Singularity of random Bernoulli 0/1 matrices

Alexander Litvak

University of Alberta

based on a work in progress with

K. Tikhomirov

EIMI, St. Petersburg, 2019

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

What is the probability that the vectors are linearly dependent?

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

What is the probability that the vectors are linearly dependent?

The trivial lower bound:

$$P_n \geq \mathbb{P}\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o(1))\, 2n^2\, 2^{-n}.$$

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \dots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

What is the probability that the vectors are linearly dependent?

The trivial lower bound:

$$P_n \geq \mathbb{P}\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o(1))\, 2n^2\, 2^{-n}.$$

A natural conjecture: This is the main reason for singularity.

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

What is the probability that the vectors are linearly dependent?

The trivial lower bound:

$$P_n \geq \mathbb{P}\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o(1))\, 2n^2\, 2^{-n}.$$

A natural conjecture: This is the main reason for singularity.

Conjecture 1: $$P_n = (1/2 + o(1))^n = 2^{-(1+o(1))n}.$$

Random ± 1 matrices

An old problem: Let B be an $n \times n$ random matrix with i.i.d. ± 1 entries. What is

$$P_n := \mathbb{P}\{B \text{ is singular}\}?$$

Equivalently: Let $X_1, X_2, \dots, X_n$ be independent random vectors uniformly distributed on the vertices of the n-dimensional cube $[-1, 1]^n$.

What is the probability that the vectors are linearly dependent?

The trivial lower bound:

$$P_n \geq \mathbb{P}\{\text{Two rows/columns of } B \text{ are equal up to a sign}\} \geq (1 - o(1))\, 2n^2\, 2^{-n}.$$

A natural conjecture: This is the main reason for singularity.

Conjecture 1: $P_n = (1/2 + o(1))^n = 2^{-(1+o(1))n}.$

Conjecture 2: $P_n = (1+o(1))\, 2n^2\, 2^{-n}.$

Known results

Komlós (**67**): $P_n \to 0.$

Known results

Komlós (**67**): $P_n \to 0.$

Kahn, Komlós and Szemerédi (**95**): $P_n \leq 0.999^n.$

Known results

Komlós (**67**): $P_n \to 0$.

Kahn, Komlós and Szemerédi (**95**): $P_n \le 0.999^n$.

Tao–Vu (**07**): $P_n \le (3/4 + o(1))^n$.

Known results

Komlós (**67**): $P_n \to 0.$

Kahn, Komlós and Szemerédi (**95**): $P_n \le 0.999^n.$

Tao–Vu (**07**): $P_n \le (3/4 + o(1))^n.$

Bourgain–Vu–P.M. Wood (**10**): $P_n \le (1/\sqrt{2} + o(1))^n.$

Known results

Komlós (**67**): $P_n \to 0$.

Kahn, Komlós and Szemerédi (**95**): $P_n \le 0.999^n$.

Tao–Vu (**07**): $P_n \le (3/4 + o(1))^n$.

Bourgain–Vu–P.M. Wood (**10**): $P_n \le (1/\sqrt{2} + o(1))^n$.

K. Tikhomirov (**19+**): $P_n \le (1/2 + o(1))^n$, solving Conjecture 1.

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2]$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p.

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2]$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Note that B_p can be viewed as the adjacency matrix of Erdős–Rényi graph — a random graph on n vertices whose edges appear independently of others with probability p.

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2]$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Note that B_p can be viewed as the adjacency matrix of Erdős–Rényi graph — a random graph on n vertices whose edges appear independently of others with probability p.

Question: What is

$$P_n := \mathbb{P}\{B_p \text{ is singular}\}?$$

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2]$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Note that B_p can be viewed as the adjacency matrix of Erdős–Rényi graph — a random graph on n vertices whose edges appear independently of others with probability p.

Question: What is

$$P_n := \mathbb{P}\{B_p \text{ is singular}\}?$$

Conjecture:

$$P_n = (1 + o(1))\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (1 + o(1))\, 2n(1 - p)^n.$$

Geometrically the condition means that either $\exists$ a zero column or $\exists$ a *coordinate* hyperplane such that all columns belong to it.

