ANTI-CONCENTRATION INEQUALITIES

MANJUNATH KRISHNAPUR
(DEPARTMENT OF MATHEMATICS, IISC.)

CONTENTS

1. About the lectures

These are notes I made for a set of five lectures I gave from 4-8 January at the the ATM workshop at IIT, Bombay. I thank the organizers Vivek Borkar, Suresh Kumar, Rajesh Sundaresan and also Mallikarjuna Rao for inviting me and for organizing the workshop. Thanks also to the other speakers for their lectures and the audience for paying attention to a topic that was certainly not of central interest to most of them. I learned much of this material with the help of Sourav Sarkar and Mokshay Madiman. It is a pleasure to acknowledge that.

These notes are not meant to be complete, either in the subject matter or the references given. A lot is borrowed from the papers referred to in the footnotes. It is almost faithful to the lectures given. Lecture 1 covered sections 2-6, Lecture 2 covered sections 7-9, Lectures 3 and 4 covered sections 10-13 and Lecture 5 covered sections 14-17.

2. A matter of scale

Let $X_n = \xi_1 + \ldots + \xi_n$ where ξ_k are i.i.d. Ber$(1/2)$ random variables[1]. If I_n is an interval centered at $\mathbf{E}[X_n] = n/2$, then $\mathbf{P}\{X_n \in I_n\}$ is close to 1 if the length of I_n is much larger than $\sqrt{n}$. If the length is $c\sqrt{n}$, then $\mathbf{P}\{X_n \in I_n\} \approx \Phi(c) - \Phi(-c)$ can be any number between 0 and 1 depending on the value of c. If the length of I_n is constant, then the probability is only of order $1/\sqrt{n}$. Decreasing the length further does not decrease the probability, since X_n has an atom of size $1/\sqrt{n}$ at $\lfloor n/2 \rfloor$. More precisely,

(1) $\mathbf{P}\{X_n \in [\frac{1}{2}n + \frac{1}{2}a\sqrt{n}, \frac{1}{2}n + \frac{1}{2}b\sqrt{n}]\} = \mathbf{P}\{Z \in [a,b]\} + o(1)$. This is central limit behaviour.

(2) $\mathbf{P}\{|X_n - \frac{1}{2}n| \geq an\} \leq Ce^{-c_a n}$ for $a > 0$ fixed. More generally, for $1 << a_n \lesssim \sqrt{n}$, we have $\mathbf{P}\{|X_n - \frac{1}{2}n| \geq a_n\sqrt{n}\} \leq Ce^{-ca_n^2}$. The two inequalities are often called *large deviation* and *moderate deviation* (or if you like, Bernstein, Chernoff, Hoeffding, etc.). These are concentration inequalities (supremely important, but not the subject of these lectures).

(3) $\mathbf{P}\{X_n \in [a, a+1)\} \leq \frac{10}{\sqrt{n}}$ for any $a \in \mathbb{R}$. To see this, observe that there is at most one integer in the interval $[a, a+1)$ and the largest atom of X_n has size $\binom{n}{\lfloor n/2 \rfloor}\frac{1}{2^n}$. By an application of Stirling's formula, this quantity is bounded by $10/\sqrt{n}$. Such inequalities that give an upper bound on the probability that can be packed into a short interval are called anti-concentration inequalities.

Exercise 1. Show that $\binom{n}{\lfloor n/2 \rfloor}\frac{1}{2^n} \sim \frac{1}{\sqrt{\pi}\sqrt{n}}$.

Moral: Everything depends on the scale at which we look. At very large scales, the random variable looks like a constant. At intermediate scales (of the order of the standard deviation), it

[1] We say $\xi \sim$ Ber$(1/2)$ if $\mathbf{P}\{\xi = 0\} = \mathbf{P}\{\xi = 1\} = \frac{1}{2}$. We say $\xi \sim$ Ber$_\pm(1/2)$ if $\mathbf{P}\{\xi = -1\} = \mathbf{P}\{\xi = 1\} = \frac{1}{2}$. Always, Z denotes a standard Gaussian random variable, having density $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$.

looks like a random variable with density. At shorter scales, the graininess of the distribution starts showing itself. At that scale the random variable has "anti-concentration".

3. Lévy's concentration function

Definition 2. If X is a real-valued random variable, define its *concentration function* $Q_X(t) := \sup_{a\in\mathbb{R}} \mathbf{P}\{|X-a| \le t\}$ for $t \ge 0$. For an $\mathbb{R}^d$-valued random vector, we define $Q_X(t) := \sup_{a\in\mathbb{R}^d} \mathbf{P}\{\|X-a\| \le t\}$ where $\|x\|$ denotes the standard Euclidean norm.

Some simple observations. Unless otherwise stated, random variables are real-valued (usually for simplicity only).

► $0 \le Q_X(t) \le 1$ and $t \mapsto Q_X(t)$ is increasing. $Q_X(0)$ is the size of the largest atom in the distribution of X.

► $Q_{\lambda X+b}(t) = Q_X(t/\lambda)$ for any $\lambda > 0$, $b \in \mathbb{R}$ and $t \ge 0$.

► For random vectors, we could define concentration function using any other norm, for example, $\|x\|_\infty := \max_{1\le i\le d} |x_i|$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$. More generally, we can define Q_X for a random variable X taking values in any normed linear space.

► If X and Y are independent, then $Q_{X+Y}(t) \le Q_X(t) \wedge Q_Y(t)$. To see this, observe that $\mathbf{P}\{|X-(a-Y)| \le t \,\big|\, Y\} \le Q_X(t)$ $a.s.$. Take expectations over X to get $\mathbf{P}\{|X+Y-a| \le t\} \le Q_X(t)$ and supremum over a to get $Q_{X+Y}(t) \le Q_X(t)$. This completes the proof (where did we use independence?).

4. Anti-concentration inequalities

An upper bound for $Q_X(t)$ is called an anti-concentration inequality. Sometimes, an upper bound for $\mathbf{P}\{|X-a| \le t\}$ for a specific $a \in \mathbb{R}$ is also referred to as an anti-concentration inequality. In general, such inequalities restrict the amount of probability that can be packed into a short interval or a small ball. They assert that the distribution is not too concentrated.

In these lectures we shall consider the following examples.

► Let X_i be independent (not necessarily identical) random variables and $S_n = X_1 + \ldots + X_n$. Most of our study will be on the concentration function of S_n.

► Let $M_n = (X_{i,j})_{i,j\le n}$ where $X_{i,j}$ are i.i.d. Bernoullis, $\mathbf{P}\{X = 0\} = \frac{1}{2} = \mathbf{P}\{X = 1\}$. Then what is $p_n = \mathbf{P}\{M_n \text{ is singular}\}$? Clearly $p_n \ge 2^{-n}$, since the first column can be zero (or the first two columns can be equal).

