THE LITTLEWOOD-OFFORD PROBLEM

KISHAN JANI

ABSTRACT. The Littlewood-Offord problem deals with finding the small ball probability, defined $\mathbf{P}_d(\sum_{i=1}^n \xi_i \mathbf{v}_i \in \mathbb{B})$, for vectors $\mathbf{v}_i \in V$ in $\mathbb{R}^d$ and i.i.d Bernoulli random variables ξ_i. This corresponds to finding the concentration of the random linear combination, which can be interpreted as a form of random walk, into a d-ball $\mathbb{B}$ of arbitrary radius. In this expository paper, we solve this problem in full generalization using combinatorial and anlalytic methods. Furthermore, we also discuss the faculty of Fourier analytic methods in proving various results pertaining to the problem, and its versatility in dealing with restrictions on the set of vectors.

1. Random Walks

A **stochastic** or **random process** is defined as a collection of random variables $\{X_t\}_{t \in T}$ indexed by a set T which is often a subset of $\mathbb{R}_{\geq 0}$ or $\mathbb{Z}_{\geq 0}$ and is usually interpreted as time. Stochastic processes are central to the field of probability theory, further entailing a wide variety of applications in the sciences by proving to be a suitable model to many natural phenomena. One particularly interesting example of a stochastic process that has been extensively studied is the random walk.

Definition 1.1. A **random walk** is a series of successive random steps taken in some mathematical space. Explicitly, for independent and identically distributed random variables X_k, a random walk is the stochastic process $\{S_n\}_{n \in \mathbb{Z}_{\geq 0}}$ with $S_0 = 0$ such that $S_k = \sum_{i=0}^k X_k$.

Example 1.1. *The Drunkard's Walk* is a famous one-dimensional example of a random walk. Consider a drunk man who is a distance d away from a cliff. If after each second he randomly either takes one step towards the cliff with a probability p or away from the cliff with the remaining probability, what are the chances of him falling off? After n steps, where is he likely to be?

Arising from the research of Littlewood and Offord on the distribution of real roots of random polynomials, the eponymous problem analyses a specific kind of random walk.

Question 1.1. Given a random walk of length n in a d-dimensional Hilbert space $\mathbb{H}_d$, characterized as a linear combination $X = \sum_{i=1}^n \xi_i \mathbf{v}_i$ where ξ_i are independent, identically distributed (i.i.d) random variables and $\mathbf{v}_i$ are vectors such that $\|\mathbf{v}_i\| \geq 1$, what is the probability $\mathbf{P}_d(n, \Delta) = \mathbf{P}(X \in \mathbb{B})$, where $\mathbb{B}$ is a d-dimensional ball of radius Δ? We refer to this as the **small ball probability**.

Intuitively, the Littlewood-Offord problem can be thought of as measuring the concentration of a random walk of length n with step sizes greater than one into a ball of radius Δ. Littlewood and Offord in [LO39] provided the first upper bound $\mathcal{N}_2(n, 1) \leq c \cdot \frac{2^n}{\sqrt{n}} \log(n)$ in a circle of radius one for complex $\mathbf{v}_i$, but through a rather involved approach. Note that $\mathcal{N}_2(n, 1)$, the number of linear combinations, leads to the probability simply by dividing by the 2^n total possible linear combinations. Erdos in [Erd45] was able to remove the $\log(n)$ term for real $\mathbf{v}_i$ by a simple yet ingenious combinatorial argument using Sperner's Theorem. Before showing Erdos's argument, we first prove the result by Sperner, which places a bound on the size of **antichains** of the set $\{1, 2, 3, \dots, n\}$.

Definition 1.2. A family $\mathscr{F}$ of subsets of the set $N = \{1, 2, \dots, n\}$ is called an **antichain** if no set of $\mathscr{F}$ contains another set in $\mathscr{F}$.

Example 1.2. The family of subsets $\mathscr{F}_1 = \{\{1,2\},\{2,3\},\{3,1\}\}$ of the set $S = \{1,2,3\}$ is an antichain of size 3. The family $\mathscr{F}_2 = \{\{1\},\{2,3\}\}$ is another antichain of size 2.

Lemma 1 (Sperner's Theorem)**.** *The antichains of an n-element set cannot be longer than $\binom{n}{\lfloor n/2 \rfloor}$*

Proof. For proving this theorem, we count the antichains with respect to chains of the form $\emptyset = C_0 \subset C_1 \subset C_2 \subset \cdots \subset C_n = N$, where $|C_i| = i$ for $1 \leq i \leq n$. We can construct such a chain by appending one element to C_k to give us the C_{k+1} subset. We have n choices for C_1, then $(n-1)$ choices for the second element in C_2, and so on, giving a total of $n!$ ways to make the chain $C_0 \subset C_1 \subset C_2 \subset \cdots \subset C_n$.
Next, let S be an element of the antichain $\mathscr{F}$. We wish to determine how many chains of the form above contain S of size k. Once again, we can start from $\emptyset$ and append elements one after the other until we reach S, and then to complete the chain, we consider the remaining combinations from S to N. The first step is equivalent to counting the permutations of the set S, and the second step involves permuting the remaining elements; as a result, the total number of chains containing S must be $k!(n-k)!$. Since $\mathscr{F}$ is an antichain, no chain can contain two elements S and T of $\mathscr{F}$.
We divide the antichain $\mathscr{F}$ by the sizes of the subsets in it. Let m_k denote the number of sets of size k, then clearly we have $|\mathscr{F}| = \sum_{k=0}^{n} m_k$. Now since each element of the antichain is contained in a total of $k!(n-k)!$ chains that do not contain any other element, the total number of chains containing some element of $\mathscr{F}$ is given by $\sum_{k=0}^{n} m_k k!(n-k)!$, which naturally cannot exceed the total number of chains $n!$. As a result, we get the inequality

$$\sum_{k=0}^{n} m_k \cdot \frac{k!(n-k)!}{n!} \leq 1, \text{ or equivalently, } \sum_{k=0}^{n} m_k \cdot \frac{1}{\binom{n}{k}} \leq 1$$

