SIAM J. DISC. MATH.
Vol. 7, No. 1, pp. 90–101, February 1994

009

# LITTLEWOOD–OFFORD INEQUALITIES FOR RANDOM VARIABLES*

I. LEADER† AND A. J. RADCLIFFE‡

**Abstract.** The *concentration* of a real-valued random variable $X$ is

$$c(X) = \sup_{t \in \mathbb{R}} \mathbf{P}(t < X < t + 1).$$

Given bounds on the concentrations of $n$ independent random variables, how large can the concentration of their sum be?

The main aim of this paper is to give a best possible upper bound for the concentration of the sum of $n$ independent random variables, each of concentration at most $1/k$, where $k$ is an integer. Other bounds on the concentration are also discussed, as well as the case of vector-valued random variables.

**Key words.** Littlewood–Offord problem, concentration, normed spaces

**AMS subject classifications.** 60G50, 06A07, 52A40

**Introduction.** In 1943, Littlewood and Offord [8], concerned with estimating the number of real zeros of random polynomials, proved that, given complex numbers $(a_i)_1^n$ of modulus at least 1, not too many of the sums $s_A = \sum_{i \in A} a_i$, $A \subset \{1, 2, \ldots, n\}$ lie in any open disc of diameter 1. They showed that the maximum number is $O(2^n n^{-1/2} \log n)$.

In 1945, Erdös [2] noted that, if the $a_i$ are real numbers, then Sperner's theorem—on the maximum size of an antichain in the poset $\mathcal{P}(n) = \mathcal{P}(\{1, 2, \ldots, n\})$—implies a best possible upper bound. Indeed, suppose first that the $a_i$ are all positive. Then, given an open interval $I$ of length 1, the set system $\mathcal{A}_I = \{A \subset \{1, 2, \ldots, n\} : s_A \in I\}$ is an antichain, since, if $B \supset A$, then $s_B - s_A = s_{B \setminus A} \geq |B \setminus A| \geq 1$. Thus, for all $I$, $\mathcal{A}_I \leq \binom{n}{\lfloor n/2 \rfloor}$ by Sperner's theorem [9]. The result for positive reals immediately implies that the same conclusion follows for all reals. Kleitman [5] and Katona [4] independently showed that the same bound, of $\binom{n}{\lfloor n/2 \rfloor}$, holds for $(a_i)_1^n$ in $\mathbb{C}$, thus giving a best possible improvement of the lemma of Littlewood and Offord. In [6] Kleitman proved a considerable extension of this result, namely, to sums of vectors $(a_i)_1^n$ of norm at least 1 in an arbitrary normed space, thus setting a conjecture of Erdös.

Jones [3] suggested a probabilistic framework for these questions, regarding a vector $a \neq 0$ in a normed space $E$ as being naturally associated with an $E$-valued random variable $X_a$ with $\mathbf{P}(X_a = 0) = \frac{1}{2}$ and $\mathbf{P}(X_a = a) = \frac{1}{2}$. So, if $\delta_a$ is the delta measure on $E$ concentrated at $a$, then the distribution of $X_a$ is $\frac{1}{2}(\delta_0 + \delta_a)$. Kleitman's result can then be stated as follows.

THEOREM A (see [6]). *Let $(a_i)_1^n$ be vectors in a normed space $E$ of norm at least 1 and let $(X_i)_1^n$ be independent random variables with $X_i$ having distribution $\frac{1}{2}(\delta_0 + \delta_{a_i})$. Then, for any open set $U \subset X$ of diameter at most 1, we have*

$$\mathbf{P}\left(\sum_1^n X_i \in U\right) \leq 2^{-n} \binom{n}{\lfloor n/2 \rfloor}.$$

Note that this bound is clearly best possible, equality being attained if, for instance, all the $a_i$ are equal.

* Received by the editors November 11, 1991; accepted for publication (in revised form) November 4, 1992.

† Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England.

‡ Department of Mathematics, Carnegie–Mellon University, Pittsburgh, Pennsylvania 15213-3890.

The conclusion of Theorem A gives a bound on the extent to which the values of the random variable $\sum X_i$ are concentrated in one place. This prompts the following definition.

Let $E$ be a normed space. The *concentration* of an $E$-valued random variable $X$ is $c(X) = \sup \mathbf{P}(X \in U)$, where the supremum is taken over all open subjects $U \subset E$ having diameter at most 1.

The hypotheses of Theorem A can also be stated in terms of concentration, and in this form it reads as follows.

THEOREM A′. *Let $(X_i)_1^n$ be independent $E$-valued random variables that are essentially two-valued and have concentration at most $\frac{1}{2}$. Then*

$$c\left(\sum_1^n X_i\right) \leq 2^{-n}\binom{n}{\lfloor n/2 \rfloor}.$$

The main result of this paper is a result that extends Theorem A′ in the case when $E = \mathbb{R}$ by removing the restriction that each $X_i$ be essentially two-valued.

THEOREM 1. *Let $(X_i)_1^n$ be independent real-valued random variables of concentration at most $\frac{1}{2}$. Then*

$$c\left(\sum_1^n X_i\right) \leq 2^{-n}\binom{n}{\lfloor n/2 \rfloor}.$$

Our technique is closely related to that of Kleitman, being based on symmetric chain decompositions. In §1 we give a proof of Theorem 1, as well as presenting some background about symmetric chain decompositions.

In §2 we consider sums of random variables of concentration at most $1/q$, where $q$ is an integer, and we generalise some results of Jones [3]. To state our result, we need some fairly standard notation. We write $[q]$ for the set $\{0, 1, \ldots, q-1\}$ and also for the poset with that ground set and the natural ordering. We write $[q]^n$ for the product of $n$ copies of $[q]$ with the usual product ordering, i.e., $(x_i)_1^n \leq (y_i)_1^n$ if and only if $x_i \leq y_i$ for each $i$. Finally, we write $W$ for the size of the largest level set in the ranked poset $[q]^n$ as follows:

$$W = W_{q,n} = |\{(x_i)_1^n \in [q]^n : \textstyle\sum x_i = \lfloor n(q-1)/2 \rfloor\}|.$$

THEOREM 2. *Let $(X_i)_1^n$ be independent real-valued random variables with $c(X_i) \leq 1/q$, where $q \in \mathbb{N}$. Then $c(\sum_1^n X_i) \leq W/q^n$.*

This bound is clearly best possible. Equality is attained when, for instance, each $X_i$ has distribution $(1/q)(\delta_0 + \delta_1 + \cdots + \delta_{q-1})$.