Open problem: For any $\lambda > \frac{1}{2}$, show that $p_n \le \lambda^n$ for large enough n.

History: Komlos: $p_n \to 0$. Kahn-Komlos-Szemeredi: $p_n \le \lambda^n$ for some $\lambda < 1$. Tao and Vu: $p_n \le (3/4)^n$. Bourgain, Vu and Wood: $p_n \le (1/\sqrt{2})^n$ (all for large enough n).

Bounding p_n can be interpreted as bounding the maximal atom of $\det(M_n)$, or of $s_n(M_n)$, the minimal singular value. The Littlewood-Offord problem (anti-concnetration for sums of independent random variables) has a direct bearing on this problem, as we shall see[2].

▶ If the entries have continuous distribution in the previous example, then M_n is non-singular with probability 1. A better formulation of the problem that does not give unfair advantage to continuous distributions is to ask for $\mathbf{P}\{s_n(M_n) \le \epsilon\}$, either for firxed ϵ or for some $\epsilon = \epsilon_n \to 0$.

▶ Let X_i be i.i.d. Let Z_n be the number of real zeros of the random polynomial $X_0 + X_1 t + \ldots + X_n t^n$. Then $\mathbf{E}[Z_n] \le C_0\sqrt{n}$. Proving this requires one to use the anti-concentration inequality for sums of independent random variables.

▶ In the previous problem, it is also true that $\mathbf{E}[Z_n] \le C_0 \log n$, but proving it requires anti-concentration inequalities proved very recently and that lie beyond the scope of these lectures. Towards the end, we shall mention this and other anti-concentration inequalities, mostly open.

5. LITTLEWOOD-OFFORD-ERDŐS

Motivated by a problem in random polynomials, Littlewood and Offord[3] showed that for any strictly positive $v_1, \ldots, v_n$ and any $t \in \mathbb{R}$, (let $\mathbf{v} = (v_1, \ldots, v_n)$ and $\langle \mathbf{v}, \mathbf{x}\rangle = v_1 x_1 + \ldots + v_n x_n$)

$$\#\{\mathbf{x} \in \{0,1\}^n : \langle \mathbf{v}, \mathbf{x}\rangle = t\} \le C.2^n.n^{-\frac{1}{2}}.\log n \tag{1}$$

and conjectured that the $\log n$ factor could be removed. This was proved spectacularly by Erdős[4]. In our language, (1) has the following equivalent formulation.

Lemma 3 (Littlewood-Offord-Erdős)**.** *Let X_i be i.i.d. Ber$(1/2)$ and let $v_i > 0$ and let $S_{\mathbf{v}} = v_1 X_1 + \ldots + v_n X_n$. Then, $Q_{S_{\mathbf{v}}}(0) \le \frac{1}{2^n}\binom{n}{\lfloor n/2 \rfloor} \le \frac{C}{\sqrt{n}}$.*

Proof. We define a partial order on $\{0,1\}^n$ by setting $\mathbf{x} \le \mathbf{y}$ if $x_i \le y_i$ for each i. An anti-chain is a subset of $\{0,1\}^n$ such no two distinct elements of which are comparable.

Sperner's lemma: The maximal size of an anti-chain in the above poset is $\binom{n}{\lfloor n/2 \rfloor}$.

For any strictly positive v_is and any $t \in \mathbb{R}$, the collection of $\mathbf{x} \in \{0,1\}^n$ such that $\sum_i v_i x_i = t$ is an anti-chain. Therefore, by Sperner's lemma such a set has at most $\binom{n}{\lfloor n/2 \rfloor}$ elements. This completes the proof of (1). ■

Exercise 4. If v_i are strictly positive real numbers, show that $Q_S(v_{\min}) \le Cn^{-1/2}$.

[2]These things and many other things we have discussed here may be found in the survey article- Nguyen, Hoi H.; Vu, Van H, *Small ball probability, inverse theorems, and applications*, Erds centennial, 409-463, Bolyai Soc. Math. Stud., **25**, Jnos Bolyai Math. Soc., Budapest, 2013.

[3]Littlewood, J. E.; Offord, A. C. *On the number of real roots of a random algebraic equation*. III. Rec. Math. [Mat. Sbornik] N.S. **12(54)**, (1943). 277-286.

[4]Erdős, P. *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51**, (1945), 898-902.

6. WHAT DID LITTLEWOOD AND OFFORD DO?

Given the strikingly beautiful, elementary and optimal proof of Erdős, the original Littlewood-Offord proof of their weaker inequality is rarely mentioned. But it is a nice lesson in how to approach a problem in analysis in three steps: (1) Consider the extreme cases, (2) Prove the theorem (usually by different methods) for each extreme case, (3) Break any case into extreme cases and invoke the results.

Assume without loss of generality that $v_1 \leq \ldots \leq v_n$. Also assume (with loss of generality), that $v_1 \geq 1$ and give a bound for $Q_{S_{\mathbf{v}}}(1)$ (or equivalently for $Q_{S_{\mathbf{v}}}(v_{\min})$).

Extreme case 1: Suppose $A \leq v_i \leq 2A$ for all i for some $0 < A < \infty$. Then, $v_i X_i$ are independent random variables of comparable magnitude and the Berry-Esséen theorem gives

$$\sup_{a,b} |\mathbf{P}\{S_{\mathbf{v}} \in [a,b] - \mathbf{P}\{\sigma_n Z \in [a,b]\}| \leq \frac{1}{\sigma_n^3} \sum_{k=1}^{n} v_k^3$$

where $\sigma_n^2 = v_1^2 + \ldots + v_n^2$. From this, it is easily deduced that $Q_{S_{\mathbf{v}}}(1) \leq \frac{8}{\sqrt{n}}$.

Extreme case 2: Suppose $1 \leq v_1 \leq \frac{1}{2}v_2 \leq \frac{1}{4}v_3 \leq \ldots \leq \frac{1}{2^{n-1}}v_n$. In this case central limit behaviour is actually false (because $v_n X_n$ is as big as $S_{\mathbf{v}}$), in the sense that the right hand side of Berry-Esséen theorem, $\sigma_n^{-3} \sum v_k^3$, is quite large. However, a much simpler argument shows that $S_{\mathbf{v}}$ takes 2^n distinct values each with probability 2^{-n} and these values are separated by at least 1 (think of binary expansion). Thus, $Q_{S_{\mathbf{v}}}(1) \leq 2^{-n+1}$.