Now replacing each $\binom{n}{k}$ by the largest binomial coefficient $\binom{n}{\lfloor n/2 \rfloor}$, we obtain $\frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \sum_{k=0}^{n} m_k \leq 1$ and from that the desired result

$$|\mathscr{F}| = \sum_{k=0}^{n} m_k \leq \binom{n}{\lfloor n/2 \rfloor}$$

□

Theorem 1 (Erdos 1945)**.** *Let $v_1, v_2, \dots v_n$ be real numbers such that $|v_i| \geq 1$. Then the number of sums $\sum_{i=1}^{n} \xi_i v_i$ that lie inside an arbitrary interval I of length 2 cannot be more than $\binom{n}{\lfloor n/2 \rfloor}$.*

Proof. Without loss of generality we can assume that $v_i \geq 1$, since for $v_i < 0$, we can change v_i to $-v_i$ and ξ_i to $-\xi_i$ to give the same product $\xi_i v_i$ but for positive v_i. Let $S = \{\sum_{i=1}^{n} \xi_i v_i : \sum_{i=1}^{n} \xi_i v_i \in I\}$ be the set of linear combinations contained within an interval I of length 2. We construct sets S that are subsets of $N = \{1, 2, \dots, n\}$ for each linear combination $\sum_{i=1}^{n} \xi_i v_i$ as follows: $k \in N$ belongs to the set S if and only if $\xi_k = 1$.
We claim that the family $\mathscr{F}$ of sets S forms an antichain. Aiming for a contradiction suppose we have for $S_1, S_2 \in \mathscr{F}$ that $S_1 \subset S_2$. But then if we take the difference of the linear combinations $\sum_{i=1}^{n} \xi_i^{(1)} v_i$ and $\sum_{i=1}^{n} \xi_i^{(2)} v_i$ corresponding to S_1 and S_2 respectively, we get

$$\sum_{i=1}^{n} \xi_i^{(2)} v_i - \sum_{i=1}^{n} \xi_i^{(1)} v_i = 2 \sum_{i \in S_2 \setminus S_1} v_i \geq 2$$

which raises a contradiction. Now since $\mathscr{F}$ is an antichain, by Sperner's Theorem, its size is limited by $\binom{n}{\lfloor n/2 \rfloor}$. But by our construction, elements of the antichain lie in a one-to-one correspondence with the linear combinations that are contained in I, and as a result the number of linear combinations also cannot be more than $\binom{n}{\lfloor n/2 \rfloor}$, proving the theorem. □

In fact, in the same paper, Erdos was able to extend his result to arbitrary radii with the following theorem. We avoid the rather involved proof, but it can be found in [Erd45].

Theorem 2 (Erdos 1945)**.** *With the same notation as Theorem 2, the number of sums that lie inside an arbitrary interval I of length $2r$ cannot be more than the sum of the r largest binomial coefficients of n.*

2. Asymptotics and Stirling's Approximation

The $\binom{n}{\lfloor n/2 \rfloor}$ upper bound introduced by Erdos can be shown to be equivalent to the bound $\frac{2^n}{\sqrt{n}}$ by Stirling's Formula, a powerful tool in asymptotics and analysis that approximates the factorial. However, the derivation of this approximation is rather involved, and so we devote this section to introducing the asymptotic notation that will be used throughout the paper and proving and implementing Stirling's Approximation.

Definition 2.1. We will be using the conventional Bachmann-Landau asymptotic notation to denote various asymptotic scenarios for functions $f(x)$ and $g(x)$:

(1) **Big O Notation:** $f(x) = O(g(x))$ means that there exist C, x_0 such that $|f| \leq Cg$ for all $x \geq x_0$.
(2) **Little o Notation:** We say $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.
(3) Finally, we say $f(x) \sim g(x)$ or f is **asymptotic** to g if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Example 2.1. .

(1) The inequality in Theorem 1 $\mathcal{N}_1(n,1) \leq \binom{n}{\lfloor n/2 \rfloor}$ can be written as a bound: $\mathcal{N}_1(n,1) = O(\binom{n}{\lfloor n/2 \rfloor})$.
(2) For functions $f(x) = x^2$ and $g(x) = 2^x$, since $\lim_{x\to\infty} \frac{x^2}{2^x} = 0$, we can say that $x^2 = o(2^x)$.
(3) We can say 2^x is asymptotic to $(2^x + 2)$ since $\lim_{x\to\infty} \frac{2^x}{2^x+2} = 1$

Theorem 3 (Stirling's Approximation)**.** *The factorial of n has the following asymptotic form:*

$$n! \sim \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$$

and based on our asymptotic notation, another equivalent way of stating this theorem is

$$n! = (1 + o(1))\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Example 2.2. We can use Stirling's formula to interpret Theorem 1 as a small ball probability. The total number of linear combinations within I is no more than

$$\binom{n}{\lfloor n/2 \rfloor} = \frac{n!}{(\lfloor n/2 \rfloor !)^2} \sim \frac{\sqrt{2\pi n} \cdot (\frac{n}{e})^n}{\left[\sqrt{\pi \cdot n} \cdot (\frac{n}{2e})^{n/2}\right]^2} = \frac{\sqrt{2\pi n} \cdot (\frac{n}{e})^n}{\pi n \cdot (\frac{n}{2e})^n} = \sqrt{\frac{2}{\pi n}} \cdot 2^n \tag{1}$$

Thus, as a consequence of Stirling's formula, for some constant c, we have $\mathcal{N}_1(n,1) \leq c \cdot \frac{2^n}{\sqrt{n}}$. Dividing by the 2^n linear combinations,

$$\mathbf{P}_1(n,1) = O(n^{-1/2})$$

Stirling's Approximation will prove to be essential in some proofs in the paper as it allows us to establish a relationship between the combinatorial and analytic results. So as the final item in this section, we now provide a proof for Stirling's Formula. For proving the formula, we will manipulate the extended definition of $n!$ to obtain the desired result in the limiting case. The Dominated Convergence Theorem from real analysis will enable us to use the limiting scenario on the integral.

