_r, B_r)\|f_{X-Y}\|_\infty.$$

It is clear that $N(B_r, B_r) \geqslant 1$. Since M can be arbitrarily large, we have

$$\|f_{X-Y}\|_\infty = \infty.$$

In this case the statement trivially holds. □

Remark. For the Euclidean space $\mathbb{R}^d$, the first assumption follows from Lebesgue's differentiation theorem and the second assumption trivially holds since $N(B_r, B_r)$ is independent of the radius r. For $d = 1$ we know that $N(B_r, B_r) = 2$. Therefore for real-valued i.i.d. random variables X and Y we have

$$h_\infty(X+Y) - h_\infty(X-Y) \geqslant -\log 2.$$

This is somewhat surprising, since such a result does not hold for Rényi entropies of order 0 and 1. Indeed, the comparison of $h_p(X+Y)$ and $h_p(X-Y)$ for $p = 0, 1$ is closely related to the study of more-sums-than-differences sets in additive combinatorics. This is discussed in [1], where it is also shown that $h_1(X+Y) - h_1(X-Y)$ can be made to take any real value if X and Y are identically distributed on the real line.

If G is a topological vector space, we can apply Theorem 5.1 to the study of concentration functions of linear combinations of i.i.d. random variables.

Corollary 7.3. *Let F and K be measurable subsets of a topological vector space V over a field $\mathbb{F}$. Suppose that K is symmetric and contains the zero vector in its interior. Let a and*

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

b be non-zero elements of $\mathbb{F}$. For any i.i.d. random variables X and Y taking values in V, we have

$$Q(aX + bY, F) \leqslant N(a, b, F, K) \cdot Q(X - Y, K).$$

Proof. By the definition of generalized Lévy's concentration function (7.1), for any $\epsilon > 0$, there exists $x \in V$ such that

$$\begin{aligned} Q(aX + bY, F) &< \mathbb{P}(aX + bY \in x + F) + \epsilon \\ &\leqslant N(a, b, F, K) \cdot \mathbb{P}(X - Y \in K) + \epsilon \\ &\leqslant N(a, b, F, K) \cdot Q(X - Y, K) + \epsilon. \end{aligned}$$

In the second inequality we use Theorem 5.1. Since $\epsilon > 0$ is arbitrary, the statement follows by letting $\epsilon \to 0$. □

Remark. It is natural to consider possible extensions of the result for more than two i.i.d. random variables. For instance, it is an interesting question to study the variation of (7.3) with the left-hand side replaced by $Q(a_1X_1 + \cdots + a_nX_n, F)$, when $X_1, \ldots, X_n$ take values in a vector space and $a_1, \ldots, a_n$ are arbitrary coefficients in a field. Such questions are related to the Littlewood–Offord phenomenon (see *e.g.* [59, 65]). The simplest manifestation of the latter is the fact that when X_i are Bernoulli random variables taking values 0 and 1 with equal probability, and $a_1, \ldots, a_n$ are positive integers, then

$$Q(a_1X_1 + \cdots + a_nX_n, 1/2) \leqslant Q(X_1 + \cdots + X_n, 1/2) = O(n^{-1/2}).$$

Moreover, when the sequence of coefficients is not allowed to have additive structure, the asymptotic behaviour of the concentration function changes: for example, if the coefficients are forced to all be distinct, the maximal probability decays like $O(n^{-3/2})$. While there is a long history of work in this direction, as surveyed for example in [29, 65], it would be very interesting to understand these questions better in general settings, particularly by identifying extremal coefficients, which are known in only a very few cases.

7.2. Moment inequalities

Unless otherwise specified, we let V denote a topological vector space over the field $\mathbb{F}$, which may be either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. Let $\|\cdot\|$ and $|||\cdot|||$ be two equivalent norms on V. We let I denote the identity operator from $(V, \|\cdot\|)$ to $(V, |||\cdot|||)$. By convention its norm $|||I|||$ is defined as

$$|||I||| = \sup_{\|x\|=1} |||x|||.$$

Let $B_1(r)$, $B_2(r)$ be the closed balls centred at the origin of radius r under the gauges $\|\cdot\|$, $|||\cdot|||$, respectively. It is clear that $|||I|||$ has the following geometric interpretation:

$$|||I||| = \inf\{r > 0 : B_1(1) \subseteq B_2(r)\}. \tag{7.4}$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Theorem 7.4. *Let $a, b \in \mathbb{F}$ be non-zero numbers, and let $q \geqslant p$ be real numbers such that $pq > 0$. For any i.i.d. random variables X and Y taking values in V, we have*

$$(\mathbb{E}|||X - Y|||^p)^{1/p} \leqslant 2|||I||| \max\{|a|^{-1}, |b|^{-1}\} \cdot (\mathbb{E}\|aX + bY\|^q)^{1/q}. \tag{7.5}$$

