Anticoncentration via the Strong Perfect Graph Theorem

Tomas Juškevičius (with V. Kurauskas)

Institute of Computer Science of the Czech Academy of Sciences

2023-06-13

The Littlewood-Offord problem

Theorem (Erdős 45)

Let $\varepsilon_1, \dots, \varepsilon_n$ be independent random variables such that $\mathbb{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$. Then for $a_i \in \mathbb{R}$ with $|a_i| \geq 1$ we have

$$\mathbb{P}(a_1\varepsilon_1 + \cdots + a_n\varepsilon_n \in (x, x+2]) \leq \mathbb{P}(\varepsilon_1 + \cdots + \varepsilon_n \in (-1, 1]).$$

Answering a question of Erdős, Kleitman (70) extended the result to arbitrary normed spaces.

Extension to other random variables

Theorem (Leader-Radcliffe 94)

Let $X_1, \ldots, X_n$ be independent random variables such that for all $x \in \mathbb{R}$

$$\mathbb{P}(X_i \in (x, x+2)) \leq \frac{1}{k}.$$

Then

$$\mathbb{P}(X_1 + \cdots + X_n \in (x, x+2)) \leq \mathbb{P}(U_1 + \cdots + U_n \in (-1, 1]),$$

where U_i's are independent uniform random variables on the k-point set $\{-k+1, -k+3, \ldots, k-3, k-1\}$.

Henceforth we shall denote the latter uniform distribution by ν^k.

Other values of concentration 1

Leader and Radcliffe asked about the situation with concentration bound being not an inverse of an integer.

Notation. For $\alpha \in [\frac{1}{k+1}, \frac{1}{k}]$ denote by $T(\alpha)$ a real random variable having distribution

$$p\nu^{k+1} + (1-p)\nu^k, p = k(k+1)\alpha - k,$$

where $p = k(k+1)\alpha - k$. The latter choise ensures that the sum of two consecutive atoms of $T(\alpha)$ is exactly α.

Other values of concentration 2

Theorem (J., 2015)

Let $X_1, \ldots, X_n$ be independent random variables such that for all $x \in \mathbb{R}$

$$\mathbb{P}(X_i \in (x, x+2)) \leq \alpha_i.$$

Then

$$\mathbb{P}(X_1 + \cdots + X_n \in (x, x+2)) \leq \mathbb{P}(T_1(\alpha_1) + \cdots + T_n(\alpha_n) \in (-1, 1]).$$

Setting $\alpha_i = \frac{1}{k}$ recovers the previous result by Leader-Radcliffe.

A conjecture by Leader-Radcliffe

Let $\mathcal{M}$ be a normed space. Define a notion of concentration of a random vector X by

$$\tilde{Q}(X, D) = \sup_A \mathbb{P}(X \in A),$$

where A runs through all open sets of diameter D in $\mathcal{M}$.

Conjecture. Let $X_1, \ldots, X_n$ be independent random vectors in some normed space $\mathcal{M}$ such that for all i we have

$$\tilde{Q}(X_i, 2) \leq \frac{1}{2}.$$

Then

$$\tilde{Q}(X_1 + \cdots + X_n, 2) \leq \mathbb{P}(\varepsilon_1 + \cdots + \varepsilon_n \in (-1, 1]).$$

A conjecture by Lee Jones 78

Definition. We shall call a set of k points in a normed space $\mathcal{M}$ a k-**block** if their pairwise distances are at least 2. We sometimes omit the parameter k when it is convenient and just call such sets **blocks**.

Conjecture. Let $X_1, \ldots, X_n$ be independent random vectors in a normed space $\mathcal{M}$. Assume that each of them is uniform on some $k-$block. Then

$$\tilde{Q}(X_1 + \cdots + X_n, 2) \leq \tilde{Q}(U_1 + \cdots + U_n, 2),$$

where U_i's are independent and each distributed uniformly on the $k-$block $\{-k+1, -k+3, \ldots, k-3, k-1\}$.

Main result

Theorem (J., V. Kurauskas 2023+)

Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbb{R}^d$ with a norm $||\cdot||$. Assume that $\tilde{Q}(X_i, 2) \leq \alpha$ for some $\alpha \in [0, 1]$. Then

$$\tilde{Q}(X_1 + \cdots + X_n, 2) \leq (1 + o(1))\tilde{Q}(T_1(\alpha) + \cdots + T_n(\alpha), 2),$$

where the $o(1)$ term depends on d and the underlying norm.