Bernoulli random matrices

One can ask a similar question about Bernoulli 0/1 random matrices:

Let $p \in (0, 1/2]$ and let B_p be an $n \times n$ random matrix with i.i.d. $0/1$ random variables taking value 1 with probability p. Note that B_p can be viewed as the adjacency matrix of Erdős–Rényi graph — a random graph on n vertices whose edges appear independently of others with probability p.

Question: What is

$$P_n := \mathbb{P}\{B_p \text{ is singular}\}?$$

Conjecture:

$$P_n = (1 + o(1))\mathbb{P}\{\exists \text{ a zero row or a zero column}\} = (1 + o(1))\, 2n(1 - p)^n.$$

Geometrically the condition means that either $\exists$ a zero column or $\exists$ a *coordinate* hyperplane such that all columns belong to it.

Many works on different models of sparse matrices (with iid entries): Götze–A. Tikhomirov, Costello–Vu, Basak–Rudelson, Rudelson–K. Tikhomirov, Tao–Vu.

Bernoulli random matrices

Basak–Rudelson (**17**): $P_n \le \exp(-cnp)$ for $p = p(n) \ge (C \ln n)/n$,

Bernoulli random matrices

Basak–Rudelson (**17**): $P_n \le \exp(-cnp)$ for $p = p(n) \ge (C\ln n)/n$, moreover

$$\mathbb{P}\left\{s_n(B_p) \le c\exp(-C\ln(1/p)/\ln(np))\, t\sqrt{p/n}\right\} \le t + \exp(-cnp),$$

where

$$s_n(M) = \inf_{|x|=1} |Mx|.$$

Bernoulli random matrices

Basak–Rudelson (**17**): $P_n \le \exp(-cnp)$ for $p = p(n) \ge (C \ln n)/n$, moreover

$$\mathbb{P}\left\{ s_n(B_p) \le c \exp(-C \ln(1/p)/\ln(np))\, t\sqrt{p/n} \right\} \le t + \exp(-cnp),$$

where

$$s_n(M) = \inf_{|x|=1} |Mx|.$$

K. Tikhomirov (**19+**): $P_n \le (1 - p + o(1))^n$ for $p \in (0, 1/2]$ (independent of n).

Bernoulli random matrices

Basak–Rudelson (**17**): $P_n \le \exp(-cnp)$ for $p = p(n) \ge (C \ln n)/n$, moreover

$$\mathbb{P}\left\{ s_n(B_p) \le c \exp(-C \ln(1/p)/\ln(np))\, t\sqrt{p/n} \right\} \le t + \exp(-cnp),$$

where

$$s_n(M) = \inf_{|x|=1} |Mx|.$$

K. Tikhomirov (**19+**): $P_n \le (1 - p + o(1))^n$ for $p \in (0, 1/2]$ (independent of n).

Moreover, $\forall \varepsilon > 0$, $\forall n \ge n(p, \varepsilon)$,

$$\mathbb{P}\left\{ s_n(B_p) \le t\sqrt{p/n} \right\} \le C(p, \varepsilon)t + (1 - p + \varepsilon)^n.$$

Main result

L-K.T.

There is a (small) absolute constant $c > 0$ such that the following holds.
Let $0 < q < c$ be a parameter and $q < p < c$. Then,

$$P_n \leq (1 + o_q(1))\, 2n(1-p)^n.$$

Main result

L-K.T.

There is a (small) absolute constant $c > 0$ such that the following holds.
Let $0 < q < c$ be a parameter and $q < p < c$. Then,

$$P_n \leq (1 + o_q(1))\, 2n(1-p)^n.$$

Moreover, $\forall \varepsilon > 0$, $\forall n \geq n(p, \varepsilon)$,

$$\mathbb{P}\left\{ s_n(B_p) \leq t\, n^{-C_q} \right\} \leq t + (1 + o_q(1))\, 2n(1-p)^n.$$

Main result

L-K.T.