Proof. We assume that the right-hand side of (7.5) is finite. Otherwise the theorem yields a trivial result. For $0 < p < q$ and $p < q < 0$ (equivalently $p < q$ and $pq > 0$), the following Hölder's inequality holds:

$$(\mathbb{E}\|aX + bY\|^p)^{1/p} \leqslant (\mathbb{E}\|aX + bY\|^q)^{1/q}.$$

Therefore the theorem follows from the case $q = p$. For $p > 0$ we have

$$\begin{aligned}\mathbb{E}\|aX + bY\|^p &= p\int_0^\infty t^{p-1}\mathbb{P}(\|aX + bY\| > t)dt \\ &= p\int_0^\infty t^{p-1}(1 - \mathbb{P}(\|aX + bY\| \leqslant t))dt.\end{aligned}$$

By the geometric interpretation (7.4) of $\|I\|$ and Theorem 5.1, we have

$$\mathbb{P}(\|aX + bY\| \leqslant t) \leqslant \mathbb{P}(|||aX + bY||| \leqslant t|||I|||) \leqslant \mathbb{P}(|||X - Y||| \leqslant Ct),$$

where the constant $C = 2|||I||| \max\{|a|^{-1}, |b|^{-1}\}$. Hence we have

$$\begin{aligned}\mathbb{E}\|aX + bY\|^p &\geqslant p\int_0^\infty t^{p-1}(1 - \mathbb{P}(|||X - Y||| \leqslant Ct))dt \\ &= p\int_0^\infty t^{p-1}\mathbb{P}(|||X - Y||| > Ct)dt \\ &= C^{-p} \cdot \mathbb{E}|||X - Y|||^p,\end{aligned}$$

which is equivalent to the desired statement. For $p < 0$, we have

$$\begin{aligned}\mathbb{E}\|aX + bY\|^p &= -p\int_0^\infty t^{-p-1}\mathbb{P}(\|aX + bY\| \leqslant t^{-1})dt \\ &\leqslant -p\int_0^\infty t^{-p-1}\mathbb{P}(|||aX + bY||| \leqslant |||I|||t^{-1})dt \\ &\leqslant -p\int_0^\infty t^{-p-1}\mathbb{P}(|||X - Y||| \leqslant Ct^{-1})dt \\ &= C^{-p}\left(-p\int_0^\infty t^{-p-1}\mathbb{P}(|||X - Y||| \leqslant t^{-1})dt\right) \\ &= C^{-p} \cdot \mathbb{E}|||X - Y|||^p.\end{aligned}$$

The first inequality follows from the geometric interpretation of $|||I|||$. We apply Theorem 5.1 in the second inequality. The statement follows by taking the $(1/p)$th root. □

As an immediate consequence of Theorem 7.4, we have the following result.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Corollary 7.5. *Let $(E, \|\cdot\|)$ denote a Banach space. For $\gamma > 0$ and any pair of E-valued i.i.d. random variables X and Y, we have*

$$\mathbb{E}\|X - Y\|^{\gamma} \leqslant 2^{\gamma}\mathbb{E}\|X + Y\|^{\gamma} \tag{7.6}$$

and

$$\mathbb{E}\|X + Y\|^{-\gamma} \leqslant 2^{\gamma}\mathbb{E}\|X - Y\|^{-\gamma}.$$

The positive moment case of Corollary 7.5 should be compared with a result of Buja, Logan, Reeds and Shepp [12], who showed that if p and γ are positive reals with $0 < \gamma \leqslant p$ and $1 \leqslant p \leqslant 2$, then

$$\mathbb{E}\|X - Y\|_p^{\gamma} \leqslant \mathbb{E}\|X + Y\|_p^{\gamma} \tag{7.7}$$

holds for any i.i.d. random variables X and Y in $\mathbb{R}^d$, where $\|\cdot\|_p$ refers to the usual ℓ_p-norm on $\mathbb{R}^d$. The negative moment case of Corollary 7.5 should be compared with the recent result for the Euclidean norm of Gao [28], who showed that for $0 < \gamma < d$,

$$\mathbb{E}\|X + Y\|_2^{-\gamma} \leqslant \mathbb{E}\|X - Y\|_2^{-\gamma} \tag{7.8}$$

holds for any i.i.d. random variables X and Y in $\mathbb{R}^d$. Corollary 7.5 has the advantage that it works for any norm and in infinite dimensions, but the inequalities (7.7) and (7.8) have the advantage that when valid, they work with better constants.

Related results were also proved by Mattner [62]. For $0 < \gamma \leqslant 2$, Mattner [62] proved that if T is any orthogonal linear transformation on $\mathbb{R}^d$, then

$$\mathbb{E}\|X - Y\|_2^{\gamma} \leqslant \mathbb{E}\|X - TY\|_2^{\gamma},$$

where $\|\cdot\|_2$ is the Euclidean norm.