General case: Let $k_1 = 1$ and for $j \geq 2$, let k_j be the first index (if any) such that $v_{k_{j-1}} \leq \frac{1}{2}v_{k_j}$. This gives us some indices $k_1, k_2, \ldots, k_m$. Let $T = \sum_{j=1}^{m} v_{k_j} X_{k_j}$. Then, $S_{\mathbf{v}} = T + T'$ where T and T' are independent. Therefore, by the second extreme case,

$$Q_{S_{\mathbf{v}}}(1) \leq Q_T(1) \leq 2^{-m+1}.$$

If $m \geq \log n$, then this is smaller than $2/n$ which is is better than the bound $1/\sqrt{n}$ that we are after. Otherwise, $m < \log n$ and hence there is some $2 \leq j \leq m$ such that $k_j - k_{j-1} \geq n/\log n$. Note that the v_i for $k_{j-1} \leq i \leq k_j - 1$ are all between A and $2A$ with $A = v_{k_{j-1}}$. Hence if $T_j = \sum_{i=k_{j-1}}^{k_j-1} v_i X_i$, then again $S_{\mathbf{v}} = T_j + T_j'$ where T_j and T_j' are independent, and hence by the first extreme case,

$$Q_{S_{\mathbf{v}}}(1) \leq Q_{T_j}(1) \leq 2\frac{1}{\sqrt{k_j - k_{j-1}}} \leq 2\frac{\sqrt{\log n}}{\sqrt{n}}.$$

Putting everything together, we have proved the (sub-optimal) bound[5] of $3\frac{\sqrt{\log n}}{\sqrt{n}}$. ■

[5] Actually Littlewood and Offord got $\log n/\sqrt{n}$ because they used a weaker precursor of Berry-Esséen theorem due to Lyapunov.

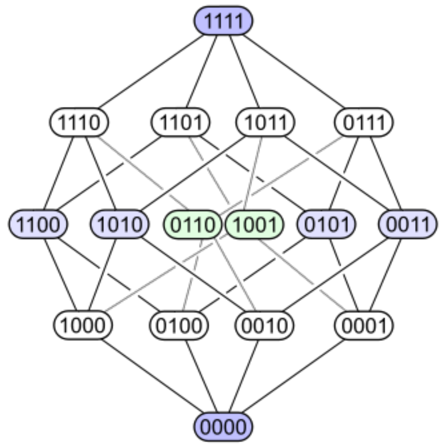

FIGURE 1. Hasse diagram of the poset $\mathcal{P}_4$ (picture taken from Wikipedia)

7. SPERNER'S LEMMA: A SIMPLE AND A NOT SO SIMPLE PROOF

Let $\mathcal{P}_n$ be the collection of all subsets of $[n]$ with the inclusion partial order. Sperner's lemma is the assertion that no anti-chain of $\mathcal{P}_n$ has more elements than the anti-chain consisting of subsets of size $\lfloor n/2 \rfloor$.

It is usually helpful to visualize a poset through its Hasse diagram. For $\mathcal{P}_n$, it may be described as follows: Let L_k be the collection of subsets with k elements, for $0 \le k \le n$. We refer to L_k as the kth layer of $\mathcal{P}_n$. If $A \in L_k$ and $B \in L_{k+1}$, then $A \le B$ if and only if B can be got from A by adding one element. If we define the order like this and extend it by transitivity, we get back the poset $\mathcal{P}_n$ with the inclusion order. The Hasse diagram is a graph with elements of L_{k+1} written above the elements of L_k and with an edge from $A \in L_k$ to $B \in L_{k+1}$ if and only if $A \le B$. See Figure 7.

First we present what may be the shortest proof of Sperner's lemma (I got this way of phrasing it from Mokshay Madiman).

Katona's proof of Sperner's lemma. On the Hasse diagram of the poset $\mathcal{P}_n$, start from $0 \dots 0$ at the bottom and do an up-ward random walk till it hits $1 \dots 1$ at the top. This means that at each step, one of the neighbours in the layer immediately above is chosen.

If $\{A_1, \dots, A_\ell\}$ is an anti-chain, let p_k be the probability of the event that the random walk passes through A_k. The anti-chain property implies that these events are pairwise disjoint and hence $\sum_{k=1}^{\ell} p_k \le 1$. By symmetry it is also clear that $p_k = 1/\binom{n}{|A_k|}$ which is at least $1/\binom{n}{\lfloor n/2 \rfloor}$. Putting both these together, we get $\ell \le \binom{n}{\lfloor n/2 \rfloor}$. ■

Katona's proof of Sperner's lemma. For each $\mathbf{x}$, let $\mathcal{S}_{\mathbf{x}}$ be the collection of all permutations of $[n]$ such that each i with $x_i = 1$ precedes each j with $x_j = 0$. If $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(\ell)}\}$ is an anti-chain, then the sets $\mathcal{S}_{\mathbf{x}^{(j)}}$, $j \le \ell$ are pairwise disjoint, and hence their cardinalities sum to at most $n!$. But if

$\mathbf{x}$ has k ones, then $\mathcal{S}_{\mathbf{x}}$ has cardinality $k!(n-k)!$ which is minimized when $k = \lfloor n/2 \rfloor$. Thus, $n! \geq \ell(\lfloor n/2 \rfloor)!(n - \lfloor n/2 \rfloor)!$ which is the same as $\ell \leq \binom{n}{\lfloor n/2 \rfloor}$. ■

Actually the above proof showed a stronger inequality known as the LYM (or LYMB inequality). If $\{A_1, \ldots, A_\ell\}$ is an anti-chain in $\mathcal{P}_n$, show that $\sum_{k=1}^{\ell} \frac{1}{\binom{n}{|A_k|}} \geq 1$ (exercise).

Now we give what may be the least straightforward proof of Sperner's lemma. Its advantage is that it generalizes to several other posets where no other method is known. This is borrowed from a beautiful paper of Proctor[6].

Stanley-Proctor proof of Sperner's lemma. For simplicity of notation, let $M = \lfloor n/2 \rfloor$.

Step -1: Sperner's lemma is implied if we show that for each $k < M$, there is an injective map $I_k : L_k \mapsto L_{k+1}$ such that $A \leq I_k(A)$ for each A. By symmetry, this implies that for $k \geq M$, there is a surjective map $I_k : L_k \mapsto L_{k+1}$.

Reason: Draw a picture of the poset with layers one above another and the maps I_k indicated by arrows upwards. Stare at it till you see that the whole poset has been broken into disjoint chains each of which passes through exactly one element of the middle layer. Hence the proof.