Based on Theorem 2, we are perhaps tempted to guess that the sum of $n$ independent random variables, each of concentration at most $p/q$ ($p$ and $q$ coprime integers), has concentration bounded by the proportion of $[q]^n$ occupied by the largest $p$ layers. Unfortunately, very simple examples show that this is not the case. Rather surprisingly, given this, the result does hold when $p = 2$. Both the examples and the proof are given in §3.

Finally, in §4, we turn our attention to the vector-valued case. We consider some of the problems raised by Jones [3] and answer some of his questions.

**1. Sums of random variables of concentration at most $\frac{1}{2}$.** Before considering the details of our proof of Theorem 1, some discussion of symmetric chain decompositions is in order. These will prove to be vital for our results, as they were for Kleitman's.

A *symmetric chain decomposition* of the power set $\mathscr{P}(n) = \mathscr{P}\{1, 2, \ldots, n\}$ is a partition of $\mathscr{P}(n)$ into chains (totally ordered subsets) in such a way that each chain

$\mathcal{A} = \{A_1, A_2, \ldots, A_r\}$ with $A_1 \subset A_2 \subset \cdots \subset A_r$ satisfies $|A_{k+1}| = |A_k| + 1$ and $|A_1| + |A_r| = n$. Thus each $\mathcal{A}$ is arrayed symmetrically about the "middle layer" of $\mathcal{P}(n)$ and contains one set from each layer between the extremes. In particular, of course, $\mathcal{A}$ must contain one set of size $\lfloor n/2 \rfloor$. de Bruijn, Tengbergen, and Kruyswijk [1] showed that symmetric chain decompositions do exist, and Kleitman's beautiful result, Theorem A, was based on that proof.

Their proof goes as follows. Suppose that $(\mathcal{A}_j)_1^s$ is a symmetric chain decomposition of $\mathcal{P}(n-1)$. We can construct a symmetric chain decomposition of $\mathcal{P}(n)$ in the following manner. Take a copy of $(\mathcal{A}_j)_1^s$ in each layer of $\mathcal{P}(n)$: the bottom layer, which is exactly $\mathcal{P}(n-1)$, and the top layer (of sets containing $n$). This is very definitely not a symmetric chain decomposition of $\mathcal{P}(n)$, but, by transferring the top element of each chain in the top layer to the corresponding chain downstairs, everything can be fixed. More precisely, for each chain $\mathcal{A}_j = \{A_1, A_2, \ldots, A_r\}$ with $A_1 \subset A_2 \subset \cdots \subset A_r$ set

$$\mathcal{A}_j' = \{A_1, A_2, \ldots, A_r, A_r \cup \{n\}\},$$

$$\mathcal{A}_j'' = \{A_1 \cup \{n\}, A_2 \cup \{n\}, \ldots, A_{r-1} \cup \{n\}\}.$$

The collection $\{\mathcal{A}_j', \mathcal{A}_j''\}_1^s$ forms a symmetric chain decomposition of $\mathcal{P}(n)$, after the removal of those $\mathcal{A}_j''$ that are empty.

A sequence $(m_j)_1^s$ is called a *symmetric profile* for $\mathcal{P}(n)$ if, for some (and therefore, up to rearrangement, every) symmetric chain decomposition of $\mathcal{P}(n)$, say $(\mathcal{A}_j)_1^s$, we have $m_j = |\mathcal{A}_j|$. Note that $s = \binom{n}{\lfloor n/2 \rfloor}$.

As the above proof shows, we get a symmetric profile for $\mathcal{P}(n+1)$ by taking $(m_i)_1^s$ and replacing each $m_j$ by the pair $m_j - 1$, $m_j + 1$ and then discarding zeros. At first, this may seem not to the point, since we can write the symmetric profile for $\mathcal{P}(n)$ easily and explicitly: A sequence $(m_j)_1^s$ is a symmetric profile for $\mathcal{P}(n)$ if $s = \binom{n}{\lfloor n/2 \rfloor}$ and the number of $j$ with $m_j = n + 1 - 2i$ is $\binom{n}{i} - \binom{n}{i-1}$ (with the convention that $\binom{n}{-1} = 0$). However, symmetric chain decompositions will arise in more complicated situations, in which finding an explicit expression is much harder. Fortunately, all that we need for the proofs is the total number of chains in a decomposition and the way in which the symmetric profile changes as the poset grows. To illustrate this, below is Kleitman's proof of Theorem A, using symmetric profiles.

*Proof of Theorem* A. The values of $X = \sum_1^n X_i$ are exactly those vectors in $E$ of the form $x_A = \sum_{i \in A} a_i$, where $A$ is any subset of $\{1, 2, \ldots, n\}$. The distribution of $X$ is $2^{-n} \sum_{A \subset \{1,2,\ldots,n\}} \delta_{x_A}$. To show that $c(X)$ is small, we partition $\mathcal{P}(n)$ into subsets $(\mathcal{A}_j)_1^s$ with $s = \binom{n}{\lfloor n/2 \rfloor}$ and

$$(*) \qquad A, B \in \mathcal{A}_j \Rightarrow \|x_A - x_B\| \geq 1.$$

To do this, we in fact do more, namely, prove that the partition can be chosen with $(|\mathcal{A}_j|)_1^s$ being a symmetric profile for $\mathcal{P}(n)$. Once this is proved, the theorem follows easily, since

$$\begin{aligned}
\mathbf{P}(X \in U) &= 2^{-n} \sum_{A \subset \{1,2,\ldots,n\}} \delta_{x_A}(U) \\
&= 2^{-n} \sum_{j=1}^{s} \sum_{A \in \mathcal{A}_j} \delta_{x_A}(U) \\
&\leq 2^{-n} s \\
&= 2^{-n} \binom{n}{\lfloor n/2 \rfloor}.
\end{aligned}$$

The last inequality holds, since, by $(*)$, at most one $x_A$ with $A \in \mathcal{A}_j$ belongs to $U$.