The question of the sharpness of Corollary 7.5 is interesting and open. For $p > 2$, Buja, Logan, Reeds and Shepp [12] constructed i.i.d. random variables X and Y in $\mathbb{R}^d$ such that

$$\mathbb{E}\|X - Y\|_p > \mathbb{E}\|X + Y\|_p,$$

but the ratio their construction gives is just a little bit greater than 1, and so does not reveal whether estimate (7.6) is tight when $\gamma = 1$.

One can also consider a more general class of functions of sums and differences than norms. A function $\varphi : V \to \mathbb{R}$ is said to be *quasi-concave* if the super-level set

$$\{x \in V : \varphi(x) > t\}$$

is convex for all $t \in \mathbb{R}$. When $V = \mathbb{R}$, this coincides with the notion of a *unimodal* function.

Proposition 7.6. *Let $a, b \in \mathbb{F}$ be non-zero numbers and let φ be a non-negative, symmetric, quasi-concave function on V. For any i.i.d. random variables X and Y taking values in V, we have*

$$\mathbb{E}\varphi(aX + bY) \leqslant \mathbb{E}\varphi\left(\frac{X - Y}{2\max\{|a|^{-1}, |b|^{-1}\}}\right).$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Proof. Since φ is non-negative, we have

$$\begin{aligned}\mathbb{E}\varphi(aX+bY) &= \int_0^\infty \mathbb{P}(\varphi(aX+bY)>t)dt \\ &= \int_0^\infty \mathbb{P}(aX+bY \in \varphi^{-1}(t,\infty))dt,\end{aligned}$$

where we let $\varphi^{-1}(t,\infty)$ denote the set $\{x \in V : \varphi(x) > t\}$. Since φ is symmetric and quasi-concave, the set $\varphi^{-1}(t,\infty)$ is symmetric and convex. Then we can apply Theorem 5.1 to have

$$\mathbb{P}(aX+bY \in \varphi^{-1}(t,\infty)) \leqslant \mathbb{P}(X-Y \in 2\max\{|a|^{-1},|b|^{-1}\}\varphi^{-1}(t,\infty)).$$

This implies that

$$\begin{aligned}\mathbb{E}\varphi(aX+bY) &\leqslant \int_0^\infty \mathbb{P}\left(\frac{X-Y}{2\max\{|a|^{-1},|b|^{-1}\}} \in \varphi^{-1}(t,\infty)\right)dt \\ &= \mathbb{E}\varphi\left(\frac{X-Y}{2\max\{|a|^{-1},|b|^{-1}\}}\right).\end{aligned}$$

□

Remark. Taking $\varphi(x) = \|x\|^p$ for $p<0$ in Proposition 7.6, we can recover the negative moment case of Theorem 7.4 when the two norms are identical. It is natural (given the use of different norms in Theorem 7.4) to consider the comparison of $\mathbb{E}\varphi(aX+bY)$ and $\mathbb{E}\phi(X-Y)$ for distinct non-negative, symmetric, quasi-concave functions φ and ϕ. When φ and ϕ are homogeneous, one can still apply Theorem 5.1 to prove results analogous to Proposition 7.6. Without homogeneity we need information on the behaviour of $N(a,b,\varphi^{-1}(t,\infty),\phi^{-1}(t,\infty))$, which is a function of $t>0$ rather than a constant; we therefore do not bother to consider this question further.

We now give a moment inequality in the spirit of Theorem 3.6, where the comparison is to the original distribution.

Theorem 7.7. *Let $p>0$. For any i.i.d. random variables X and Y taking values in V, we have*

$$(\mathbb{E}|||X-Y|||^p)^{1/p} \leqslant 2^{1+1/p}|||I||| \cdot (\mathbb{E}\|X\|^p)^{1/p}. \tag{7.9}$$

Proof. We can assume that the right-hand side of (7.9) is finite, so that

$$\begin{aligned}\mathbb{E}\|X\|^p &= p\int_0^\infty t^{p-1}\mathbb{P}(\|X\|>t)dt \\ &= p\int_0^\infty t^{p-1}(1-\mathbb{P}(\|X\|\leqslant t))dt.\end{aligned}$$

Using the geometric interpretation (7.4) of $|||I|||$ and Theorem 3.6, we have

$$\mathbb{P}(\|X\|\leqslant t) \leqslant \mathbb{P}(|||X||| \leqslant |||I|||t) \leqslant (\mathbb{P}(|||X-Y||| \leqslant 2|||I|||t))^{1/2}.$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Hence we have

$$\begin{aligned}\mathbb{E}\|X\|^p &\geqslant p\int_0^\infty t^{p-1}(1-(\mathbb{P}(|||X-Y|||\leqslant 2|||I|||t))^{1/2})dt\\ &\geqslant \frac{p}{2}\int_0^\infty t^{p-1}(1-\mathbb{P}(|||X-Y|||\leqslant 2|||I|||t))dt\\ &= \frac{p}{2}\int_0^\infty t^{p-1}\mathbb{P}(|||X-Y|||> 2|||I|||t)dt\\ &= 2^{-p-1}|||I|||^{-p}\cdot\mathbb{E}|||X-Y|||^p.\end{aligned}$$