Step -2: Suppose that for each $k < M$ we find an $L_{k+1} \times L_k$ matrix X_k with full column rank and such that the the (B, A) entry of X_k is positive if and only if $A \leq B$. Then the injective map I_k of Step-1 exists.

Reason: X_k has full column rank and hence there exists an $L_k \times L_k$ sub-matrix of X_k whose determinant is not zero. Expand the determinant as a sum over permutations to see that at least one of these summands must be non-zero. This permutation gives the injective mapping I_k.

Step -3: Define X_k as the $L_{k+1} \times L_k$ matrix with the (B, A) entry equal to 1 is $A \leq B$ and 0 otherwise. Then X_k has full column rank.

Reason: For $k < M$, check that $X_k X_k^* - X_k^* X_k$ is the scalar matrix $\theta_k I_{L_k}$ where $\theta_k = 2k - n$. Hence, X_k must have full column rank. ■

8. Extensions of Littlewood-Offord

One can try to extend Littlewood-Offord inequality in different ways. One is to take more general distributions in place of Bernoullis. But more interesting is the relationship between the coefficients $v_1, \ldots, v_n$ and the largest atom of $S_{\mathbf{v}}$. We stay with Bernoullis for now. If $\mathbf{v}_i^* = 1$ for all i, then the Littlewood-Offord-Erdős lemma may be restated as follows:

[6] Proctor, Robert A. *Solution of two difficult combinatorial problems with linear algebra*, Amer. Math. Monthly, **89** (1982), no. 10, 721734. Richard Stanley's book on Algebraic combinatorics explains all these things in detail. After the lectures, I came to know of Zeilberger's very nice paper.

Littlewood-Offord-Erdős lemma: If v_i are strictly positive, then, (1) $Q_{S_{\mathbf{v}}}(0) \le Q_{S_{\mathbf{v}^*}}(0)$ and (2) $Q_{S_{\mathbf{v}^*}}(0) \le 10/\sqrt{n}$.

Next suppose we restrict v_i to be distinct. Then, how large can $Q_{S_{\mathbf{v}}}(0)$ be? To get a guess, let us take X_i to be i.i.d. $N(0,1)$ and investigate $Q_{S_{\mathbf{v}}}(1)$ (since there are no atoms anyway). Suppose that v_i satisfy $v_{i+1} - v_i \ge 1$ for all i (without such a condition, we can replace v_i by ϵv_i and get $Q_{S_{\mathbf{v}}}(1)$ to be as close to 1 as we want by making ϵ small!).

Exercise 5. Assume that X_i are i.i.d. $N(0,1)$. If v_i are positive and $v_{i+1} - v_i \ge 1$ for all $i \le n-1$, then show that $Q_{S_{\mathbf{v}}}(1) \le 10.n^{-3/2}$.

This suggests that even for $X_i \sim \text{Ber}(1/2)$, i.i.d., we must have $Q_{S_{\mathbf{v}}}(0) \lesssim n^{-3/2}$ for distinct v_is. This was conjectured by Erdős and Offord but it was their turn to lose a $\log n$ factor! The optimal inequality was proved by Sárközy and Szemeredi. It was shown by Stanley that the optimal choice of v_is is $1, 2, \dots, n$.

9. Stanley's theorem

Theorem 6. *Let X_i be i.i.d. Ber$(1/2)$ and let v_i be distinct positive real numbers. Let $\mathbf{v}^* = (1, 2, \dots, n)$. Then, (1) $Q_{S_{\mathbf{v}}}(0) \le Q_{S_{\mathbf{v}^*}}(0)$ and (2) $Q_{S_{\mathbf{v}^*}}(0) \le 10.n^{-3/2}$.*

We shall leave the second part as an exercise later. The proof of the first part is an absolute gem! We present the proof in the beautifully simplified form given by Proctor (the paper was referred to earlier).

Step 1: A new poset. Let $\mathcal{P}_n = \{0,1\}^n$ and define the level of $\mathbf{x} \in \mathcal{P}_n$ to be $\sum_{i=1}^n ix_i$. Levels range over all integers from 0 to $N = n(n+1)/2$. Let L_k be the set of all $\mathbf{x}$ for whose level is k, for $0 \le k \le N$. For $\mathbf{x} \in L_k$ and $\mathbf{y} \in L_{k+1}$, define $\mathbf{x} \le \mathbf{y}$ if you can get $\mathbf{y}$ from $\mathbf{x}$ by moving a 1 to right or by converting the left-most bit of $\mathbf{x}$ from 0 to 1. For example, with $n = 5$, the string $01101 \in L_{10}$ and the only elements of L_{11} that are above 01101 are 01011 and 11101.

By extending the partial order transitively, we get a poset structure on $\mathcal{P}_n$. Compared to the earlier partial order, fewer things are comparable in this poset.

Step 2: We claim that the first part of Stanley's theorem is equivalent to the statement that no anti-chain in $\mathcal{P}_n$ is larger than $L_{\lfloor N/2 \rfloor}$.

To prove this, suppose $v_1, \dots, v_n$ are distinct positive numbers and $t \in \mathbb{R}$. If $\mathbf{x} \in L_k$, $\mathbf{y} \in L_{k+1}$ and $\mathbf{x} \le \mathbf{y}$, then $\langle \mathbf{v}, \mathbf{x} \rangle$ is strictly smaller than $\langle \mathbf{v}, \mathbf{x} \rangle$ (why?). Therefore, the set of $\mathbf{x}$ such that $\langle \mathbf{v}, \mathbf{x} \rangle = t$ forms an anti-chain in $\mathcal{P}_n$. Thus, if we prove that $L_{\lfloor N/2 \rfloor}$, then is follows that

$$\#\{\mathbf{x} : \langle \mathbf{v}, \mathbf{x} \rangle = t\} \le \#\{\mathbf{x} : \langle \mathbf{v}^*, \mathbf{x} \rangle = \lfloor N/2 \rfloor\}.$$

If we divide both sides by 2^n and maximize over t we get $Q_{S_{\mathbf{v}}}(0) \le Q_{S_{\mathbf{v}^*}}(0)$.

With these two steps, the problem is reduced to finding the maximal size of an anti-chain in the poset $\mathcal{P}_n$. For simplicity of presentation, we give the proof for a modified poset defined as follows:

Let $0 \le \ell \le n$ and let $\mathcal{P}_{n,\ell} = \{\mathbf{x} \in \{0,1\}^n : \sum_{i=1}^n x_i = \ell\}$. The levels in $\mathcal{P}_{n,\ell}$ range from $N_1 = 1 + 2 + \ldots + \ell$ (achieved when all 1s are to the left) to $N_2 = (n - \ell + 1) + \ldots + n$ (achieved when all 1s are to the right). On $\mathcal{P}_{n,\ell}$, define an order in exactly the same way as in $\mathcal{P}_n$. For example, in $01101 \in \mathcal{P}_{5,3}$ has level 10 but the only element above it in L_{11} is 01011 (in general, cannot add a 1 from the left because that would take $\mathbf{x}$ out of $\mathcal{P}_{n,\ell}$).