There is a (small) absolute constant $c > 0$ such that the following holds.
Let $0 < q < c$ be a parameter and $q < p < c$. Then,

$$P_n \leq (1 + o_q(1))\, 2n(1-p)^n.$$

Moreover, $\forall \varepsilon > 0$, $\forall n \geq n(p, \varepsilon)$,

$$\mathbb{P}\left\{ s_n(B_p) \leq t\, n^{-C_q} \right\} \leq t + (1 + o_q(1))\, 2n(1-p)^n.$$

Remark. It seems that using technique from LLTTY (Lytova, Tomczak-Jaegermann, Youssef + LT) papers on random regular matrices one can substitute the assumption $q < p < c$ with $\frac{C \ln n}{n} < p < c$ and to remove dependence of the constants on q.
(A $0/1$ matrix is regular if the sums of 1 in all columns and in all rows are the same — it is the adjacency matrix of a regular directed graph).

Some ideas of the proof.

It is well-understood by now that to deal with the smallest singular number one needs to split S^{n-1} into several parts and to work separately on each part.

Some ideas of the proof.

It is well-understood by now that to deal with the smallest singular number one needs to split S^{n-1} into several parts and to work separately on each part.

This idea goes back to Kashin **77**, where, in order obtain an orthogonal decomposition of ℓ_1^n, he split the sphere into two classes according to the ratio of ℓ_1^n and ℓ_2^n norms. In a similar context it was used by Schehtman **04**.

Some ideas of the proof.

It is well-understood by now that to deal with the smallest singular number one needs to split S^{n-1} into several parts and to work separately on each part.

This idea goes back to Kashin **77**, where, in order obtain an orthogonal decomposition of ℓ_1^n, he split the sphere into two classes according to the ratio of ℓ_1^n and ℓ_2^n norms. In a similar context it was used by Schehtman **04**.

Since we want to provide a lower bound on the smallest singular value of a random matrix M, we need to show that $|Mx|$ is not very small for all $x \in S^{n-1}$. Usually it is done using the union bound – to prove a good probability bound for an individual vector x and then to find a good net in order to apply approximation. The main point is to have a good balance between the probability and the cardinality of a net.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}$ be a Bernoulli random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_i e_i \rangle = \sum_i \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}$ be a Bernoulli random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_i e_i \rangle = \sum_i \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}$ be a Bernoulli random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_i e_i \rangle = \sum_i \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}$ be a Bernoulli random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_i e_i \rangle = \sum_i \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net. This leads to splitting the sphere into *compressible* vectors – those closed to sparse, and *incompressible* vectors – the rest.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}$ be a Bernoulli random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_i e_i \rangle = \sum_i \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net. This leads to splitting the sphere into *compressible* vectors – those closed to sparse, and *incompressible* vectors – the rest. For compressible vectors we have a net of small cardinality, therefore relatively poor individual probability bounds work, while incompressible vectors are well spread and therefore have very good anti-concentration properties.

Some ideas of the proof.

But vectors behave differently. Consider the following example, let $X = \{\varepsilon_i\}$ be a Bernoulli random vector with ± 1 independent entries. Then

$$\langle X, e_1 + e_2 \rangle = \varepsilon_1 + \varepsilon_2 = 0 \quad \text{with probability} \quad 1/2.$$

On the other hand,

$$\langle X, \sum_i e_i \rangle = \sum_i \varepsilon_i = 0 \quad \text{with probability at most} \quad 1/\sqrt{n}$$

by the Erdős–Littlewood–Offord anti-concentration lemma.

Usually, it is hard to get good individual bounds for vectors of small support, so-called *sparse vectors*. However, the set of such vectors is essentially of lower dimension, hence admit a very good net. This leads to splitting the sphere into *compressible* vectors – those closed to sparse, and *incompressible* vectors – the rest. For compressible vectors we have a net of small cardinality, therefore relatively poor individual probability bounds work, while incompressible vectors are well spread and therefore have very good anti-concentration properties. This approach was used in L–Pajor–Rudelson–Tomczak-Jaegermann (**05**) for rectangular matrices and was later developed in series of works by Rudelson–Vershynin.