The proof goes by induction on $n$. The result is trivial for $n = 1$, so we turn to the induction step. Take an appropriate partition $\mathscr{P}(n-1) = \cup_1^{s'} \mathscr{A}_j$ (where $s' = \binom{n-1}{\lfloor (n-1)/2 \rfloor}$). Take a support functional $f \in X^*$ for $a_n$, a functional with $\|f\| = 1$ and $f(a_n) = \|a_n\| \geq 1$. For any $\mathscr{A}_j = \{A_1, A_2, \ldots, A_r\}$, choose $l$ with

$$f(x_{A_l}) \geq f(x_{A_k}), \qquad k = 1, 2, \ldots, r$$

and set

$$\mathscr{A}_j' = \{A_1, A_2, \ldots, A_r, A_l \cup \{n\}\},$$

$$\mathscr{A}_j'' = \{A_1 \cup \{n\}, A_2 \cup \{n\}, \ldots, A_{l-1} \cup \{n\}, A_{l+1} \cup \{n\}, \ldots, A_r \cup \{n\}\}.$$

The partition of $\mathscr{P}(n)$ that is needed consists of all the nonempty $\mathscr{A}_j'$ and $\mathscr{A}_j''$. Clearly, each $\mathscr{A}_j''$ satisfies $(*)$, since $\mathscr{A}_j$ did originally. In $\mathscr{A}_j'$, we need only check that, for each $A_k \in \mathscr{A}_j$, the norm of $x_{A_l \cup \{n\}} - x_{A_k}$ is large. This follows by applying $f$ as follows:

$$\begin{aligned} \|x_{A_l \cup \{n\}} - x_{A_k}\| &\geq f(x_{A_l \cup \{n\}} - x_{A_k}) \\ &= f(a_n) + f(x_{A_l}) - f(x_{A_k}) \\ &\geq f(a_n) \geq 1. \end{aligned}$$

The profile of the new partition is a symmetric profile of $\mathscr{P}(n)$, since each $\mathscr{A}_j$ of size $m$ splits into two, of sizes $m+1$ and $m-1$. Thus by induction the theorem is proved. □

In the proof of Theorem 1, we will be dealing with random variables and their distributions, treating the latter similarly to the finite subsets of $E$ that arise in the proof of Theorem A. Indeed, we often regard a finite subset of $\mathbb{R}$ as corresponding to a random variable that assigns equal mass to those points and none to all others. We are interested in the distribution of the sum of these random variables, that is, in the convolution of their distributions.

More generally, we will be dealing with finite (positive Borel) measures on $\mathbb{R}$—the collection of all such we denote by $\mathscr{M}$. However, we wish to stress that we will not really be using any measure theory. Indeed, a reader who considers only measures of finite support will not be losing much.

For $\mu \in \mathscr{M}$, the *mass* of $\mu$ is $|\mu| = \mu(\mathbb{R})$. Just as before, the *concentration* of $\mu$ is $c(\mu) = \sup \mu(I)$, where the supremum is over all open intervals of length 1. The convolution of $\mu, \lambda \in \mathscr{M}$ is denoted by $\mu * \lambda$, and we write $\mathscr{M}(m, c)$ for the set of all $\mu \in \mathscr{M}$ of mass $m$ and concentration at most $c$.

Two elementary facts are summarised in the following lemma.

LEMMA 3. *If $\mu$ has mass $m$ and $\lambda$ has concentration at most $c$, then $\mu * \lambda$ has concentration at most $mc$. Also, if $(\mu_i)_1^n \subset \mathscr{M}$, then $c(\sum_1^n \mu_i) \leq \sum_1^n c(\mu_i)$.*

*Proof.* Given any interval $I = (t, t+1)$, we have

$$\begin{aligned} \mu * \lambda(I) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_I(x+y)\, d\lambda(x)\, d\mu(y) \\ &= \int_{\mathbb{R}} \lambda(I - y)\, d\mu(y) \\ &\leq \int_{\mathbb{R}} c\, d\mu(y) = mc, \end{aligned}$$

proving the first statement. The second is immediate. □

A standard approach in the proofs will be to split up a measure $\mu$ into parts with almost disjoint support. We need notation for these parts and therefore we define

$$\text{Left}(\mu, m) = \mu|_{(-\infty, t]} - x\delta_t,$$

where $t = \sup\{x : \mu(-\infty, x) \leq m\}$ and $x = \mu(-\infty, t] - m$. Thus the support of Left $(\mu, m)$ is contained in $(-\infty, t]$ and $|\text{Left}(\mu, m)| = m$. We define Right $(\mu, m)$ in a similar fashion.

The next lemma, which is rather technical, enables us to peel off, from a convolution of measures of concentration at most 1, a part that also has concentration at most 1. This process is analogous to the transfer that occurs in the de Bruijn / Tengbergen / Kruyswijk proof of the existence of symmetric chain decompositions.

LEMMA 4. *If $\mu$ and $\lambda$ are measures of concentration* 1 *and mass at least* 1, *then, writing $\mu_R$ for* Right $(\mu, 1)$ *and $\lambda_L$ for* Left $(\lambda, 1)$, *the measure $\nu = \mu_R * \lambda + \mu * \lambda_L - \mu_R * \lambda_L$ belongs to $\mathcal{M}(m(\mu) + m(\lambda) - 1, 1)$.*

*Proof.* Let $I = (t, t+1)$ be an interval of length 1 in $\mathbb{R}$. We wish to show that $\nu(I) \leq 1$. Set $\mu_L = \mu - \mu_R$ and $x_\mu = \inf\{t : \mu(t, \infty) \leq 1\}$. Similarly, let $\lambda_R = \lambda - \lambda_L$ and $x_\lambda = \sup\{t : \lambda(-\infty, t) \leq 1\}$. Then we can also write $\nu$ as $\mu_L * \lambda_L + \mu_R * \lambda_L + \mu_R * \lambda_R$. If $\mu_L * \lambda_L(I) = 0$, then $\nu(I) = \mu_R * \lambda(I) \leq 1$, the last since $\mu_R$ has mass 1 and $\lambda$ has concentration 1. Similarly, we are finished if $\mu_R * \lambda_R(I) = 0$. If neither is zero, then necessarily $t \leq x_\lambda + x_\mu \leq t + 1$. In this case, we split $\lambda$ yet further. Write

$$\lambda_{LL} = \lambda_L|_{(-\infty, t - x_\mu]}, \qquad \lambda_{LR} = \lambda_L|_{(t - x_\mu, \infty)},$$

$$\lambda_{RL} = \lambda_R|_{(-\infty, t + 1 - x_\mu]}, \qquad \lambda_{RR} = \lambda_R|_{[t + 1 - x_\mu, \infty)}.$$