Claim: The middle layer L_M where $M = \lfloor (N_1 + N_2)/2 \rfloor$ is a maximal anti-chain in $\mathcal{P}_{n,\ell}$.

Step -1: The claim will be proved if we show that for $k \le M$, there is an injective map $I_k : L_k \mapsto L_{k+1}$ such that $I_k(\mathbf{x}) \ge \mathbf{x}$ for each $\mathbf{x}$. Then by symmetry, for $k > M$, there is an injective map from L_k into L_{k-1} such that $I_k(\mathbf{x}) \le \mathbf{x}$.

Reason: Given these maps I_k, the poset can be decomposed into chains, each of which passes through the middle layer.

Step -2: The previous step will be achieved if for each $k < M$ we find an $L_{k+1} \times L_k$ matrix X_k such that the $(\mathbf{y}, \mathbf{x})$ entry of X_k is positive if and only if $\mathbf{x} \le \mathbf{y}$.

Reason: By injectivity, there is an $L_k \times L_k$ submatrix of X_k with non-zero determinant. Expand this determinant over permutations and note that at least one of the summands must be non-zero. The corresponding permutation gives the injective map I_k.

Step -3: Define X_k by setting the $(\mathbf{y}, \mathbf{x})$ entry to be $\sqrt{r(n-r)}$ if $\mathbf{y} = \mathbf{x} + e_{r+1} - e_r$ for some $1 \le r \le n-1$. Then, X_k is injective for $k < M$.

Reason: $X_{k-1}X_{k-1}^* - X_k^* X_k$ is an $L_k \times L_k$ matrix that is in fact equal to the scalar matrix $\theta_k I_{L_k}$ where $\theta_k = 2(k - M)$ (a straightforward calculation). Thus $X_k^* X_k = X_{k-1}X_{k-1}^* + \theta_k I_{L_k}$ is strictly positive definite for $k < M$. Thus X_k must have full column rank if $k < M$. ■

10. Further extensions: Halász's theorem

The more general idea coming from Littlewood-Offord and its extension by Sárközy-Szemeredi and Stanley is that the more the restriction on v_is, the smaller the smaller the atoms of $S_{\mathbf{v}}$. More precisely, it is the arithmetic structure of v_is that determines how concentrated $S_{\mathbf{v}}$ can be. Halász proved such a theorem[7]. We give a restricted version of his theorem where v_i are all assumed to be integers and X_i are $\text{Ber}_{\pm}(1)$. Later we shall see theorems even more powerful than Halász's full theorem.

Theorem 7 (Halász). *Suppose $v_1, \ldots, v_n$ are non-zero integers. For $k \ge 1$, let R_k be the number of solutions to the equation $\epsilon_1 v_{i_1} + \ldots + \epsilon_{2k} v_{i_{2k}} = 0$ over $1 \le i_1, \ldots, i_{2k} \le n$ and $\epsilon_j \in \{-1, 1\}$. Let X_i be*

[7] Halász, G., *Estimates for the concentration function of combinatorial number theory and probability*, Period. Math. Hungar. **8** (1977), no. 3–4, 197-211.

i.i.d. $\text{Ber}_{\pm}(1/2)$ *and let* $S_{\mathbf{v}} = v_1X_1 + \ldots + v_nX_n$. *Then,*

$$Q_{S_{\mathbf{v}}}(0) \le C_k \frac{R_k}{n^{2k+\frac{1}{2}}}.$$

Here are some examples to illustrate the power of the theorem.

▶ Suppose v_i are any non-zero integers. Then, $R_k \le n^{2k}2^k$ for any k. Thus we get the Littlewood-Offord bound $Q_{S_{\mathbf{v}}}(0) \le C_1/\sqrt{n}$ (without sharp constants, but we don't care anymore).

▶ Restrict v_i to be distinct integers. Take $k = 1$ and observe that $R_k \le 2n$ (we must take $i_2 = i_1$ and $\epsilon_2 = -\epsilon_1$). Hence, $Q_{S_{\mathbf{v}}}(0) \le C_1 n^{-3/2}$, the Sárközy-Szemeredi-Stanley bound.

▶ If we further restrict v_is so that $v_i + v_j \ne v_k + v_\ell$ unless $\{i, j\} = \{k, \ell\}$, then $R_4 \le Cn^2$ (why?) and we get $Q_{S_{\mathbf{v}}}(0) \le C_2 n^{-5/2}$.

▶ If all triple sums $v_i + v_j + v_k$ for $i < j < k$ are distinct, then $R_3 \le Cn^3$ and we get a bound of $n^{-7/2}$ for the largest atom. So on and so forth...

Exercise 8. Let $v_k = k^2$ (if convenient, take $v_k = k(k+1)$ or any quadratic you prefer). Get a tight bound for R_2.

11. The most important identity in analysis

Fourier was perhaps the first to realize that

$$\int_{-1/2}^{1/2} e^{2\pi i k\theta} d\theta = \delta_0(k) \qquad \text{for } k \in \mathbb{Z}.$$

This identity allows us to write integer identities in terms of integrals and we shall use it to prove Halász's theorem. As a simple illustration, first we sketch an estimate that is very close to what we needed in the second part of Stanley's theorem.

Claim: If X_i are i.i.d. $\text{Ber}_{\pm}(1/2)$, then $\mathbf{P}\{X_1 + 2X_2 + \ldots + nX_n = 0\} \le 10.n^{-3/2}$.