Some ideas of the proof.

For $0/1$ matrices an additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$.

Some ideas of the proof.

For $0/1$ matrices an additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

Some ideas of the proof.

For $0/1$ matrices an additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

This leads to our splitting. The first class will be sparse vectors shifted by constants vectors. The second class will be the remaining vectors.

Some ideas of the proof.

For $0/1$ matrices an additional problem is caused by constant vectors. Indeed, while properly normalized centered random matrices (say with entries ± 1) have norm of order $\sqrt{n}$, the norm $\|B_p\| \approx pn$. Fortunately, this large norm is only in the direction of $\mathbf{1} = (1, 1, ..., 1)$. On the subspace orthogonal to $\mathbf{1}$ the norm is of the order $\sqrt{pn}$.

This leads to our splitting. The first class will be sparse vectors shifted by constants vectors. The second class will be the remaining vectors.

For the first class standard anti-concentration technique works, since the set is essentially of lower dimension (although there are many cases).

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$
2. $x^*_i \le (n/i)^L$
3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \ge h$.

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$
2. $x^*_i \le (n/i)^L$
3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \ge h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$
2. $x^*_i \leq (n/i)^L$
3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular number to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq 1$.

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$

2. $x^*_i \leq (n/i)^L$

3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular number to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq 1$.

This distance is a projection on a (random) normal vector to H_i.

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$

2. $x^*_i \le (n/i)^L$

3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \ge h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular number to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \le 1$.

This distance is a projection on a (random) normal vector to H_i.

Thus, we have an inner product of X_i and the normal (note that they are independent).

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$
2. $x^*_i \leq (n/i)^L$
3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular number to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq 1$.

This distance is a projection on a (random) normal vector to H_i.

Thus, we have an inner product of X_i and the normal (note that they are independent).

Then we apply an anti-concentration property (such a property says that an inner product of a random vector with a flat vector can't concentrate around a number).

Some ideas of the proof.

For the second class we show that it is contained in *gradual non-constant vectors*, that is, vectors (after certain normalization and for some parameters r, δ, L, h) s.t.

1. $x^*_{rn} = 1$
2. $x^*_i \leq (n/i)^L$
3. If $(y_i)_i$ is a non-increasing rearrangement of $(x_i)_i$ then $y_{\delta n} - y_{n-\delta n} \geq h$.

To work with this class we partially follow Rudelson–Vershynin scheme.

First, one reduces estimating the smallest singular number to estimating distances between a column X_i to the span of remaining columns, say H_i, $i \leq 1$.

This distance is a projection on a (random) normal vector to H_i.

Thus, we have an inner product of X_i and the normal (note that they are independent).

Then we apply an anti-concentration property (such a property says that an inner product of a random vector with a flat vector can't concentrate around a number).

To make this scheme work, Rudelson–Vershynin introduced LCD (*least common denominator*), which, in a sense, measures how close a proportional coordinate projection of a vector to the properly rescaled integer lattice. They also had to develope Littlewood–Offord theory.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Second idea is to substitute LCD with another parameter, which we call balancing degree of a vector, and which is more directly related to the Esseen lemma, used to prove an anti-concentration.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Second idea is to substitute LCD with another parameter, which we call balancing degree of a vector, and which is more directly related to the Esseen lemma, used to prove an anti-concentration.

Next we have to prove a Littlewood–Offord type anti-concentration property for this new parameter.

Some ideas of the proof.

In our case both, the LCD, and the known anti-concentration results are not strong enough, so we need to develop new tools.

First idea is to pass from a Bernoulli random vector, which may have many zeros, to a random $0/1$ vector with prescribed number of ones, say, with m ones, where m is of the order pn. Note that pn is an average number of ones in a Bernoulli vector.