The statement follows by taking the $(1/p)$th root. In the second inequality, we use the simple fact that $1-t\leqslant 2(1-\sqrt{t})$ for $0\leqslant t\leqslant 1$. □

Combining with a result of Guédon [31], Theorems 7.4 and 7.7 imply a reverse Hölder-type inequality for log-concave random variables. A random variable X taking values in a Banach space $(E,\|\cdot\|)$ is called *log-concave* if, for any non-empty Borel sets $A,B\subseteq E$ and $0<\lambda<1$, we have

$$\mathbb{P}(X\in\lambda A+(1-\lambda)B)\geqslant\mathbb{P}(X\in A)^\lambda\mathbb{P}(X\in B)^{1-\lambda},$$

where $\lambda A+(1-\lambda)B$ stands for the Minkowski sum $\{\lambda a+(1-\lambda)b : a\in A, b\in B\}$. Log-concave distributions are a large class, and include all Gaussian distributions, exponential distributions, and the uniform distribution over any compact convex set.

A reverse Hölder inequality asserts the equivalence of higher and lower moments of random variables. There are many different varieties of reverse Hölder inequalities, such as Khinchin inequalities or inequalities relating different L^p-norms of functions; the survey [7] contains a discussion of several of these classes. In particular, reverse Hölder inequalities have found considerable use in recent years at the interface of probability theory and convex geometry (see *e.g.* [8, 9, 10, 25, 26, 32]), For our purposes, we focus on situations where there exists a constant $C(p,q)$ depending only on $q\geqslant p$ such that

$$(\mathbb{E}\|X\|^q)^{1/q}\leqslant C(p,q)(\mathbb{E}\|X\|^p)^{1/p}$$

holds for random variables X in a normed measurable space. In general such an inequality does not hold, but it does hold for random variables with log-concave distributions. For example, Borell [11] showed the equivalence between the pth and qth moments of log-concave random variables for $q\geqslant p\geqslant 1$, while Latała [49] demonstrated that the constant $C(p,q)$ can be independent of p. Guédon [31] established a reverse Hölder inequality involving negative moments, which we now describe. Defining $M_0:=\lim_{p\to 0}(\mathbb{E}\|X\|^p)^{1/p}=\exp(\mathbb{E}\log\|X\|)$, Guédon [31] showed that there is an absolute constant C_p such that for any $-1<p\leqslant 0$, one has

$$(\mathbb{E}\|X\|^p)^{1/p}\geqslant C_pM_0.$$

In fact, for $-1<p\leqslant 0$, Guédon [31] proved the explicit form

$$(\mathbb{E}\|X\|^p)^{1/p}\geqslant\frac{1+p}{4e}\mathbb{E}\|X\|.$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

For $p > 0$, a corresponding result is not explicitly stated in the literature, but combining the estimates A (with $q = 0$) and B in [31, Theorem 4], one can prove

$$(\mathbb{E}\|X\|^p)^{1/p} \leqslant C_p M_0,$$

with

$$C_p = \min_{0 \leqslant x \leqslant 1} (1-x)^{-2} \left(x + \Gamma(p+1)(1-x)^{1/2} \left(\frac{2}{-\log(1-x)} \right)^p \right)^{1/p}.$$

An important fact implied by Prékopa–Leindler inequality is that the sum and difference of independent log-concave random variables are still log-concave. Therefore it is reasonable to expect that one may have reverse Hölder-type inequalities relating $aX + bY$ and $X - Y$ for i.i.d. log-concave random variables X and Y.

Corollary 7.8. *Let V be a real topological vector space associated with two equivalent norms $\|\cdot\|$ and $|||\cdot|||$. Let a, b be reals such that not both of them are zero. Let p, q be reals such that $0 < p \leqslant q$ or $-1 < p < q < 0$. There exists an absolute constant $C(a, b, p, q)$ such that*

$$(\mathbb{E}|||X - Y|||^q)^{1/q} \leqslant C(a, b, p, q) \cdot (\mathbb{E}\|aX + bY\|^p)^{1/p}$$

holds for any i.i.d. log-concave random variables X and Y taking values in V.

Proof. First we assume that a, b are non-zero. For $-1 < p < q < 0$, Guédon's result implies the existence of a constant C_p such that

$$(\mathbb{E}|||X - Y|||^p)^{1/p} \geqslant C_p (\mathbb{E}|||X - Y|||^q)^{1/q}.$$

Then the statement follows from Theorem 7.4. The case $0 < p < q$ can be proved in the same manner. When one of a, b is zero, the proof proceeds in the same way with the application of Theorem 7.7 instead. □

Remark. We point out that Hölder or reverse Hölder inequalities of the form

$$(\mathbb{E}\|X + Y\|^q)^{1/q} \leqslant c \cdot (\mathbb{E}|||X - Y|||^p)^{1/p}$$

do not hold, as can be seen by taking real-valued random variables X and Y to be independent and uniform on the interval $[n, n+1]$ with n sufficiently large.