Note that $\mu_R * \lambda_{RR}(I) = \mu_L * \lambda_{LL}(I) = 0$, so

$$\begin{aligned}\nu(I) &= (\mu_R * \lambda_{LL} + \mu_L * \lambda_{LR} + \mu_R * \lambda_{LR} + \mu_R * \lambda_{RL})(I) \\ &= (\mu * \lambda_{LR})(I) + (\mu_R * (\lambda_{LL} + \lambda_{RL}))(I).\end{aligned}$$

Since $\mu$ has concentration 1, the first term is at most $|\lambda_{LR}| = \lambda_L(t - x_\mu, x_\lambda]$. The measure $\mu_R$, on the other hand, has mass 1, so, to prove the lemma, it suffices to show that the concentration of $\lambda_{LL} + \lambda_{RL}$ is at most $1 - \lambda_L(t - x_\mu, x_\lambda]$.

By considering the various ways in which an interval of length 1 could overlap with $(t - x_\mu, x_\lambda]$, it is easy to see that

$$c(\lambda_{LL} + \lambda_{LR}) \leq \max\{|\lambda_{LL}|, |\lambda_{RL}|, 1 - |\lambda_{LR}|\}.$$

Now, the first and last of these terms are $|\lambda_{LL}| = 1 - |\lambda_{LR}| = 1 - \lambda_L(t - x_\mu, x_\lambda]$, while

$$\begin{aligned}|\lambda_{RL}| &= \lambda(t - x_\mu, t - x_\mu + 1) - \lambda_L(t - x_\mu, x_\lambda] \\ &\leq 1 - \lambda_L(t - x_\mu, x_\lambda],\end{aligned}$$

since $\lambda$ has concentration 1. Thus the result is proved. □

With this lemma, it is simple to deduce the following, more comprehensible version.

LEMMA 5. *If $\mu$ is a measure of mass $m \geq 1$ and $\lambda$ is a measure of mass* 2, *and each has concentration at most* 1, *then the convolution $\mu * \lambda$ can be written as a sum of two measures of concentration at most* 1, *$\mu * \lambda = \nu' + \nu''$, with $|\nu'| = m + 1$ and $|\nu''| = m - 1$.*

*Proof.* With the same notation as in Lemma 4, set $\nu' = \nu$ and $\nu'' = \mu * \lambda - \nu$. Then, by that lemma, we have $\nu' \in \mathcal{M}(m + 1, 1)$, and certainly $\mu * \lambda = \nu' + \nu''$. Also, $\nu'' = \mu_L * \lambda_R$ and, since $\mu_L$ has concentration at most 1 while $\lambda_R$ has mass 1, Lemma 3 shows that $\nu'' \in \mathcal{M}(m - 1, 1)$. □

These tools suffice for the proof of Theorem 1.

*Proof of Theorem* 1. Let $\mu_{X_i}$ be the distribution of $X_i$ and set $\mu_i = 2\mu_{X_i}$. The theorem states that, whenever $U$ is an open subset of $\mathbb{R}$ with diameter at most 1, then $(\otimes_1^n \mu_i)(U) \leq \binom{n}{\lfloor n/2 \rfloor}$. More is true; in fact, $\otimes_1^n \mu_i$ can be written as a sum $\sum_1^s \nu_j$, of

measures of concentration at most 1, where $(|\nu_j|)_1^s$ is a symmetric profile for $\mathscr{P}(n)$. Again, the proof goes by induction on $n$. If $\otimes_1^{n-1}\mu_i = \sum_1^s \nu_j$, where each $\nu_j$ has concentration 1 and $(|\nu_j|)_1^s$ is a symmetric profile of $\mathscr{P}(n-1)$, then Lemma 5 gives

$$\begin{aligned}\bigotimes_1^n \mu_i &= \left(\sum_1^s \nu_j\right) * \mu_n \\ &= \sum_1^s \nu_j * \mu_n \\ &= \sum_1^s \nu_j' + \nu_j''.\end{aligned}$$

Since each $\nu_j$ splits into two new measures, of masses $|\nu_j| + 1$ and $|\nu_j| - 1$, the masses of the new decomposition form a symmetric profile of $\mathscr{P}(n)$. So

$$\begin{aligned}c\left(\sum_1^n X_i\right) &= 2^{-n} c\left(\bigotimes_1^n \mu_i\right) \\ &= 2^{-n} c\left(\sum_{j=1}^s \nu_j\right) \\ &\le 2^{-n} \sum_{j=1}^s c(\nu_j) \\ &\le 2^{-n} s \\ &= 2^{-n}\binom{n}{\lfloor n/2 \rfloor}.\end{aligned}$$

Thus the result is proved. □

**2. Concentration at most $1/q$.** In this section, we extend Theorem 1 to measures of concentration at most $1/q$ for some fixed integer $q$. The techniques used are a straightforward extension of those in the proof of Theorem 1.

The first step is to note that the poset $[q]^n$ has a symmetric chain decomposition. This poset is ranked by *weight*: $w(x) = \sum_1^n x_i$ for $x \in [q]^n$. A chain $x^{(1)} \le x^{(2)} \le \cdots \le x^{(r)}$ is *symmetric* if $w(x^{(k+1)}) = w(x^{(k)}) + 1$ and $w(x^{(1)}) + w(x^{(r)}) = n(q-1)$. It was proved by de Bruijn, Kruyswijk, and Tengbergen [1] that $[q]^n$ has a symmetric chain decomposition. Again, we say that a sequence $(m_j)_1^s$ is a *symmetric profile* for $[q]^n$ if, for some (and hence, up to rearrangement, for any) symmetric chain decomposition $(\mathscr{S}_j)_1^s$, we have $|\mathscr{S}_j| = m_j$. The required information about how symmetric profiles change is here stated as a lemma.