Definition 2.2. The Gamma Function $\Gamma(\alpha)$ is an extension of the factorial function to $\mathbb{R}$ and $\mathbb{C}$, defined by the following convergent improper integral

$$\Gamma(\alpha + 1) = \int_0^\infty x^\alpha e^{-x} dx \tag{2}$$

For $n \in \mathbb{N}$, integrating $\Gamma(n+1)$ by parts, we get the recurrence relation

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = -x^n e^{-x}\Big|_0^\infty + n\int_0^\infty x^{n-1}e^{-x} = 0 + n\Gamma(n)$$

We can calculate the initial value as $\Gamma(1) = \int_0^\infty e^{-x}dx = 1$, and then it is easy to show by induction that $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$.

Proof. For the purpose of proving Stirling's formula, we make the substitution $x = n + t\sqrt{n}$ into $\Gamma(n+1)$. For this substitution we have $dx = dt\sqrt{n}$ and upper and lower limits $t = \infty$ and $t = -\sqrt{n}$ respectively.

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = \sqrt{n}\cdot\int_{-\sqrt{n}}^\infty (n+t\sqrt{n})^n e^{-n-t\sqrt{n}}dt = \sqrt{n}\cdot\left(\frac{n}{e}\right)^n\int_{-\sqrt{n}}^\infty \left(1+\frac{t}{\sqrt{n}}\right)^n e^{-t\sqrt{n}}dt$$

We claim that as $n \to \infty$,

$$\int_{-\sqrt{n}}^\infty \left(1+\frac{t}{\sqrt{n}}\right)^n e^{-t\sqrt{n}}dt \longrightarrow \int_\infty^\infty e^{-t^2/2}dt, \tag{3}$$

where the RHS of (3) is the Gaussian integral, which will give us the missing $\sqrt{2\pi}$ term and complete our proof. However, note that the commutation of the limit and the integral is not necessary, and in some cases can lead to a contradictory result. Consequently, we need to take a more rigorous approach in proving that we can apply the limit inside the integral, for which we will require the following theorem from real analysis.

Theorem 4 (Dominated Convergence Theorem)**.** *Suppose we have a sequence $\{f_n(t)\}$ of functions that are integrable over $\mathbb{R}$. Additionally, suppose we have that $\lim_{n\to\infty} f_n(t) = f(t)$ for an absolutely integrable function f. Then if f_n is dominated by an absolutely integrable function g i.e, $f_n(t) \leq g(t)$ for all n and t, then*

$$\lim_{n\to\infty}\int_a^b f_n(t)dt = \int_a^b f(t)dt$$

A more rigorous and proper statement of this theorem in terms of Lebesgue integrals can be found in Chapter 4 of [RF10] as the Lebesgue Dominated Convergence Theorem. To implement this theorem, we first define the function

$$f_n(t) := \begin{cases} 0 & \text{if } x \leq -\sqrt{n} \\ \left(1+\frac{t}{\sqrt{n}}\right)^n e^{-t\sqrt{n}} & \text{if } x \geq -\sqrt{n} \end{cases} \text{ and } f(t) = e^{-t^2/2}$$

Limit of the sequence f_n: First, we need to prove that $\lim_{n\to\infty} f_n(t) = f(t)$. Taking the natural logarithm of $f_n(t)$, we obtain $\ln(f_n) = n\ln\left(1+\frac{t}{\sqrt{n}}\right) - t\sqrt{n}$. Next, considering the limit, we see that

$$\lim_{n\to\infty} n\ln\left(1+\frac{t}{\sqrt{n}}\right) - t\sqrt{n} = \lim_{n\to\infty}\left(n\Big[\frac{t}{\sqrt{n}} - \frac{t^2}{2n} + O\big(\frac{t^3}{n\sqrt{n}}\big)\Big] - t\sqrt{n}\right) = \lim_{n\to\infty}\big(-\frac{t^2}{2} + O(t^3/n\sqrt{n})\big) = -t^2/2$$

from which we can get the desired result by exponentiating.

Domination by a function g: To satisfy the second criterion for the dominated convergence theorem, we consider the following integrable function $g(t) \geq 0$:

$$g(t) := \begin{cases} e^{-t^2/2} & \text{if } t < 0 \\ (1+t)e^{-t} & \text{if } t \geq 0 \end{cases}$$

To prove that $g(t) \geq f_n(t)$ for all n and t, we consider the following set of cases:

(1) $t \leq -\sqrt{n}$: This case follows directly from our definition of $f_n(t)$ and $g(t)$, since $f_n(t) = 0$ over this interval.