7.3. Positive definite functions

This section is focused on the estimate of $\mathbb{E}\varphi(aX + bY)$, where φ is a positive definite function. The study is independent of the small ball inequalities developed in the previous sections.

Let G be an abelian group, and $\varphi : G \to \mathbb{C}$ be a Hermitian function, *i.e.* a function satisfying $\overline{\varphi(x)} = \varphi(-x)$. The function φ is said to be *positive definite* if, for any $x_1, \ldots, x_n \in G$ and $c_1, \ldots, c_n \in \mathbb{C}$, we have

$$\sum_{i,j=1}^{n} \varphi(x_i - x_j) c_i \overline{c_j} \geqslant 0.$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

Similarly the Hermitian function φ is said to be *negative definite* if the reversed inequality holds under the condition $\sum_{i=1}^{n} c_i = 0$. For example, for $0 < p \leqslant 2$, the function $e^{-\|x\|^p}$ is positive definite over the Euclidean space $\mathbb{R}^d$.

Bochner's well-known theorem asserts that a continuous positive definite function φ on a locally compact abelian group G can be uniquely represented as the Fourier transform of a positive finite Radon measure μ on the Pontryagin dual group G^*, that is,

$$\varphi(x) = \int_{G^*} \xi(x)\mu(d\xi). \tag{7.10}$$

The counterpart of Bochner's theorem is the Lévy–Khinchin representation formula for a continuous negative definite function φ on $\mathbb{R}^d$, that is,

$$\varphi(x) = c + i\langle y_0, x\rangle + q(x) + \int_{\mathbb{R}^d\setminus\{0\}} \left(1 - e^{-i\langle x,y\rangle} - \frac{i\langle x, y\rangle}{1 + \|y\|^2}\right)\mu(dy) \tag{7.11}$$

where $c \in \mathbb{R}$, $y_0 \in \mathbb{R}^d$, $q(x)$ is some quadratic form on $\mathbb{R}^d$ and μ is a Lévy measure. The close relations between these two types of functions have been well studied. For example, a function φ is negative definite if and only if $e^{-t\varphi}$ is positive definite for all $t > 0$. This observation goes back to Schoenberg. They are also closely related to another important type of functions, the so-called completely monotone functions. We refer to [5, 6] for more details in this direction.

Theorem 7.9. *Let G be a locally compact abelian group and let $\varphi : G \to \mathbb{C}$ be a continuous positive definite function. For independent random variables X and Y taking values in G and $m, n \in \mathbb{Z}$, we have*

$$|\mathbb{E}\varphi(mX + nY)|^2 \leqslant \mathbb{E}\varphi(mX - mX')\mathbb{E}\varphi(nY - nY'),$$

where X' and Y' are independent copies of X and Y, respectively.

Proof. Using Bochner's theorem (7.10), we have

$$|\mathbb{E}\varphi(mX + nY)| = \left|\mathbb{E}\int_{G^*} \xi(mX + nY)\mu(d\xi)\right| = \left|\int_{G^*} \mathbb{E}\xi(mX)\mathbb{E}\xi(nY)\mu(d\xi)\right|.$$

The second equation follows from Fubini's theorem and the assumption that X, Y are independent. By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&\left|\int_{G^*} \mathbb{E}\xi(mX)\mathbb{E}\xi(nY)\mu(d\xi)\right| \\
&\quad\leqslant \left(\int_{G^*} |\mathbb{E}\xi(mX)|^2 d\mu(\xi) \int_{G^*} |\mathbb{E}\xi(nY)|^2 d\mu(\xi)\right)^{1/2} \\
&\quad= \left(\int_{G^*} \mathbb{E}\xi(mX)\overline{\mathbb{E}\xi(mX)}\mu(d\xi) \int_{G^*} \mathbb{E}\xi(nY)\overline{\mathbb{E}\xi(nY)}\mu(d\xi)\right)^{1/2} \\
&\quad= \left(\int_{G^*} \mathbb{E}\xi(mX)\mathbb{E}\overline{\xi(mX)}\mu(d\xi) \int_{G^*} \mathbb{E}\xi(nY)\mathbb{E}\overline{\xi(nY)}\mu(d\xi)\right)^{1/2}
\end{aligned}$$

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

$$
\begin{aligned}
&= \left(\int_{G^*} \mathbb{E}\xi(mX)\mathbb{E}\xi(-mX)\mu(d\xi) \int_{G^*} \mathbb{E}\xi(nY)\mathbb{E}\xi(-nY)\mu(d\xi) \right)^{1/2} \\
&= \left(\int_{G^*} \mathbb{E}\xi(mX)\mathbb{E}\xi(-mX')\mu(d\xi) \int_{G^*} \mathbb{E}\xi(nY)\mathbb{E}\xi(-nY')\mu(d\xi) \right)^{1/2} \\
&= \left(\mathbb{E} \int_{G^*} \xi(mX - mX')\mu(d\xi)\mathbb{E} \int_{G^*} \xi(nY - nY')\mu(d\xi) \right)^{1/2} \\
&= (\mathbb{E}\varphi(mX - mX')\mathbb{E}\varphi(nY - nY'))^{1/2}.
\end{aligned}
$$

We let $\overline{\mathbb{E}\xi(mX)}$ denote the conjugate of $\mathbb{E}\xi(mX)$. The equation $\overline{\xi(mX)} = \xi(-mX)$ follows from the fact that $\xi \in G^*$. □

Let V be a locally compact topological vector space over a field $\mathbb{F}$. Repeating the above argument we can prove the following result.