Second idea is to substitute LCD with another parameter, which we call balancing degree of a vector, and which is more directly related to the Esseen lemma, used to prove an anti-concentration.

Next we have to prove a Littlewood–Offord type anti-concentration property for this new parameter.

In particular, we also extend the Littlewood–Offord theory to the case of dependent random variables (in our case – the coordinates of a vector with fixed number of ones).

Balancing degree

Recall the definition of Lévy concentration function:
$\mathcal{Q}(\xi, t) = \max_\lambda \mathbb{P}\left(|\xi - \lambda| < t\right)$.

Balancing degree

Recall the definition of Lévy concentration function:
$\mathcal{Q}(\xi, t) = \max_\lambda \mathbb{P}\left(|\xi - \lambda| < t\right).$

Esseen Lemma (**66**):

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \leq C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E}\exp(2\pi \mathbf{i} \xi_i s/\tau)|\, ds.$$

Balancing degree

Recall the definition of Lévy concentration function:
$\mathcal{Q}(\xi, t) = \max_\lambda \mathbb{P}\left(|\xi - \lambda| < t\right).$

Esseen Lemma (**66**):

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \leq C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E}\exp(2\pi \mathbf{i} \xi_i s/\tau)|\, ds.$$

For a finite integer subset S, let $\eta[S]$ denotes a r.v. uniformly distributed on S.

Balancing degree

Recall the definition of Lévy concentration function:
$\mathcal{Q}(\xi, t) = \max_\lambda \mathbb{P}\left(|\xi - \lambda| < t\right)$.

Esseen Lemma (**66**):

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \leq C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E}\exp(2\pi \mathbf{i} \xi_i s/\tau)|\, ds.$$

For a finite integer subset S, let $\eta[S]$ denotes a r.v. uniformly distributed on S. Then

$$\mathbf{Bal}_n(v, m, K) := \sup\left\{t > 0 : \frac{1}{N} \sum_{(S_1,\dots,S_m)} \int_{-t}^{t} \prod_{i=1}^{m} \left|\mathbb{E}\exp\left(2\pi \mathbf{i}\, v_{\eta[S_i]}\, m^{-1/2} s\right)\right| ds \leq K\right\},$$

where the sum is taken over all sequences $(S_i)_{i=1}^m$ of disjoint subsets $S_1, \dots, S_m \subset [n]$, each of cardinality $\lfloor n/m \rfloor$, N is the number of such sequences, $K \geq 1$ is a parameter.

Balancing degree

Recall the definition of Lévy concentration function:
$\mathcal{Q}\left(\xi, t\right) = \max_\lambda \mathbb{P}\left(|\xi - \lambda| < t\right).$

Esseen Lemma (**66**):

$$\mathcal{Q}\Big(\sum_{i=1}^{m} \xi_i, \tau\Big) \leq C' \int_{-1}^{1} \prod_{i=1}^{m} |\mathbb{E}\exp(2\pi\mathbf{i}\xi_i s/\tau)|\, ds.$$

For a finite integer subset S, let $\eta[S]$ denotes a r.v. uniformly distributed on S. Then

$$\mathbf{Bal}_n(v, m, K) := \sup\left\{t > 0 : \frac{1}{N} \sum_{(S_1,\ldots,S_m)} \int_{-t}^{t} \prod_{i=1}^{m} \left|\mathbb{E}\exp\left(2\pi\mathbf{i}\, v_{\eta[S_i]}\, m^{-1/2} s\right)\right| ds \leq K\right\},$$

where the sum is taken over all sequences $(S_i)_{i=1}^m$ of disjoint subsets $S_1, \ldots, S_m \subset [n]$, each of cardinality $\lfloor n/m \rfloor$, N is the number of such sequences, $K \geq 1$ is a parameter. We prove that

$$\mathcal{Q}\Big(\sum_{i=1}^{n} v_i X_i, \sqrt{m}\, t\Big) \leq C\left(t + 1/\mathbf{Bal}_n(v, m, K)\right) \quad \text{for all } t > 0,$$