Proof. Let $T = 1X_1 + 2X_2 + \ldots + nX_n$. Using the identity of Fourier and taking expectations, we get (henceforth $I = [-\frac{1}{2}, \frac{1}{2}]$)

$$\mathbf{P}\{T = 0\} = \int_I \mathbf{E}[e^{2\pi i\theta T}]d\theta = \int_I \prod_{k=1}^n \mathbf{E}[e^{2\pi i\theta k X_k}]d\theta = \int_I \prod_{k=1}^n \cos(2\pi\theta k)\ d\theta.$$

We claim that only the integral in a neighbourhood of length $1/n$ around 0 actually contributes. For instance, if $\theta = 2/n$, then about half of the $\cos(2\pi k\theta)$ values are less than 0.9 and hence the integrand is at most 0.9^n, a very small quantity. One needs to fill in some details, but with this, we come to

$$\mathbf{P}\{T = 0\} = \int_{-1/n}^{1/n} \prod_{k=1}^n \cos(2\pi\theta k)\ d\theta + O(e^{-cn}).$$

For small u, we have $\cos(u) \approx 1 - \frac{1}{2}u^2 \approx e^{-\frac{1}{2}u^2}$. If we make that approximation, then we get the above integral to be

$$\int_{-1/n}^{1/n} \exp\{-\frac{1}{2}\sum_{k=1}^{n} 4\pi^2 k^2\theta^2\}d\theta = \int_{-1/n}^{1/n} \exp\{-\frac{2}{3}\pi^2 n^3\theta^2\}d\theta.$$

If we multiply this by $cn^{3/2}$, we get the integral of the $N(0, n^{-3})$ density integrated over an interval of length $1/n$ around the mean. That integral is bounded by 1 (since it is a probability). Thus, $\mathbf{P}\{T = 0\} \lesssim n^{3/2}$. ■

12. Fourier analysis to the rescue: A proof of Halász's theorem

The proof of Halász's theorem used analysis, namely the Fourier identity[8]. I am not aware of any purely combinatorial proof (even in the restricted setting we have of Bernoulli variables and integer v_is).

Write $I = [-\frac{1}{2}, \frac{1}{2}]$ henceforth in this section. For $S_{\mathbf{v}} = v_1X_1 + \ldots + v_nX_n$ with $v_i \in \mathbb{Z}$ (also remember that X_i takes values ± 1), we deduce that for $t \in \mathbb{Z}$,

$$\mathbf{P}\{S_{\mathbf{v}} = t\} = \mathbf{E}\left[\int_I e^{2\pi i(S_{\mathbf{v}}-t)\theta}d\theta\right] \le \left|e^{-it\theta}\int_I \prod_{k=1}^{n} \mathbf{E}[e^{2\pi i v_k X_k\theta}]d\theta\right| \le \int_I \prod_{k=1}^{n} |\cos(2\pi v_k\theta)|d\theta.$$

Since $\cos(2\pi\theta)$ vanishes exactly when θ is an integer, we can use the cosine function to measure the distance of a number from the closest integer as follows (both inequalities are elementary and left as exercise).

$$1 - 2\pi^2\theta^2 \le \cos(2\pi\theta) \le e^{-8\theta^2} \quad \text{for } \theta \in \left[-\frac{1}{2}, \frac{1}{2}\right], \tag{2}$$

Hence $|\cos(2\pi v_k\theta)| \le e^{-8\|v_k\theta\|^2}$ where $\|x\| := \min\{|x - n| : n \in \mathbb{Z}\}$ (it is not a norm!). Thus,

$$\int_I \prod_{k=1}^{n} |\cos(2\pi v_k\theta)|d\theta = \int_0^\infty |A_t| 8e^{-8t}dt$$

where $A_t = \{\theta \in I : \sum_{k=1}^{n} \|v_k\theta\|^2 \le t\}$ and $|A_t|$ is the Lebesgue measure of A_t. Since $|A_t| \le 1$ for all t, we can write

$$Q_{S_{\mathbf{v}}}(0) \le \int_0^{n/4\pi^2} |A_t|\, 8e^{-8t}dt \;+\; e^{-2n/\pi^2}.$$

Now we want to bound $|A_t|$ for $t \le n/4\pi^2$. If $\theta_i \in A_t$ then $\theta_1 + \ldots + \theta_m \in A_{m^2t}$ by triangle inequality. Using the Cauchy-Davenport[9] inequality $|A + A| \ge 2|A|$ we get $|A_t| \le \frac{1}{m}|A_{m^2t}|$. In

[8] This proof is taken piecewise and slightly edited from the book- Tao, Terence; Vu, Van H., *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, **105**, Cambridge University Press, Cambridge, 2010.

[9] *Proof:* Assume A and B are compact and translate them so that $\sup A = \inf B = 0$. Then $A + B \supseteq A \cup B$ while A and B are almost disjoint. Hence $|A + B| \ge |A| + |B|$.

particular, for any $t < \frac{n}{4\pi^2}$ we take $m = \lfloor \frac{\sqrt{n}}{2\pi\sqrt{t}} \rfloor$ and get $|A_t| \leq 100\frac{\sqrt{t}}{\sqrt{n}}|A_{n/4\pi^2}|$. It remains to bound the latter.

If $t = \frac{n}{4\pi^2}$ and $\theta \in A_t$ then by the inequality $\cos(2\pi x) \geq 1 - 2\pi^2\|x\|^2$ we see that $\sum_{k=1}^n \cos(2\pi v_k\theta) \geq \frac{n}{2}$. Thus setting $B_n^* = \{\theta \in I : \sum_{k=1}^n \cos(2\pi v_k\theta) \geq \frac{n}{2}\}$, we have the bound $|A_{n/4\pi^2}| \leq |B_{\mathbf{v}}^*|$.

Putting everything together we get

$$Q_{S_{\mathbf{v}}}(0) \leq \frac{100}{\sqrt{n}}|B_n^*| \int_0^{n/4\pi^2} \sqrt{t}8e^{-8t}dt + e^{-n/5}$$

$$\leq 1000\frac{1}{\sqrt{n}}|B_n^*| + e^{-n/5}. \tag{3}$$

From assumptions on arithmetical structure of $\mathbf{v}$ we can get precise bounds on $|B_n^*|$ by computing moments of $\sum_{k=1}^n \cos(2\pi v_k\theta)$ and using Markov's inequality.

$$\int_I \Big| \sum_{j=1}^n \cos(2\pi v_j\theta) \Big|^{2k} d\theta = \int_I \Big| \sum_{j=1}^n \left(e^{2\pi i v_j\theta} + e^{-2\pi i v_j\theta} \right) \Big|^{2k} d\theta$$

$$= \sum_{\epsilon_i=\pm 1} \sum_{j_1,\ldots j_{2k}\leq n} \int_I e^{2\pi i\theta \sum_{i=1}^{2k} \epsilon_i v_{j_i}} d\theta \; = \; \sum_{\epsilon_i=\pm 1} \sum_{j_1,\ldots j_{2k}\leq n} \mathbf{1}_{\epsilon_1 v_{j_1}+\ldots+\epsilon_{2k}v_{j_{2k}}=0}$$

$$= R_k.$$

Hence by Markov's inequality $|B_n^*| \leq R_k \frac{2^{2k}}{n^{2k}}$. The conclusion is that $Q_{S_{\mathbf{v}}}(0) \leq \frac{10^3}{n^{2k+\frac{1}{2}}} R_k + e^{-n/5}$.