Theorem 7.10. *Let $\varphi : V \to \mathbb{C}$ be a continuous positive definite function. For independent random variables X and Y taking values in V and $a, b \in \mathbb{F}$, we have*

$$|\mathbb{E}\varphi(aX + bY)|^2 \leqslant \mathbb{E}\varphi(aX - aX')\mathbb{E}\varphi(bY - bY'),$$

where X' and Y' are independent copies of X and Y, respectively.

Taking $m = n = 1$ and $m = 1, n = 0$ in Theorem 7.9, we have the following easy consequences.

Corollary 7.11. *Let G be a locally compact abelian group and let $\varphi : G \to \mathbb{C}$ be a continuous positive definite function. For i.i.d. random variables X and Y taking values in G, we have*

$$|\mathbb{E}\varphi(X + Y)| \leqslant \mathbb{E}\varphi(X - Y).$$

Corollary 7.12. *Let G be a locally compact abelian group and let $\varphi : G \to \mathbb{C}$ be a continuous positive definite function such that $\varphi(0) = 1$. For any random variable X taking values in G, we have*

$$|\mathbb{E}\varphi(X)|^2 \leqslant \mathbb{E}\varphi(X - X'),$$

where X' is an independent copy of X.

Remark. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a continuous negative definite function. Lifshits, Schilling and Tyurin [54] showed that for any i.i.d. random variables X and Y,

$$\mathbb{E}\varphi(X - Y) \leqslant \mathbb{E}\varphi(X + Y).$$

Moreover, they showed that $\mathbb{E}\varphi(X + Y) - \mathbb{E}\varphi(X - Y)$ is the variance of an integrated centred Gaussian process by employing the Lévy–Khinchin representation (7.11).

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

References

[1] Abbe, E., Li, J. and Madiman, M. (2017) Entropies of weighted sums in cyclic groups and an application to polar codes. *Entropy* **19** 235.

[2] Alon, N. and Yuster, R. (1995) The 123 theorem and its extensions. *J. Combin. Theory Ser. A* **72** 322–331.

[3] Ambrosio, L., Fusco, N. and Pallara, D. (2000) *Functions of Bounded Variation and Free Discontinuity Problems*, Clarendon Press.

[4] Bateman, P. and Erdős, P. (1951) Geometrical extrema suggested by a lemma of Besicovitch. *Amer. Math. Monthly* **58** 306–314.

[5] Berg, C. (2008) Stieltjes–Pick–Bernstein–Schoenberg and their connection to complete monotonicity. In *Positive Definite Functions: From Schoenberg to Space-Time Challenges* (J. Mateu and E. Porcu, eds), University Jaume I, pp. 15–45.

[6] Berg, C., Christensen, J. P. R. and Ressel, P. (1984) *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*, Springer.

[7] Bobkov, S., Fradelizi, M., Li, J. and Madiman, M. (2018) When can one invert Hölder's inequality? (and why one may want to). Preprint.

[8] Bobkov, S. and Madiman, M. (2011) Concentration of the information in data with log-concave distributions. *Ann. Probab.* **39** 1528–1543.

[9] Bobkov, S. and Madiman, M. (2011) The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory* **57** 4940–4954.

[10] Bobkov, S. and Madiman, M. (2012) Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures. *J. Funct. Anal.* **262** 3309–3339.

[11] Borell, C. (1974) Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252.

[12] Buja, A., Logan, B. F., Reeds, J. A. and Shepp, L. A. (1994) Inequalities and positive-definite functions arising from a problem in multidimensional scaling. *Ann. Statist.* **22** 406–438.

[13] Burago, Y. D. and Zalgaller V. A. (1988) *Geometric Inequalities*, Vol. 285 of Grundlehren der Mathematischen Wissenschaften, Springer.

[14] Burchard, A. (2009) A short course on rearrangement inequalities. http://www.math.utoronto.ca/almut/rearrange.pdf

[15] Burchard, A. and Schmuckenschläger, M. (2001) Comparison theorems for exit times. *Geom. Funct. Anal.* **11** 651–692.

[16] Cohn, H., Kumar, A., Miller, S., Radchenko, D. and Viazovska, M. S. (2017) The sphere packing problem in dimension 24. *Ann. Math.* **185** 1017–1033.