Remark. Apart from the $1 + o(1)$ multiplicative factor, the latter inequality gives exactly the
1) bound in the Leader-Radcliffe conjecture taking $\alpha = 2$;
2) bound in the conjecture of Lee Jones taking $\alpha = \frac{1}{k}$.

Difficulties and a negative result

Difficulties of the proof - circumventing the Krein-Milman theorem (board).

It might be tempting to believe that the latter inequality is true without the parasitic $1 + o(1)$ factor. We now show that this natural conjecture is false, even for $n = 2$. A counterexample for 2-dimensional Euclidean space is when both X_1 and X_2 are chosen uniformly at random from the vertices of a regular octagon with radius $2(\sqrt{2}+2)^{-\frac{1}{2}} \approx 0.5412$. Then

$$\tilde{Q}(X_1 + X_2, 2) = \tilde{Q}(X_1, 2) = \frac{3}{8}.$$

This value is sharp when $\alpha = \frac{3}{8}$ since $\tilde{Q}(X_1 + X_2, 2) \leq \tilde{Q}(X_1, 2)$. It is also strictly larger than $\tilde{Q}(T_1(\frac{3}{8}) + T_2(\frac{3}{8}), 2)$.

Setup for today

In order to present the proof ideas we shall consider the problem in a more restricted setting:

1) Only the l_2 norm;

2) Finitely supported random variables with rational probabilities.

Splitting into cases (idealized)

Case 1 (structured): all distributions lie close to a single line;

Case 2 (unstructured): distributions are substantially "high-dimensional", i.e., the spread of the sum is happening in multiple directions.

Unstructured case

Use Halasz's concentration bound that gives probabilities of magnitude $o(n^{-1/2})$.

Structured case

Now every distribution is close to a single line (to be quantified later). Take any random vector X in $\mathbb{R}^d$ with finite support and rational probabilities. We can assume that X is uniformly distributed on some multiset M (find a common denominator for the probabilities ant take each point appropriately many times).

The Perfect Graph Theorem

Definition. A (finite simple) graph is called **perfect** iff the chromatic number and the clique number coincide on all of its induced subgraphs.

Theorem (Chudnovsky, Robertson, Seymour, Thomas 2003)

A graph G is perfect iff neither G nor its complement have induced odd cycles of length ≥ 5.

The use in our setting

Recall that we are working with random vectors X concentrated on some multiset $M \subset \mathbb{R}^d$ such that $\tilde{Q}(X, 2) \leq \alpha$.

Define a distance graph on M in the following way - treat all elements of M as distinct (repeating elements are distinguished) vertices of a graph G. Two vertices $x, y \in M$ are joined iff $d(x, y) < 2$.

Theorem (J., V. Kurauskas)

Let G be a distance graph of points $x_1, \dots, x_N$ in $\mathbb{R}^d$, and assume each of the points is of distance at most $\frac{\sqrt{3}}{2}$ from the line generated by the vector $(1, 0, \dots, 0)^T$ (the "x axis"). Then neigher G nor its complement contain an induced cycles of odd length 5 or more.

So what? Quite a lot, actually.

We now know that in the structured case our distance graph of the supporting multiset of each variable is perfect. The optimal coloring of any such graph produces a certain number of independent sets (color classes) which we can control via the clique number. The condition $\tilde{Q}(X, 2) \leq \alpha$ means that the largest clique in G has at most α-proportion of the vertices which is also a bound on the chromatic number.

Why care about the chromatic classes (independent sets)? **They decompose the distribution of X into blocks!** And we can control the number of them (and thus, the sizes).

Where are the bodies burried?

Lee Jones actually has proved his conjecture when the underlying normed space satisfies a certain partitioning condition (called the De Bruijn, Tengbergen, Kruyswijk or B.K.T. condition) of the sumset of blocks. Regrettably, that condition fails for l_2. Yet we can justify it in the structured case and we do not need it in the unrestricted case.

Finishing up

To get the desired result we split the underlying product measure into product measures of blocks and use a more general version of Lee Jones's result when the B.K.T. condition is satisfied. This makes every distribution on a block into a distribution ν^k. Putting it all back into a single product structure gives us a convolution of measures on the line by keeping the concentration condition $\tilde{Q}(X_i, 2) \leq \alpha$. Then we just apply what we know about the linear case.

THE END