LEMMA 6. *Let $(m_j)_1^s$ be a symmetric profile for $[q]^{n-1}$, and, for each $j = 1, 2, \ldots, s$, set $r_j = \min\{q, m_j\}$. Then the sequence obtained by replacing $m_j$ by the $r_j$ values $m_j + q + 1 - 2k$ for $k = 1, 2, \ldots, r_j$ is a symmetric profile for $[q]^n$.*

*Proof.* Consider a chain $\mathscr{S} = (x^{(k)})_1^r$ belonging to a symmetric chain decomposition of $[q]^{n-1}$. For $x \in [q]^{n-1}$ and $h \in [q]$, denote by $x + he_n$ the element of $[q]^n$ formed by appending $h$ to $x$. For $0 \le l \le \min(q, r) - 1$, let $\mathscr{S}^{(l)} = \{x^{(k)} + he_n : \min(r-k, h) = l\}$. Then each $\mathscr{S}^{(l)}$ is a symmetric chain in $[q]^n$, and the union of all the chains arising in this way forms a symmetric chain decomposition of $[q]^n$. See Fig. 1 and [1] for more details. □

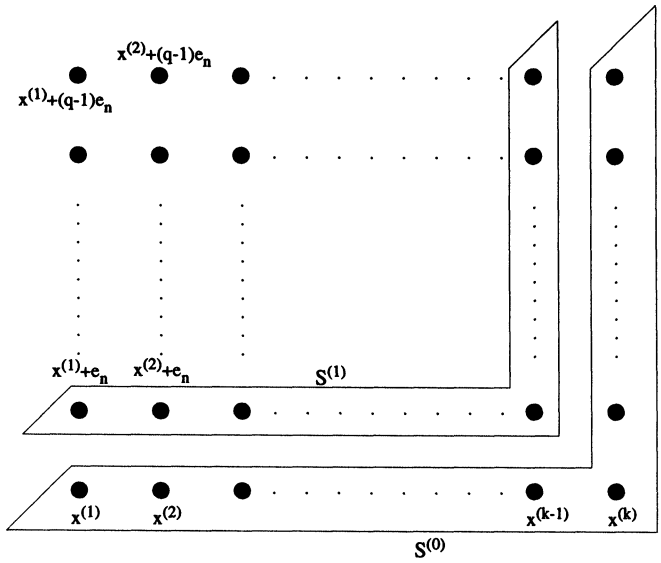


FIG. 1. *The construction of symmetric chains.*

The next lemma extends Lemma 5 to cover the present case. It states, in essence, that we can treat a measure of mass $m$ and concentration 1 much like the poset $[m]$. In particular, the convolution of an element of $\mathcal{M}(m, 1)$ and one of $\mathcal{M}(l, 1)$ behaves similarly to the poset $[m] \times [l]$ and can be "peeled apart" in the same fashion. This peeling process will allow us to write a convolution $\otimes_1^n \mu_i$ (where $\mu_i \in \mathcal{M}(q, 1)$) as a sum of measures whose mass profile mimics that of a symmetric chain decomposition of the poset $[q]^n$.

LEMMA 7. *Given* $m, l \in \mathbb{N}$, *set* $r = \min\{m, l\}$. *If* $\mu \in \mathcal{M}(m, 1)$ *and* $\lambda \in \mathcal{M}(l, 1)$, *then* $\mu * \lambda$ *can be written as a sum* $\sum_{k=1}^{r} \nu^{(k)}$ *in which* $\nu^{(k)} \in \mathcal{M}(m + l + 1 - 2k, 1)$.

*Proof.* The case $l = 2$ of this lemma is precisely Lemma 5, and the case where $l = 1$ is trivial: the splitting $\nu^{(1)} = \mu * \lambda$ will do. The general case is proved by induction on $l$. Using the result and notation of Lemma 4, set $\nu^{(1)} = \mu_R * \lambda + \mu * \lambda_L - \mu_R * \lambda_L$, with $|\nu^{(1)}| = m + l - 1$ and $c(\nu^{(1)}) \leq 1$. After some judicious relabelling, the induction hypothesis states that $\mu * \lambda - \nu^{(1)} = \mu_L * \lambda_R$ can be written as $\sum_{k=2}^{r} \nu^{(k)}$, with $\nu^{(k)} \in \mathcal{M}(m + l + 1 - 2k, 1)$. □

*Proof of Theorem* 2. We prove by induction that the convolution $\otimes_1^n q\mu_{X_i}$ can be decomposed as a sum of measures of concentration at most 1 whose mass profile is a symmetric profile for $[q]^n$. Indeed, this is trivially possible if $n = 1$. For the induction step, let $(\nu_j)_1^s$ be a decomposition for $\otimes_1^{n-1} q_i \mu_{X_i}$. Then

$$\mu = \left( \sum_{j=1}^{s} \nu_j \right) * (q\mu_{X_n})$$

$$= \sum_{j=1}^{n} \nu_j * (q\mu_{X_n}).$$

Now, by Lemma 7, each convolution $\nu_j * (q\mu_{X_n})$ can be written as a sum $\sum_{k=1}^{r_j} \nu_j^{(k)}$, where $r_j = \min\{q, |\nu_j|\}$ and $\nu_j^{(k)} \in \mathcal{M}(|\nu_j| + q + 1 - 2k, 1)$ for $k = 1, 2, \ldots, r_j$. The collection of all nonzero $\nu_j^{(k)}$ is a decomposition of $\mu$ into measures of concentration 1, and by Lemma 6 their masses form a symmetric profile for $[q]^n$. □

*Remark*. The proof of Theorem 2 can easily be extended to prove rather more. Indeed, if $(X_i)_1^n$ are independent real-valued random variables with $c(X_i) \leq 1/q_i$ for all $i$ (where $q_1, \ldots, q_n$ are integers), we can show that the concentration of $\sum X_i$ is at most the proportion of $\prod_1^n [q_i]$ occupied by the largest layer. In fact, by using slightly more information about the symmetric profile of $\prod_1^n [q_i]$, we can show that, given $r$ open intervals $(I_k)_1^r$, each of length at most 1, the probability that $\sum_1^n X_i$ lies in $\cup_1^r I_k$ is bounded by the proportion of $\prod_1^n [q_i]$ occupied by the $r$ largest layers.