(2) $-\sqrt{n} \leq t \leq 0$: Because the natural logarithm is a monotone increasing function, $\ln(x_1) \geq \ln(x_2)$ if and only if $x_1 \geq x_2$. Taking the logarithm of functions f_n and g,

$$d_n = \ln(f_n) - \ln(g) = n\ln\left(1 + \frac{t}{\sqrt{n}}\right) - t\sqrt{n} + \frac{t^2}{2}$$

It suffices to show that $d_n \leq 0$ in the given interval. Clearly, $d_n(0) = 0$. Consequently, our approach will be to show that d_n is increasing on the interval $-\sqrt{n} \leq t \leq 0$, since that implies that d_n takes negative values over it. Taking the derivative,

$$d_n'(t) = \frac{\sqrt{n}}{1 + t/\sqrt{n}} + t - \sqrt{n} = \frac{t^2}{t + \sqrt{n}} > 0 \text{ for } -\sqrt{n} < t < 0$$

Now since $\ln(f_n) - \ln(g) \leq 0$, the desired result follows.

(3) $t \geq 0$: We take a similar approach over this interval and consider the difference of logarithms

$$d_n(t) = n\ln\left(1 + \frac{t}{\sqrt{n}}\right) - t\sqrt{n} - \ln(t+1) + t$$

In this case, we have $d_n(0) = 0$, and so it is sufficient to show that $d_n(t)$ is decreasing over the interval $(0, \infty)$ to prove that $f_n(t) \leq g(t)$. Differentiating both sides,

$$d_n'(t) = \frac{\sqrt{n}}{1 + t/\sqrt{n}} + (1 - \sqrt{n}) - \frac{1}{t+1} = \frac{(1-\sqrt{n})t^2}{(t+1)(t+\sqrt{n})} < 0 \text{ for } t > 0, n > 1$$

In the case that we have $n = 1$, we have $f_1 - g = 0$, and so in general $f_n(t) \leq g$ holds true for all n and t.

The conditions $lim_{n\to\infty} f_n(t) = f(t)$ and that $f_n(t) \leq g(t)$ for all n allow us to apply the dominated convergence theorem. In particular, that gives us

$$\lim_{n\to\infty} \frac{\Gamma(n+1)}{\sqrt{n}(\frac{n}{e})^n} = \lim_{n\to\infty} \int_0^\infty (1 + \frac{t}{\sqrt{n}})^n e^{-t\sqrt{n}} dt = \int_{-\infty}^\infty e^{-t^2/2} dt \tag{4}$$

The last integral $I = \int_\infty^\infty e^{-t^2/2} dt = 2\int_0^\infty e^{-t^2/2} dt$ is the Gaussian integral, a well-known integral in probability and statistics that bears many applications. The evaluation of this integral, which we perform in terms of variables x and y for practical reasons, relies on its property that

$$I^2 = \left(2\int_0^\infty e^{-x^2/2} dx\right)^2 = 4\int_0^\infty e^{-x^2/2} dx \int_0^\infty e^{-y^2/2} dy = 4\int_0^\infty e^{-\frac{1}{2}(x^2+y^2)} dy dx$$

Now we can convert to polar coordinates by the substitution $x^2 + y^2 = r^2$ and $dxdy = rdrd\theta$, yielding

$$I^2 = 4\int_0^{\pi/2}\int_0^\infty re^{-1/2r^2} drd\theta = 4\int_0^{\pi/2} -e^{-1/2r^2}\Big|_0^\infty d\theta = 4\int_0^{\pi/2} 1 d\theta = 4 \cdot \pi/2 = 2\pi$$

giving $I = \sqrt{2\pi}$ since $e^{-x^2/2} > 0$ for all x. In combination with (4), we get

$$\lim_{n\to\infty} \frac{\Gamma(n+1)}{\sqrt{n}(\frac{n}{e})^n} = \sqrt{2\pi}, \text{ or } \lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}(\frac{n}{e})^n} = 1 \tag{5}$$

which corresponds the desired asymptotic $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$, and finally finishes our proof of Stirling's approximation. □

An interesting feature of this proof is the choice of substitution made. Setting $x = n + t\sqrt{n}$ may seem arbitrary, rewriting it as $t = \frac{x-n}{\sqrt{n}}$ motivates the probability theoretic reasoning behind this choice of substitution. For a sequence of random variables X_n with mean n and variance $\sqrt{n}$, the central limit theorem states that $Z = \frac{X_n - n}{\sqrt{n}}$ becomes similar in distribution to the standard normal distribution in the asymptotic case. This approach leads to a rather short proof of Stirling's Approximation as shown in [Kha74] and [Won77] , although both of these methods use advanced concepts in probability. This proof, derived from [Con16], circumvents these probability theoretic aspects as we make the substitution $x = n + t\sqrt{n}$.

3. Generalizations of the Littlewood-Offord Problem

In his 1945 paper [Erd45] where he improved the small ball probability bound to $1/\sqrt{n}$ for real numbers, Erdos conjectured that his result could be extended to complex numbers and even to arbitrary unit balls in Hilbert Spaces $\mathbb{H}_d$ of any dimension. Klietman and Katona in [Kle65] and [Kat66] respectively managed to prove the complex number case; but their approach was difficult to extend to higher dimensions. It was due to an innovative method by Klietman in [Kle70] that Erdos conjecture was finally proved in full generalization.