13. Singularity probability of the Bernoulli matrix

Recall the Bernoulli matrix $M_n = (X_{i,j})_{i,j\leq n}$ and $p_n = \mathbf{P}\{M_n \text{ is singular}\}$. We use the Littlewood-Offord lemma to show that $p_n \to 0$ as $n \to \infty$.

Write $M_n = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots \mathbf{v}_n]$ and let $W_k = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ with $W_0 = \{0\}$ be subspaces of $\mathbb{R}^n$. Then, considering the first k for which v_k depends linearly on the previous columns, we get

$$p_n = \sum_{k=1}^n \mathbf{P}\{\dim(W_{k-1}) = k - 1 = \dim(W_k)\}.$$

Fix $k \leq n$ and suppose $\dim(W_{k-1}) = k-1$. Then we claim that W_{k-1} contains at most 2^{k-1} distinct vectors in $\{0,1\}^n$. Hence, $\mathbf{v}_k$ belongs to W_{k-1} with a probability of at most 2^{k-1-n}. Summing this over k gives a bad bound since 2^{k-1-n} is very large when $k = n$ (or even when k is close to n).

To get a decent bound for k close to n, we must use the randomness of W_{k-1} and its independence from $\mathbf{v}_k$. For simplicity, let $k = n$.

Condition on W_{n-1} and fix a vector $\eta \in W_{n-1}$. If $\mathbf{v}_n \in W_{n-1}$, then we must have $\langle \eta, \mathbf{v}_n \rangle = 0$. Writing the co-ordinates of $\mathbf{v}$ are $X_1, \ldots, X_n$, this is the same as $\sum_{i=1}^n \eta_i X_i = 0$. But the probability of this event (conditional on η) is at most $10/\sqrt{m}$ where m is the number of non-zero co-ordinates of η. Thus,

$$\mathbf{P}\{\mathbf{v}_n \in W_{n-1}\} \le \frac{10}{\sqrt{\log n}} + \mathbf{P}\{\eta \text{ has at most } \frac{1}{2}\log n \text{ non-zero co-ordinates}\}.$$

But, writing $m = \frac{1}{2}\log n$ for simplicity,

$$\mathbf{P}\{\eta \text{ has at most } \sqrt{\log n} \text{ non-zero co-ordinates}\} \le \binom{n}{m}.2^m.(1-2^{-m})^{n-1}$$

by choosing the m non-zero co-ordinates out of n, choosing the signs of η_i on these non-zero co-ordinates, and ensuring that none of the first $n-1$ columns have the same set of signs as η (or else it could not be orthogonal to η). Check that the last quantity is $o(e^{-0.1\sqrt{n}})$.

We considered only the case $k = n$. Work through the argument and use the bounds

$$\mathbf{P}\{\mathbf{v}_k \in W_{k-1}\} \le \begin{cases} 2^{-k+1} & \text{for } k \le n - \log n, \\ e^{-c\sqrt{n}} & \text{for } n - \log n < j \le n-1. \end{cases}$$

Put together to get the bound $p_n \le \frac{C}{(\log n)^{1/10}}$. ■

The bound for the singularity probability given by the above proof is a far cry from the conjectured λ^n for any $\lambda < \frac{1}{2}$.

14. Generalization of the basic Littlewood-Offord inequality

As mentioned earlier, the anti-concentration inequalities we have seen extend to general random variables (exact results such as optimality of $(1, \ldots, 1)$ and $(1, 2, \ldots, n)$ are special to the Bernoulli case, of course). Here we state the results and in the next section we shall see a lemma which is an essential ingredient in all proofs.

Theorem 9 (Kolmogorov-Rogozin inequality). *Let $X_1, \ldots, X_n$ be independent random variables and let $S = X_1 + \ldots + X_n$. Then, for any $t > 0$ and any $0 < t_i < t$, $i \le n$, we have*

$$Q_S(t) \le 100 \frac{t}{\sqrt{\sum_{i=1}^n t_i^2(1 - Q_{X_i}(t_i))}}.$$

Since the denominator is at most $t\sqrt{n}$, it is clear that the bound can never be better than $1/\sqrt{n}$, whatever be the random variables and however small t may be. In this sense, it is exactly like the original Littlewood-Offord-Erdős inequality. This bound is often attained but not always. One should only be careful to note that increasing t_i all the way up to t may not be the best idea, since $1 - Q_{X_i}(t_i)$ decreases.

There are two approaches that I know of to the Kolmogorov-Rogozin inequality. The second proof is due to Esséen and uses his inequality that we mention later. The first proof is due to Kolmogorov (with a loss of $\log n$ factors) and Rogozin (who found the inequality in its final form[10]). Since we shall not present details of the proof, we would like to mention a neat idea due to Kolmogorov of writing any distribution as a mixture of symmetric Bernoullis and hence transferring results for Bernoulli to general distributions. This isea, although simple, is quite useful and has been used in many contexts[11].

For every $a \le b$, let $\lambda_{a,b}$ denote the measure that puts mass $\frac{1}{2}$ at a and at b.

Claim 10. *Let μ be any non-degenerate probability measure on $\mathbb{R}$. Then, there exists a probability distribution ν on $\{(a,b) : a \le b\}$ with the property that $\mu(\cdot) = \int_{\mathbb{R}^2} \mathbf{1}_{a \le b} \lambda_{a,b}(\cdot)\, d\nu(a,b)$ and such that $\nu\{(a,a) : a \in \mathbb{R}\} < 1$.*

In terms of random variables, what this says is that if we pick (A,B) from the measure ν and conditional on (A,B) set $X = A$ or $X = B$ with equal probability, then the unconditional distribution of X is μ. The last condition says that ν puts a non-trivial mass on non-degenerate Bernoullis.

Proof. Let $U \sim \text{Uniform}[0,1]$ and $\xi \sim \text{Ber}(1/2)$ be independent of each other. From basic probability class we know that there is an increasing, measurable function $T : [0,1] \mapsto \mathbb{R}$ such that $T(Z) \sim \mu$ (in fact T is a sort of inverse of the distribution function of μ). Define $A = T(Z)$ and $B = T(-Z)$ and $X = A$ if $\xi = 0$ and $X = B$ if $\xi = 1$. Check that $X \sim \mu$. ∎

15. INVERSE LITTLEWOOD-OFFORD THEOREMS

Again we work with arbitrary distributions. Like in the result of Halász, the arithmetic structure of v_ks is closely related to the concentration of $S_{\mathbf{v}}$. Results are often stated by assuming that a specific lower bound on the concentration function of $S_{\mathbf{v}}$ forces a minimal amount of arithmetic structure on the v_ks. In this form, the results were named *inverse Littlewood-Offord theorems* by Tao and Vu. The first such results were dure to Arak in the 1980s[12], and more recently Tao and Vu and Rudelson and Vershynin. We just present one result in the form stated by Rudelson and Vershynin and simplified by Friedland and Sodin[13]. **The LCD measure of arithmetic structure:**

---

[10] Rogozin, B. A., *On the increase of dispersion of sums of independent random variables*, Teor. Verojatnost. i Primenen **6** 1961 106-108.