[17] Conway, J. H. and Sloane, N. J. A. (1999) *Sphere Packings, Lattices and Groups*, Vol. 290 of Grundlehren der Mathematischen Wissenschaften, Springer.

[18] Deshouillers, J.-M., Freiman, G. A. and Yudin, A. A. (2001) On bounds for the concentration function II. *J. Theoret. Probab.* **14** 813–820.

[19] Deshouillers, J.-M. and Sutanto. (2005) On the rate of decay of the concentration function of the sum of independent random variables. *Ramanujan J.* **9** 241–250.

[20] Doeblin, W. (1939) Sur les sommes d'un grand nombre des variables aléatoires indépendantes. *Bull. Sci. Math.* **63** 23–32, 35–64.

[21] Dong, D., Li, J. and Li, W. V. (2015) A note on distribution-free symmetrization inequalities. *J. Theoret. Probab.* **28** 958–967.

[22] Eliseeva, Y. S. and Zaitsev, A. Y. (2013) Estimates of the concentration functions of weighted sums of independent random variables. *Theory Probab. Appl.* **57** 670–678.

[23] Esseen, C. G. (1966) On the Kolmogorov–Rogozin inequality for the concentration function. *Z. Wahrsch. Verw. Gebiete.* **5** 210–216.

[24] Esseen, C. G. (1968) On the concentration function of a sum of independent random variables. *Z. Wahrsch. Verw. Gebiete.* **9** 290–308.

[25] Fradelizi, M., Li, J. and Madiman, M. (2015) Concentration of information content for convex measures. arXiv:1512.01490

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

[26] Fradelizi, M., Madiman, M. and Wang, L. (2016) Optimal concentration of information content for log-concave densities. In *High Dimensional Probability VII: The Cargèse Volume* (C. Houdré, D. Mason, P. Reynaud-Bouret and J. Rosinski, eds), Progress in Probability, Birkhäuser.
[27] Friedland, O. and Sodin, S. (2007) Bounds on the concentration function in terms of the Diophantine approximation. *CR Math. Acad. Sci. Paris* **345** 513–518.
[28] Gao, F. (2018) A probabilistic characterization of negative definite functions. Preprint.
[29] Götze, F., Eliseeva, Y. S. and Zaitsev, A. Y. (2016) Arak's inequalities for concentration functions and the Littlewood–Offord problem. *Dokl. Akad. Nauk* **467** 514–518.
[30] Giné, E. and Zinn, J. (1984) Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–998.
[31] Guédon, O. (1999) Kahane–Khinchine type inequalities for negative exponent. *Mathematika* **46** 165–173.
[32] Guédon, O. and Milman, E. (2011) Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geom. Funct. Anal.* **21** 1043–1068.
[33] Hales, T. C. (2005) A proof of the Kepler conjecture. *Ann. Math.* **162** 1065–1185.
[34] Hengartner, W. and Theodorescu, R. (1973) *Concentration Functions*, Academic Press.
[35] Jain, N. C. and Marcus, M. B. (1975) Central limit theorems for $C(S)$-valued random variables. *J. Funct. Anal.* **19** 216–231.
[36] Kanter, M. (1976) Probability inequalities for convex sets and multidimensional concentration functions. *J. Multivariate Anal.* **6** 222–236.
[37] Katona, D. and Stečkin, B. S. (1980) Combinatorial numbers, geometric constants and probabilistic inequalities. *Dokl. Akad. Nauk SSSR* **251** 1293–1296.
[38] Katona, G. O. H. (1969) Graphs, vectors and inequalities in probability theory. *Mat. Lapok.* **20** 123–127.
[39] Katona, G. O. H. (1977) Inequalities for the distribution of the length of a sum of random vectors. *Teor. Verojatnost. i Primenen* **22** 466–481.
[40] Katona, G. O. H. (1978) Continuous versions of some extremal hypergraph problems. In *Proceedings of the Colloquium Mathematical Society*, Vol. 18 of János Bolyai Mathematical Studies, pp. 653–678.
[41] Katona, G. O. H. (1980) Continuous versions of some extremal hypergraph problems II. *Acta Math. Acad. Sci. Hungar.* **35** 67–77.
[42] Katona, G. O. H. (1981) Sums of vectors and Turán's problem for 3-graphs. *European J. Combin.* **2** 145–154.
[43] Katona, G. O. H. (1982) 'Best' estimations on the distribution of the length of sums of two random vectors. *Z. Wahrsch. Verw. Gebiete* **60** 411–423.
[44] Katona, G. O. H. (1983) Sums of vectors and Turán's graph problem. In *Combinatorial Mathematics*, Vol. 75 of North-Holland Mathematics Studies, North-Holland, pp. 377–382.
[45] Katona, G. O. H. (1985) Probabilistic inequalities from extremal graph results (a survey). *Ann. Discrete Math.* **28** 159–170.
[46] Kawohl, B. (1985) *Rearrangements and Convexity of Level Sets in PDE*, Vol. 1150 of Lecture Notes in Mathematics, Springer.
[47] Kesten, H. (1969) A sharper form of the Doeblin–Lévy–Kolmogorov–Rogozin inequality for concentration functions. *Math. Scand.* **25** 133–144.
[48] Kolmogorov, A. (1958) Sur les propriétés des fonctions de concentrations de M. P. Lévy. *Ann. Inst. H. Poincaré* **16** 27–34.
[49] Latała, R. (1996) On the equivalence between geometric and arithmetic means for log-concave measures. In *Convex Geometric Analysis*, Vol. 34 of Mathematical Sciences Research Institute Publications, Cambridge University Press, pp. 123–127.
[50] Leech, J. (1956) The problem of the thirteen spheres. *Math. Gaz.* **40** 22–23.
[51] Levenšteĭn, V. I. (1979) Boundaries for packings in n-dimensional Euclidean space. *Dokl. Akad. Nauk SSSR* **245** 1299–1303.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press