**3. Other values of the concentration.** What can be said about the sum of independent random variables of concentration at most $c$ for values of $c$ not of the form $1/q$? Might it be that the sum of $n$ independent random variables, each of concentration $p/q$ ($p$, $q$ coprime integers) has concentration at most the proportion of $[q]^n$ occupied by the $p$ largest layers? It is easy to see that this cannot be the case, since we may have quite complicated fractions $p/q$ that closely approximate some simple number such as $\frac{1}{2}$. For instance, let $X_1$, $X_2$ be independent and identically distributed random variables with distribution $(\delta_0 + \delta_1)/2$. The $X_i$ certainly have concentration at most $\frac{4}{7}$. However, their sum has concentration $\frac{1}{2}$, which is greater than $\frac{24}{49}$, the proportion of $[7]^2$ occupied by the four largest layers.

On the other hand, somewhat surprisingly given this simple example, the question above does have a positive answer when $p = 2$, in other words, for concentrations of the form $2/q$ with $q$ odd. The proof proceeds by showing how to "peel apart" convolutions of measures of concentration 2.

One preliminary lemma is necessary.

LEMMA 8. *If $q \in \mathbb{N}$ and $\mu \in \mathcal{M}(q, 2)$, then there exist measures $\mu_0$, $\mu_1$, both of concentration* 1, *with $\mu = \mu_0 + \mu_1$ and $|\mu_0| = \lfloor q/2 \rfloor$ and $|\mu_1| = \lceil q/2 \rceil$.*

*Proof*. Split $\mu$ into parts of mass 1 and (almost) disjoint support going from left to right. In other words, define $\nu_1 = \text{Left}(\mu, 1)$ and for $j = 2, 3, \ldots, q$ set

$$\nu_j = \text{Left}\left(\mu - \sum_1^{j-1} \nu_k, 1\right).$$

Now collect all the $\nu_j$ for $j$ even together and similarly for all the odd $\nu_j$ as follows:

$$\mu_h = \sum_{j \equiv h (\text{mod } 2)} \nu_j, \qquad h = 0, 1.$$

Then it is clear that $\mu_0$, $\mu_1$ have the correct masses. Now let $I = (t, t+1)$ be an arbitrary interval. Since $\mu$ has concentration 2, at most three of the $\nu_k$ can have $\nu_k(I) > 0$. This is because, if four of them could detect $I$, of necessity four consecutive ones, $\nu_k$, $\nu_{k+1}$, $\nu_{k+2}$, and $\nu_{k+3}$, say, then we would have $\nu_{k+1} = \nu_{k+2} = 1$ and $\mu(I) \geq (\sum_{l=0}^3 \nu_{k+l})(I) > (\nu_{k+1} + \nu_{k+2})(I) = 2$. This contradicts the fact that $\mu$ has concentration 2.

If exactly three of the $\nu_k$ give positive measure to $I$, then similarly they are $\nu_k$, $\nu_{k+1}$ and $\nu_{k+2}$ with $\nu_{k+1}(I) = 1$. Thus

$$\nu_k(I) + \nu_{k+2}(I) = \mu(I) - \nu_{k+1}(I) \leq 2 - 1 = 1.$$

So, in the case when the support of three $\nu_k$ intersect $I$, we have $\mu_0(I)$, $\mu_1(I) \leq 1$. In the case when at most two supports are involved, it is clear that both $\mu_0$ and $\mu_1$ are at most 1 on $I$, and the proof is complete. □

We are very fortunate to have the following lemma.

LEMMA 9. *If $\mu \in \mathcal{M}(m, 2)$ and $\lambda \in \mathcal{M}(l, 2)$, with $m$ and $l$ odd, then $\mu * \lambda$ can be written as a sum of measures of concentration at most* 2 *whose masses form a symmetric profile for $[m] \times [l]$.*

*Proof.* Write $m = 2a + 1$ and $l = 2b + 1$. By Lemma 8, we can write $\mu = \mu_0 + \mu_1$ and $\lambda = \lambda_0 + \lambda_1$, measures of concentration at most 1 and masses $a$, $a + 1$ and $b$, $b + 1$, respectively. Without loss of generality, $b \leq a$, but there are two cases, when $b < a$ and when $b = a$.

*Case* 1. $b < a$.

By Lemma 7, $\mu_0 * \lambda_0$ splits into $b$ measures of concentration at most 1 and masses $a + b - 1, a + b - 3, \ldots, a - b + 1$. The convolution $\mu_1 * \lambda_0$ can be decomposed into $b$ measures of concentration at most 1 of masses $a + b, a + b - 2, \ldots, a - b + 2$. Summing in pairs, we can write $\mu * \lambda_0$ as the sum of $b$ measures of concentration at most 2 of masses $2a + 2b - 1, 2a + 2b - 5, \ldots, 2a - 2b + 3$.

In a similar fashion, both $\mu_0 * \lambda_1$ and $\mu_1 * \lambda_1$ split into $b + 1$ pieces of concentration at most 1. When paired up, these give measure of concentration at most 2 and masses $2a + 2b + 1, 2a + 2b - 1, \ldots, 2a - 2b - 3$. The two collections together provide an appropriate splitting for $\mu * \lambda$.

*Case* 2. $b = a$.

Partition $\mu * \lambda_0$ as before. Now, however, $\mu_0 * \lambda_1$ splits into only $a$ parts, of masses $a + b, a + b - 2, \ldots, 2$, whereas $\mu_1 * \lambda_1$ splits as before into $b + 1$ parts: masses $a + b + 1, a + b - 1, \ldots, 1$. Pair these measures, leaving the final measures of mass 1 unpaired. This produces $b$ measures of concentration at most 2, of masses $2a + 2b + 1$, $2a + 2b - 3, \ldots, 5$. Together with the remaining measure of mass 1 (and therefore certainly of concentration at most 2), this gives us exactly the desired splitting. □

We are ready to study the case of concentration $2/q$.