Theorem 5 (Klietman 1970)**.** *Let X represent the linear combination $\sum_{i=1}^{n} \xi_i \mathbf{v}_i$ for i.i.d Bernoulli random variables η_i and vectors $\mathbf{v}_i \in \mathbb{H}_d$ such that $\|\mathbf{v}_i\| \geq 1$. Additionally, consider the region $\bigcup_{i=1}^{k} R_i$ such that $R_i \in \mathbb{H}_d$ and $|\mathbf{x}-\mathbf{y}| < 2$ for all $\mathbf{x}, \mathbf{y} \in R_i$. Then the number of linear combinations in this region is no more than the sum of the k largest binomial coefficients of n.*

Proof. Klietman's proof hinges on the recursion of binomial coefficients via Pascal's formula, which gives the relation $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. We set $r = \lfloor \frac{n-k+1}{2} \rfloor$ and $s = \lfloor \frac{n+k-1}{2} \rfloor$ so that $\binom{n}{r}, \binom{n}{r+1}, \ldots \binom{n}{s}$ represent the k largest binomial coefficients of n. Using the aforementioned recursion,

$$\sum_{i=r}^{s} \binom{n}{i} = \sum_{i=r}^{s} \binom{n-1}{i} + \sum_{i=r}^{s} \binom{n-1}{i-1}$$

Replacing i by $i+1$, we get

$$= \sum_{i=r}^{s} \binom{n-1}{i} + \sum_{i=r-1}^{s-1} \binom{n-1}{i} = \sum_{i=r-1}^{s} \binom{n-1}{i} + \binom{n-1}{r} + \sum_{i=r-1}^{s-1} \binom{n-1}{i}$$

$$\text{and so } \sum_{i=r}^{s} \binom{n}{i} = \sum_{i=r-1}^{s} \binom{n-1}{i} + \sum_{i=r}^{s-1} \binom{n-1}{i} \tag{6}$$

This last relation suggests implementing an inductive step for solving the problem since the sum of the i largest binomial coefficients of n is the same as the sum of the $(k+1)$ and $(k-1)$ binomial coefficients of $(n-1)$.

Without any loss of generality, we can assume that R_i are disjoint regions. The base case $n = 1$ is trivial. Suppose the theorem holds true for $n-1$, i.e the number of linear combinations $\sum_{i=1}^{n-1} \xi_i \mathbf{v}_i$ lying in the union of k regions is given by the sum of the k largest binomial coefficients of $(n-1)$. Keeping in mind (6), our approach will be to show that linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ that lie in k disjoint regions exhibit a one-to-one correspondence with the linear combinations lying in $(k+1)$ or $(k-1)$ regions.

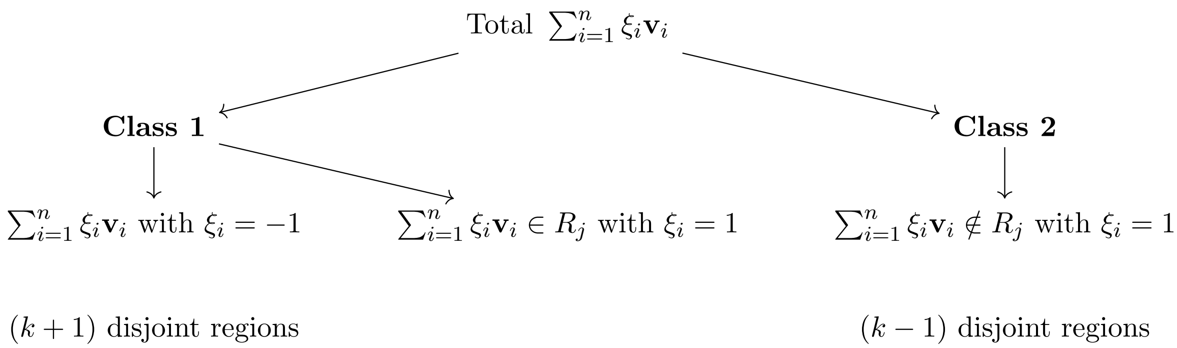
We know that $\sum_{i=1}^{n-1} \xi_i \mathbf{v}_i \in R_i$ for some i. Then we can translate the region by $\pm \mathbf{v}_i$, based on the value of ξ_n, to assert that $\sum_{i=1}^{n} \xi_i \mathbf{v}_i \in R_i \pm \mathbf{v}_n$. To establish a one-to-one correspondence, we first consider the following lemma:

Lemma 2. *Upon translating regions $R_1, R_2, \ldots R_k$ by $-\mathbf{v}_n$, at least one of these regions $R_j - \mathbf{v}_n$ is disjoint from the same regions under translation by $\mathbf{v}_i$, that is the regions $R_1 + \mathbf{v}_n, R_2 + \mathbf{v}_n, \ldots R_k + \mathbf{v}_n$*

Postponing the proof of this lemma to the end, note that as a consequence we can bijectively map linear combinations $\sum_{i=1}^{n} \xi_i \mathbf{v}_i$ into one of two classes as desired.

Class 1: Into Class 1 we add linear combinations $\sum_{i=1}^{n} \xi_i \mathbf{v}_i$ with $\xi_n = -1$, as well as linear combinations $\sum_{i=1}^{n}\xi_i\mathbf{v}_i$ with $\xi_n = 1$ that lie in the region R_j. For the first set of linear combinations, since $\sum_{i=1}^{n}\xi_i\mathbf{v}_i = \sum_{i=1}^{n-1}\xi_i\mathbf{v}_i - \mathbf{v}_n \in R_i$ for some $1 \le i \le k$ by assumption, it follows by translation that $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i \in (R_i + \mathbf{v}_n)$ for some i. For the second set of linear combinations in this class, we will have $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i + \mathbf{v}_n \in R_j$, or $\sum_{i=1}^{n-1} xi_i\mathbf{v}_i \in R_j - \mathbf{v}_n$. Consequently, the linear combinations $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i$ associated with elements in class 1 lie in $(k+1)$ disjoint regions $R_1 + \mathbf{v}_n, R_2 + \mathbf{v}_n, \ldots, R_n + \mathbf{v}_n$ and $R_j - \mathbf{v}_n$.