[52] Lévy, P. (1937) *Théorie de l'Addition des Variables Aléatoires*, Gauthier-Villars.
[53] Lieb, E. H. and Loss, M. (2001) *Analysis*, second edition, American Mathematical Society.
[54] Lifshits, M., Schilling, R. L. and Tyurin, I. (2013) A probabilistic inequality related to negative definite functions. *Progr. Probab.* **66** 73–80.
[55] Madiman, M. and Kontoyiannis, I. (2010) The entropies of the sum and the difference of two IID random variables are not too different. In *2010 IEEE International Symposium on Information Theory*, IEEE.
[56] Madiman, M. and Kontoyiannis, I. (2018) Entropy bounds on abelian groups and the Ruzsa divergence. *IEEE Trans. Inform. Theory* **64** 77–92.
[57] Madiman, M., Melbourne, J. and Xu, P. (2017) Forward and reverse entropy power inequalities in convex geometry. In *Convexity and Concentration* (E. Carlen, M. Madiman and E. M. Werner, eds.), Vol. 161 of IMA Volumes in Mathematics and its Applications, Springer, pp. 427–485.
[58] Madiman, M., Melbourne, J. and Xu, P. (2017) Rogozin's convolution inequality for locally compact groups. arXiv:1705.00642
[59] Madiman, M., Wang, L. and Woo, J. O. (2017) Entropy inequalities for sums in prime cyclic groups. arXiv:1710.00812
[60] Madiman, M., Wang, L. and Woo, J. O. (2017) Discrete entropy power inequalities via Sperner theory. arXiv:1712.00913
[61] Marcinkiewicz, J. and Zygmund, A. (1939) Quelques inégalités pour les opérations linéaires. *Fund. Math.* **32** 115–121.
[62] Mattner, L. (1997) Strict definiteness of integrals via complete monotonicity of derivatives. *Trans. Amer. Math. Soc.* **349** 3321–3342.
[63] Mirošnikov, A. L. and Rogozin, B. A. (1980) Inequalities for concentration functions. *Teor. Veroyatnost. i Primenen* **25** 178–183.
[64] Musin, O. R. (2003) The problem of the twenty-five spheres. *Uspekhi Mat. Nauk* **58** 153–154.
[65] Nguyen, H. H. and Vu V. H. (2013) Small ball probability, inverse theorems, and applications. In *Erdős Centennial* (L. Lovász, I. Z. Ruzsa and V. T. Sós, eds), Vol. 25 of Bolyai Society Mathematical Studies, Springer, pp. 409–463.
[66] Odlyzko, A. M. and Sloane, N. J. A. (1979) New bounds on the number of unit spheres that can touch a unit sphere in n dimensions. *J. Combin. Theory Ser. A*, *26* 210–214.
[67] Pollard, D. (1989) Asymptotics via empirical processes. *Statist. Sci.* **4** 341–366.
[68] Rogozin, B. A. (1961) On the increase of dispersion of sums of independent random variables. *Teor. Verojatnost. i Primenen* **6** 106–108.
[69] Rudelson, M. and Vershynin, R. (2008) The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633.
[70] Sidorenko, A. F. (1983) Extremal estimates of probability measures and their combinatorial nature. *Math. USSR-Izv.* **20** 503-533.
[71] Siegmund-Schultze, R. and von Weizsäcker, H. (2007) Level crossing probabilities I: One-dimensional random walks and symmetrization. *Adv. Math.* **208** 672–679.
[72] Viazovska, M. S. (2017) The sphere packing problem in dimension 8. *Ann. Math.* **185** 991–1015.
[73] Wang, L and Madiman, M. (2014) Beyond the entropy power inequality, via rearrangements. *IEEE Trans. Inform. Theory* **60** 5116–5137.
[74] Watanabe, T. (1983) The isoperimetric inequality for isotropic unimodal Lévy processes. *Z. Wahrsch. Verw. Gebiete* **63** 487–499.

https://doi.org/10.1017/S0963548318000494 Published online by Cambridge University Press