THEOREM 10. *Let $q \in \mathbb{N}$ be odd and let $(X_i)_1^n$ be independent real-valued random variables with $c(X_i) \leq 2/q$. Let $W_1$ and $W_2$ be the sizes of the two largest layers in $[q]^n$. Then*

$$c\left(\sum_{i=1}^{n} X_i\right) \leq (W_1 + W_2)/q^n.$$

*Proof.* Following the proof of Theorem 1, using Lemma 9 rather than Lemma 5, we can write $\mu = \otimes_1^n (q_i \mu_{X_i})$ as a sum $\sum_1^s \nu_j$, where each $\nu_j$ has concentration at most 2 and $(|\nu_j|)_1^s$ is a symmetric profile for $[q]^n$. What does this profile look like? If $W_2 < W_1$, then there must be exactly $W_1 - W_2$ 1's in it. If $W_1 = W_2$, then the corresponding chain decomposition can have no chains of length 1. In summary, exactly $W_1 - W_2$ of the $\nu_j$ have mass 1 (and therefore concentration at most 1) and the other $W_2$ have concentration at most 2. Thus

$$\begin{aligned} c(\mu) &\leq \sum_{1}^{s} c(\nu_j) \\ &\leq 2W_2 + (W_1 - W_2) \\ &= W_1 + W_2, \end{aligned}$$

and so $c(X) = c(\mu)/q^n \leq (W_1 + W_2)/q^n$. □

We note that Theorem 10 is best possible, as may be seen by taking each $X_i$ to have distribution $(1/q)(\delta_0 + \delta_1 + \cdots + \delta_{q-1})$.

The question remains as to what can be said for other values of the concentration. If $(X_i)_1^n$ are independent real-valued random variables each of concentration at most $c$, can we give good upper bounds for the concentration of $\sum X_i$?

**4. The vector-valued case.** In the first three sections, we have concentrated on the behavior of real-valued random variables. We turn now to the situation that was Jones's

[3] primary concern, the vector-valued case. Refer to [7] for general background and notation about normed spaces.

Jones studied the following question. A subset $M$ of a normed space $E$ is said to be 1-*separated* if, for any distinct $x, y \in M$, we have $\|x - y\| \geq 1$. Given sets $M_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,q}\}$ for $i = 1, 2, \ldots, n$ such that each is 1-separated, form the corresponding random variables $(X_i)_1^n$ with $X_i$ having distribution $(1/q) \sum_{j=1}^{q} \delta_{x_{i,j}}$. Is it true that the concentration of $X = \sum X_i$ is always bounded as we would wish, by the proportion of $[q]^n$ occupied by the largest layer? This question remains unanswered.

In his study of the problem, Jones introduced some useful definitions (given here in slightly less generality than in [3]). Let $M_1$ and $M_2$ be 1-separated finite subsets of a normed space $E$ with $|M_i| = m_i$. We say that the pair $M_1, M_2$ has the *B.T.K. chain property* if their sum $M_1 + M_2$, counted with multiplicities, can be partitioned into a family of 1-separated subsets whose size profile is a symmetric profile for $[m_1] \times [m_2]$. We say that $E$ has the *B.T.K. chain property* if every pair of finite 1-separated subsets of $E$ has the B.T.K. chain property. If $E$ has the B.T.K. chain property, then the above question has an affirmative answer (as may be seen by mimicking symmetric chain decompositions of $[q]^n$).

A partial ordering $\leq$ on a normed space $E$ is said to be *compatible* if it is translation invariant (i.e., satisfies $x \leq y$ if and only if $x + a \leq y + a$ for all $x, y, a \in E$) and has the property that distinct $x, y \in E$ are comparable if and only if $\|x - y\| \geq 1$. Thus, for example, $\mathbb{R}$ certainly has a compatible ordering: we let $x$ precede $y$ if $x + 1 \leq y$ in the usual order.

Jones showed that, if $X$ has a compatible order, then $X$ has the B.T.K. chain property. He proved that two-dimensional Hilbert space, and hence each higher-dimensional Hilbert space, fails to have the B.T.K. chain property and (a fortiori) has no compatible order. He asked whether $l_\infty$ has a compatible order, or at least satisfies the B.T.K. chain property.

In some sense, Jones answered this question himself, since we can find two-dimensional subspaces of $l_\infty$ isometric to Hilbert space, and so $l_\infty$ cannot have the B.T.K. chain property. In fact, compatible orders are rather hard to find: no normed space of dimension greater than 1 has a compatible ordering. Moreover, the condition that $\leq$ be translation invariant is not the reason.

PROPOSITION 11. *Let $E$ be a normed space of dimension greater than* 1. *Then there is no partial ordering $\leq$ on $E$ such that distinct $x, y \in E$ are comparable if and only if $\|x - y\| \geq 1$.*

*Proof.* It clearly suffices to show that no two-dimensional example exists, so let us suppose that $E$ is a two-dimensional space with such an ordering $\leq$. Let $\{x_1, x_2\}$, $\{x_1^*, x_2^*\}$ be an Auerbach system for $X$. Thus $x_1$ and $x_2$ have norm 1; $x_1^*$ and $x_2^*$, belonging to $E^*$, have dual norm 1; and $x_i^*(x_j) = \delta_{ij}$ (such a system can easily be found—see, e.g., [7]). Consider first the set $M = \{0, x_1, x_2\}$. Since $\|x_1 - x_2\| \geq x_1^*(x_1 - x_2) = 1$, the set $M$ is 1-separated and hence totally ordered by $\leq$.

CLAIM. *Either $x_1 \leq 0 \leq x_2$ or $x_2 \leq 0 \leq x_1$.*

Otherwise, we may suppose, without loss of generality, that $0 \leq x_1 \leq x_2$. Consider then $y = x_2/2$. We have that $\|x_1 - y\| \geq x_1^*(x_1 - y) = 1$, so either $x_1 > y$ or $x_1 < y$. In the first case, we have $x_2 > x_1 > y$, despite the fact that $\|x_2 - y\| = \frac{1}{2}$. In the second case, $y > x_1 > 0$, which again contradicts the condition on $\leq$ since $\|y\| = \frac{1}{2}$. So the claim is proved.

Exactly the same reasoning, applied to $M' = \{0, x_1, x_1 + x_2\}$, shows that $x_1$ must be $\leq$-between 0 and $x_1 + x_2$. Similarly, we must have $x_1 + x_2$ between $x_1$ and $x_2$ and also $x_2$ between 0 and $x_1 + x_2$. However, these four conditions are incompatible—the $\leq$-maximum of the four vectors $\{0, x_1, x_2, x_1 + x_2\}$ does not lie between two others. This contradiction establishes the nonexistence of $(E, \leq)$. □

Jones's main positive result was that, if $M_1$ and $M_2$ are both 1-separated subsets of Hilbert space, with $M_2$ having size at most 3, then the pair $M_1$, $M_2$ has the B.T.K. chain property. (We should note that Kleitman based his proof of Theorem A exactly on the fact that in any normed space, any pair $M_1$, $M_2$ of 1-separated subsets has the B.T.K. chain property if $M_2$ has size at most 2.)