Class 2: Into the second class we admit the remaining linear combinations $\sum_{i=1}^{n}\xi_i\mathbf{v}_i$ with $\xi_n = 1$ that are not in R_j. Since $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i + \mathbf{v}_n \in R_i$ for some $i \neq j$, it follows that $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i \in R_i - \mathbf{v}_n$, and so combinations $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i$ associated with elements of class 2 lie in $(k-1)$ disjoint regions $R_1 - \mathbf{v}_n, \ldots, R_k - \mathbf{v}_n$ excluding $R_j - \mathbf{v}_n$.



Thus, we have managed to set up a one-to-one correspondence that associates with every linear combination of the form $\sum_{i=1}^{n}\xi_i\mathbf{v}_i$ a linear combination either in class 1 lying in the union of $(k+1)$ regions, or in class 2, lying in the union of $(k-1)$ regions. Representing $\mathcal{N}(S)$ as the number of elements, we have

$$\mathcal{N}\left(\sum_{i=1}^{n}\xi_i\mathbf{v}_i \in \bigcup R_i\right) = \mathcal{N}(\text{Class 1}) + \mathcal{N}(\text{Class 2}) \tag{7}$$

However, by our induction hypothesis, the number of linear combinations $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i$ corresponding to elements of class 2 lying in $(k+1)$ regions can be at most $\sum_{i=r-1}^{s}\binom{n-1}{i}$. Similarly, the number of the linear combinations $\sum_{i=1}^{n-1}\xi_i\mathbf{v}_i$ corresponding to elements of class 1 lying in $(k-1)$ regions can be at most $\sum_{i=r}^{s-1}\binom{n-1}{i}$. Finally, due to the bijectivity of correspondence in both classes, it thus follows from (6) and (7) that

$$\mathcal{N}\left(\sum_{i=1}^{n}\xi_i\mathbf{v}_i \in \bigcup R_i\right) \le \sum_{i=r-1}^{s}\binom{n-1}{i} + \sum_{i=r}^{s-1}\binom{n-1}{i} = \sum_{i=r}^{s}\binom{n}{i}$$

which is the desired result. □

Proof of Lemma 2. The final step in our proof is to prove Lemma 2, which comes as an application of the Cauchy-Schwartz inequality. Consider the hyperplane $\mathcal{H} = \{\mathbf{x} : \langle\mathbf{v}_n, \mathbf{x}\rangle = c\}$ that is orthogonal to $\mathbf{v}_n$, where $\langle\cdot,\cdot\rangle$ represents the inner product of vectors. Consider $\mathcal{H}$ to also contain all translated regions $R_i + \mathbf{v}_n$ in the region defined as $\langle\mathbf{v}_n, \mathbf{x}\rangle \ge c$ and touches some region $R_j + \mathbf{v}_n$, which is possible due to the boundedness of our regions. We wish to show that $R_j - \mathbf{v}_n$ lies on the other side of $\mathcal{H}$.

Aiming for a contradiction, suppose for some $\mathbf{x} \in R_j$ that $\langle\mathbf{v}_n, \mathbf{x} - \mathbf{v}_n\rangle \ge c$, or equivalently, $\langle\mathbf{v}_n, \mathbf{x}\rangle \ge \|\mathbf{v}_n\|^2 + c$. Next let $\mathbf{y} + \mathbf{v}_n$ be the point where the hyperplane $\mathcal{H}$ and the region $R_j + \mathbf{v}_n$ touch. Then $\mathbf{y} \in R_j$ and satisfies $\langle\mathbf{v}_n, \mathbf{y} + \mathbf{v}_n\rangle = c$. This can also be expressed as $\langle\mathbf{v}_n, \mathbf{y} + \mathbf{v}_n\rangle =$

$\langle \mathbf{v}_n, \mathbf{y}\rangle + \langle \mathbf{v}_n, \mathbf{v}_n\rangle = c$, or $\langle \mathbf{v}_n, -\mathbf{y}\rangle = \left\|\mathbf{v}_n^2\right\| - c$. Combining this equation with the inequality obtained for $\mathbf{x}$, we get

$$\langle \mathbf{v}_n, \mathbf{x} - \mathbf{y}\rangle \geq 2\|\mathbf{v}_n\|^2$$

and then by the Cauchy-Schwartz inequality, which states that $\langle \mathbf{a}, \mathbf{b}\rangle \leq \|\mathbf{a}\| \cdot \|\mathbf{b}\|$, we see that

$$2\|\mathbf{v}_n\|^2 \leq \langle \mathbf{v}_n, \mathbf{x} - \mathbf{y}\rangle \leq \|\mathbf{v}_n\| \cdot \|\mathbf{x} - \mathbf{y}\|, \text{ or } \|\mathbf{x} - \mathbf{y}\| \geq 2$$

However, this raises a contradiction, since because $\mathbf{x}, \mathbf{y}$ were in the closure of R_j, it must hold that $\|\mathbf{x} - \mathbf{y}\| \leq 2$, and so our lemma is proved. □

The small ball probability is obtained when $k = 1$. The largest binomial coefficient is $\binom{n}{\lfloor n/2 \rfloor}$, which proves Erdos conjecture that the small ball probability generalizes to arbitrary dimension d, i.e, $\mathbf{P}_d(n, 1) = O\left(\frac{1}{\sqrt{n}}\right)$. Another interesting feature to note is that the bound is exact since equality is achieved if we consider vectors

$$\mathbf{v}_1 = \cdots = \mathbf{v}_n = \mathbf{v} = (1, 0, 0, \ldots, 0)^T$$

For even n, $\binom{n}{n/2}$ total linear combinations sum to 0, $\binom{n}{n/2+1}$ sum to $-2\mathbf{v}$, $\binom{n}{n/2-1}$ sum to $2\mathbf{v}$, and so on. If we have a ball of radius 1 containing vectors

$$-2\left\lfloor \frac{k-1}{2} \right\rfloor, \ldots - 2\mathbf{v}, \mathbf{0}, 2\mathbf{v}, \ldots 2\left\lfloor \frac{k-1}{2} \right\rfloor$$

then the sum of the total number of linear combinations that sum to these vectors is naturally given by the k largest binomial coefficients, namely

$$\binom{n}{r}, \ldots, \binom{n}{n/2-1}, \binom{n}{n/2}, \binom{n}{n/2+1}, \ldots, \binom{n}{s}$$

with a very similar result holding for odd numbers as well.