[11] For example, Aizenman, Michael; Germinet, Franois; Klein, Abel; Warzel, Simone, *On Bernoulli decompositions for random variables, concentration bounds, and spectral localization*, Probab. Theory Related Fields 143 (2009), no. 1-2, 219238.

[12] Arak's results appear to be unknown even to most experts. In recent preprints of Eliseeva-Götze-Zaitsev, the contributions of Arak as well as their relationship to the more recently published inverse Littlewood-Offord theorems are explained.

[13] On the webpages of Rudelson and of Vershynin, one can find a great many interesting papers and surveys and lecture slides on this topic. The form presented here is taken from: Friedland, Omer; Sodin, Sasha, *Bounds on the concentration function in terms of the Diophantine approximation*, C. R. Math. Acad. Sci. Paris **345** (2007), no. 9, 513-518.

Fix a parameter $\alpha > 0$ (usually taken to be $c\sqrt{n}$ for a small constant c). For a vector $\mathbf{v} \in \mathbb{R}^n$, define

$$\mathrm{Lcd}_\alpha(\mathbf{v}) := \min\left\{\theta > \frac{1}{\|\mathbf{v}\|} : \sum_{k=1}^{n} \|\theta v_k\|^2 \le \alpha^2\right\}$$

Example 11. Let $v_k = 1$ for all k. If $\frac{1}{\sqrt{n}} \le \theta \le \frac{1}{2}$, then $\sum\limits_{k=1}^{n} \|\theta v_k\|^2 = n\theta^2 \ge 1$. Hence $\mathrm{Lcd}_{\sqrt{n}}(\mathbf{v}) \ge \frac{1}{2}$.

Example 12. Let $v_k = k$. We claim that $\mathrm{Lcd}_{\sqrt{n}}(\mathbf{v}) \asymp n^{3/2}$. Suppose $\frac{1}{n^{3/2}} \le \theta \le \frac{1}{2}$.

Theorem 13 (Rudelson-Vershynin, Friedland-Sodin)**.** *Let X_k be i.i.d. random variables with $Q_{X_1}(\epsilon) \le 1 - p_\epsilon$. Let $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$. Then,*

$$Q_{S_\mathbf{v}}(\epsilon) \le \frac{10}{p_\epsilon \, \|\mathbf{v}\| \, Lcd_\alpha(\mathbf{v})} + 10.e^{-p_\epsilon^2 \alpha^2/10}.$$

For simplicity, if $p = Q_{X_1}(0)$, then we may use p in place of p_ϵ and get

$$Q_{S_\mathbf{v}}(\epsilon) \le C\left\{\frac{1}{\|\mathbf{v}\| \, \mathrm{Lcd}_\alpha(\mathbf{v})} + e^{-c\alpha^2}\right\}$$

for some constants C and c that may depend on p. In particular, on scales larger than $1/\mathrm{Lcd}_\alpha(\epsilon)$, the random variable looks like a random variable with density.

16. Characteristic functions and concentration

Fourier analysis is the standard tool to study sums of independent random variables. To go beyond ± 1 valued (or integer-valued) random variables, we cannot use Fourier series but Fourier transform or characteristic function. Recall that if X is random variable, its characteristic function if $\varphi_X(\lambda) = \mathbf{E}[e^{i\lambda X}]$. The following inequality (or similar ones) forms an essential step in almost all proofs of anti-concentration[14].

Lemma 14 (Esséen's inequality)**.** *If $X \sim \mu$ is a real-valued random variable with characteristic function ψ, then $Q_X(t) \le t \int_{-2\pi/t}^{2\pi/t} |\psi(\lambda)| d\lambda$.*

Proof. Let $T = 2\pi/L$ and consider the probability density $g(x) = \frac{T}{\pi} \frac{\sin^2(Tx/2)}{(Tx/2)^2}$ with characteristic function $\omega(\lambda) = \left(1 - \frac{|\lambda|}{T}\right)_+$ (easier to check this by Fourier inversion). If V has density g and is independent of X, then $X + V$ has characteristic function $\psi(\lambda)\omega(\lambda)$ and hence its density is given by

$$h(x) = \int_{-T}^{T} \psi(\lambda)\omega(\lambda) d\lambda \le \int_{-T}^{T} |\psi(\lambda)| d\lambda.$$

[14]Esseen, C. G., *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **5** 1966, 210-216.

But we can also use the convolution formula to write

$$h(a) = \mathbf{E}[g(a-X)] \geq \left[\min_{|t|\leq \frac{1}{2}L} g(t)\right] \mathbf{P}\{|X-a| \leq L/2\}.$$

Since $\sin u/u$ is decreasing for $u \in \frac{1}{2}\pi$, we get a lower bound of $4/\pi^2 L$ for the minimum on the right side. Thus we arrive at $Q_X(L) \leq \frac{L}{4\pi^2}\int_{-2\pi/L}^{2\pi/L} |\psi(\lambda)| d\lambda$. ■

The following exercise is a simple illustration of a use of Esséen's inequality.

Exercise 15. Let X_k be i.i.d. with characteristic function $e^{-|\lambda|^\alpha}$ where $0 < \alpha \leq 2$ (the distribution of X_1 is called the symmetric α-stable distribution). Let $S_n := X_1 + \ldots + X_n$ be the random walk with steps X_k. If $\alpha < 1$, show that $\{S_n\}$ is *transient*, i.e., $|S_n| \to \infty$ $a.s.$

17. Some questions in anti-concentration

In the lecture I mentioned the problems of singularity of *symmetric* Bernoulli matrices, the problem of quadratic Littlewood-Offord, the relative anti-concentration problem and its application to random polynomials, a permanental anti-concentration conjecture (Aaronson and Arkhipov) and that anti-concentration inequalities are used in many places such as the localization problem for random Schrodinger operator in one dimension when the potential is Bernoulli, smoothed analysis and the singularity question etc. These key words are presumably enough to find out more and read about them. I shall not write them out in detail.