This pleasant fact about Hilbert space does not, unfortunately, generalise to arbitrary normed spaces. We present here an example of a normed space $E$ and two 1-separated subsets, each of size 3, not having the B.T.K. chain property. In fact, we find 1-separated sets $M_1$, $M_2$ of size 3 such that the sum $M_1 + M_2$, far from having a partition into 1-separated subsets of size 5, 3, and 1, does not even contain a 1-separated subset of size 5.

PROPOSITION 12. *There exists a normed space $E$ and 1-separated subsets $M_1$, $M_2 \subset E$ such that $|M_1| = |M_2| = 3$ but $M_1 + M_2$ contains no 1-separated subset of size 5.*

*Proof.* We define a norm on $\mathbb{R}^4$ in such a way that the sets $M_1 = \{0, e_1, e_2\}$ and $M_2 = \{0, e_3, e_4\}$ satisfy the conclusion of the proposition. More exactly, we ensure that each distance marked on Fig. 2 is strictly less than 1, while both $M_1$ and $M_2$ are 1-separated. To ensure that the requisite vectors are short, we define our norm $\|\cdot\|$ by taking for its unit ball the absolute convex hull of these vectors. In other words, we take as the unit ball the set

$$B_{\|\cdot\|} = \text{abs-co}\,\{e_1 + e_3, (e_2 - e_1) + e_4, e_2 - e_4, e_2 - e_3,$$
$$e_1 + (e_4 + e_3), e_2 + (e_3 - e_4), (e_1 - e_2) + (e_4 - e_3)\}.$$

By definition, all these vectors have norm at most 1. Now we show that the vectors $e_1$, $e_2$, $e_3$, $e_1 - e_2$, $e_3 - e_4$ have norm strictly greater than 1 by exhibiting functionals of (dual) norm at most 1 taking large values at those vectors. For instance, for $\varepsilon > 0$ sufficiently small, the functional $f = (1 + \varepsilon, 0, -\varepsilon, -2\varepsilon)$ has dual norm at most 1, because it takes values at most 1 in absolute value at the extreme points of the $\|\cdot\|$ unit ball. However, $f(e_1) = 1 + \varepsilon > 1$, and therefore $\|e_1\| > 1$. In similar fashion, we can exhibit

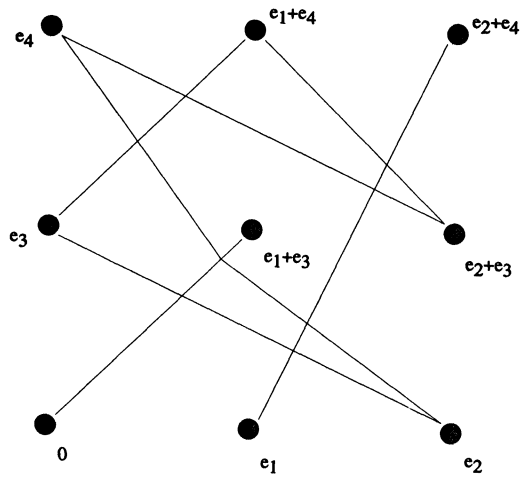


FIG. 2. *The pattern of small distances in Proposition* 12.

functionals to show that all the vectors we desire to be long are indeed long. For some small $\varepsilon > 0$, the following suffice:

$$\begin{aligned} e_1: &\quad (1+\varepsilon, 0, -\varepsilon, -2\varepsilon), \\ e_2: &\quad (3\varepsilon, 1+\varepsilon, \varepsilon, 2\varepsilon), \\ e_3: &\quad (-\varepsilon, \varepsilon, 1+\varepsilon, 3\varepsilon), \\ e_4: &\quad (2\varepsilon, \varepsilon, 3\varepsilon, 1+\varepsilon), \\ e_2 - e_1: &\quad \tfrac{1}{2}(1+\varepsilon, -1, 2\varepsilon, \varepsilon), \\ e_3 - e_4: &\quad \tfrac{1}{2}(-\varepsilon, \varepsilon, -1-\varepsilon, 1). \end{aligned}$$

The norm $\|\cdot\|$ does not behave exactly as we would like—the norms from Fig. 2 are at most 1, rather than strictly less than 1—but for some $0 < \lambda < 1$, the norm $\lambda\|\cdot\|$ will do. It is easy to check, from Fig. 2, that $M_1 + M_2$ contains no 1-separated subset of size 5. □

There are still many unanswered questions concerning the vector-valued case. The most striking and interesting one, it seems to us, is whether the following conjecture is true.

CONJECTURE 13. *Let $E$ be a normed space and let $(X_i)_1^n$ be independent $E$-valued random variables of concentrations at most $\frac{1}{2}$. Then the concentration of $\sum X_i$ is at most $\binom{n}{\lfloor n/2 \rfloor} 2^{-n}$.*

## REFERENCES


[1] N. G. DE BRUIJN, D. K. KRUYSWIJK, AND CA. VAN EBBENHORST TENGBERGEN, *On the set of divisors of a number*, Nieuw Arch. Wisk., 23 (1952), pp. 191–193.

[2] P. ERDÖS, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc., 51 (1945), pp. 898–902.

[3] L. JONES, *On the distribution of sums of vectors*, SIAM J. Appl. Math., 34 (1978), pp. 1–6.

[4] G. O. H. KATONA, *On a conjecture of Erdös and a stronger form of Sperner's theorem*, Studia Sci. Math. Hungar., 1 (1966), pp. 59–63.

[5] D. J. KLEITMAN, *On a lemma of Littlewood and Offord on the distribution of linear combinations of vectors*, Adv. in Math., 5 (1970), pp. 155–157.

[6] ———, *On a lemma of Littlewood and Offord on the distribution of certain sums*, Math. Z., 90 (1965), pp. 251–259.

[7] J. LINDENSTRAUSS AND L. TZAFRIRI, *Classical Banach Spaces* I, Springer-Verlag, Berlin, 1977.

[8] J. E. LITTLEWOOD AND C. OFFORD, *On the number of real roots of a random algebraic equation* (III), Math. USSR-Sb., 12 (1943), pp. 277–285.

[9] E. SPERNER, *Ein Satz über Untermengen einer endlichen Menge*, Math. Z., 27 (1928), pp. 544–548.