Following Klietman's comprehensive derivation of small ball probabilities for the unit d-ball, the next natural step was to tackle the problem for an arbitrary radius Δ. However, under this scenario the problem saw a substantial increase in difficulty. After the relatively strong inequality $\mathbf{P}_d(n, \Delta) \leq 2^d \lceil \Delta\sqrt{d} \rceil \cdot O(\frac{1}{\sqrt{n}})$ due to [Gri80] and [Sal83], an extremely significant breakthrough was realized by Frankl and Füredi in [FF88], giving a much better bound.

Theorem 6 (Frankl, Füredi 1988)**.** *The small ball probability for dimension d can be expressed as*

$$\mathbf{P}_d(n, \Delta) \leq (1 + o(1))2^{-n} S(n, s)$$

where $S(n, k)$ is the sum of the k largest binomial coefficients and $s := \lfloor \Delta \rfloor + 1$.

With Frankl and Füredi's result, a natural question to ask is whether we can remove the $o(1)$ term. Unfortunately, the result does not always hold true, as one can construct a counter-example for $s \geq 2$ and $\Delta > \sqrt{(s-1)^2 + 1}$ as follows:

Example 3.1. Suppose we have vectors $\mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_{n-1} = \mathbf{e}_1$ and $\mathbf{v}_n = \mathbf{e}_2$ for orthogonal unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. We consider a ball $\mathbb{B}$ of radius $\Delta > \sqrt{(s-1)^2 + 1}$ centered at $\mathbf{v} = (\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_n)/2$. We assume that n and s have the same parity. Now suppose we choose $\frac{n-s}{2} + k$ for $k > 0$ of the $(n-1)$ total ξ_i coefficients of $\mathbf{e}_1$ to be negative, and the rest positive. Then we get

$$\sum_{i=1}^{n} \xi_i \mathbf{v}_i = -\left(\frac{n-s}{2} + k\right)\mathbf{e}_1 + \left(n - 1 - \frac{n-s}{2} - k\right)\mathbf{e}_1 + \xi_n \mathbf{e}_2 = (s - 1 - 2k)\mathbf{e}_1 + \xi_n \mathbf{e}_2$$

Consequently, we see that the vector obtained by this linear combination is within the ball centered at $\mathbf{v} = \frac{(n-1)\mathbf{e}_1 + \mathbf{e}_2}{2}$ as long as $0 \leq k \leq s$, which means that we can choose between $\frac{n-s}{2}$ and $\frac{n+s}{2}$ of the $(n-1)$ total ξ_i to be negative to obtain a vector that is still within $\mathbb{B}$. So then the probability of a linear combination being inside $\mathbb{B}$ becomes

$$\mathbf{P}_d(n, \Delta_{\mathbf{v}}) = 2 \sum_{\frac{n-s}{2} \le i \le \frac{n+s}{2}} \binom{n-1}{i} / 2^n > 2^{-n} S(n,s)$$

Keeping in mind this counter-example, [FF88] conjectured that Klietman's result could be generalized for sufficiently large n and appropriately chosen Δ.

Conjecture 1. *For $n \ge n_0(d,\Delta)$, if $s-1 \le \Delta \le \sqrt{(s-1)^2+1}$, then*

$$\mathbf{P}_d(n, \Delta) \le 2^{-n} S(n,s)$$

We can show that the condition holds when $\Delta \in (n, \sqrt{n^2+1}]$ for $n \in \mathbb{N}$; this implies that the interval over which this theorem becomes smaller for large values of n since $n \sim \sqrt{n^2+1}$. Consequently, while the conjecture provides a better bound on $\mathbf{P}_d(n,\Delta)$, it is more restrictive in nature and fails to work in a general setting in comparison to Theorem 6. The conjecture remained unsolved until Tao and Vu in [TV12] managed to provide a short proof for the conjecture in full generality using the following theorem, stated below along with its contrapositive.

Theorem 7 (Tao, Vu 2012)**.** *Let $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ denote a multiset of vectors in $\mathbb{R}^d$ and let X_V represent linear combinations $\sum_{i=1}^n \xi_i \mathbf{v}_i$. Additionally, let V have the property that for any hyperplane $\mathcal{H}$, dist$(\mathcal{H}, \mathbf{v}_i) \ge 1$ for atleast k of the total n vectors $\mathbf{v}_i$. Then for any unit ball $\mathbb{B}$,*

$$\mathbf{P}_d(X_V \in \mathbb{B}) = O(k^{-d/2})$$

Theorem 8 (Contrapositive)**.** *Suppose we have $\mathbf{P}_d(X_V \in \mathbb{B}) = \Omega(k^{-d/2})$ for a unit ball $\mathbb{B}$. Then there must exist a hyperplane $\mathcal{H}$ such that dist$(\mathcal{H}, \mathbf{v}_i) \ge 1$ for atleast k of the total n vectors $\mathbf{v}_i$.*

The proof for Theorem 8 and its contrapositive are deeply involved in advanced Fourier analytic techniques, and so we refrain from providing them here. See [TV12] and [NV13] for a rigorous treatment of these theorems. Now, with this powerful theorem, we can now prove both Theorem 7 and Conjecture 1.

Proof of Theorem 7. Aiming for a contradiction, suppose we have for arbitrarily large n and some $\Delta > 0$ that for a ball of radius Δ,

$$\mathbf{P}_d(X_V \in \mathbb{B}) \ge (1+\varepsilon) 2^{-n} S(n,s) \tag{8}$$

Next, we apply Stirling's approximation to $S(n,s)$. We assume n and s are even since these will not affect our final result, which is rather weak but sufficient. For the largest binomial coefficient, we have $\binom{n}{\lfloor n/2 \rfloor} = O(1/\sqrt{n})$, and since the number of largest binomial coefficients s is taken independent of the arbitrarily large n, we have

$$\sum_{k=-s/2}^{s/2} \binom{n}{n/2+k} 2^{-n} = \sum_{k=-s/2}^{s/2} \frac{n!}{(n/2+k)!(n/2-k)!} 2^{-n} > \frac{n!}{((n/2)!)^2} \cdot 2^{-n} \sim \frac{1}{\sqrt{n}}$$

and as a result we can conclude that

$$\mathbf{P}_d(X_V \in \mathbb{B}) > 1/\sqrt{n}$$

Now we can apply the pigeonhole principle to assert that inside our ball $\mathbb{B}$, we can find a smaller ball $\mathbb{B}_0$ of radius $1/\log(n)$ such that

$$\mathbf{P}_d(X_V \in \mathbb{B}_0) > \frac{1}{n^{1/2} \log^d n}$$

since each ball $\mathbb{B}_0$ would have volume proportional to $(1/\log(n))^d$. Now we wish to express this in the form of Theorem 8 as $Ck^{-d/2}$ for some constant C and appropriately defined k. So then our task reduces to finding appropriate k such that $n^{-1/2}\log^{-d}(n)$ dominates $k^{-d/2}$ in the asymptotic scenario. Looking at the quotient of functions, we get

$$\frac{n^{-1/2}\log^{-d}(n)}{k^{-d/2}} = \left(\frac{n^{\frac{-1}{2d}}\cdot\frac{1}{\log(n)}}{k^{-1/2}}\right)^d = \frac{\sqrt{k}\cdot n^{-1/2d}}{\log(n)} \tag{9}$$

Now since n^a dominates $\log(n)$ for all $a>0$, we define $k := n^a$ such that we get a positive exponent of n in the numerator that can dominated the $\log(n)$. From (9) we must have $\frac{a}{2}-\frac{1}{2d}>0$, and since $d\geq 2$, the smallest fractional value of a for which the inequality holds true is $a=2/3$. So then by defining $k := n^{2/3}$, the desired asymptotic result for large n is obtained.

$$\mathbf{P}_d(X_V\in\mathbb{B}_0)=\Omega(k^{-d/2})$$

Then by applying a scaled version of Theorem 9 to account for the $1/\log(n)$ radius of $\mathbb{B}_0$, we confirm the existence of a hyperplane $\mathcal{H}$ such that $\mathrm{dist}(\mathcal{H},\mathbf{v}_i)\leq 1$ for atleast $n-k$ vectors $\mathbf{v}_i$. By conditioning on the sign of ξ_i for the remaining k vectors and projecting the sum X_V onto the hyperplane $\mathcal{H}$ by a map ψ, we can conclude based on (8) that there must exist a $d-1$ dimensional ball $\mathbb{B}'$ for which

$$\mathbf{P}_{d-1}(X_{\psi(V)}\in\mathbb{B}')\geq(1+\varepsilon)2^{-n}S(n,s)$$

But each of the orthogonally projected set of vectors $\psi(V)$ must have magnitude atleast $1-\frac{1}{\log(n)}$. But for sufficiently large n, this contradicts the induction hypothesis: we first rescale $\psi(\mathbf{v}_i)$ by $\frac{1}{1-1/\log(n)}$ to make the magnitude of our vectors atleast one. Next, notice that we can identify $\mathcal{H}$ with $\mathbb{R}^{d-1}$; however, under these circumstances we can apply a scaled variant of Theorem 6. for dimension $d-1$, immediately giving a contradiction. □

Proof of Conjecture 1. We only need to prove for $s>2$, since the lower case matches Theorem 5. Aiming for a contradiction, suppose we have for arbitrarily large n and some $\Delta>0$ that for a ball of radius Δ,

$$\mathbf{P}_d(X_V\in\mathbb{B})\geq 2^{-n}S(n,s) \tag{10}$$

We can iterate the process outlined in the proof of Theorem 6. to construct a descent argument, which allows us to obtain a one-dimensional subspace $\mathcal{L}$ of the original $\mathbb{R}^d$ for which now atleast $n-O(k)$ of the vectors have $\mathrm{dist}(\mathcal{L},\mathbf{v}_i)<1/\log(n)$. By rearranging, we can have $\mathrm{dist}(\mathcal{L},\mathbf{v}_i)<1/\log(n)$ for the first $n-k_1$ vectors $\mathbf{v}_i$, where $k_1 := O(k)$.
Now we consider the orthogonal projection $\varphi:\mathbb{R}^d\to L$ and divide it into two distinct cases:

1. In the case that $|\varphi(\mathbf{v}_i)|>\Delta/s$ for all i, we first use the bound

$$\mathbf{P}_d(X_V\in\mathbb{B})\leq\mathbf{P}_1(X_{\varphi(V)}\in\varphi(\mathbb{B})), \tag{11}$$

which stems from the fact that increasing the dimensionality of the small ball probability leads to a decrease in the probability. Now on the left-hand side we can use Theorem 3 after rescaling by s/Δ to match the $|v_i|\geq 1$ criterion. This gives

$$\mathbf{P}_1(X_{\varphi(V)}\in\varphi(\mathbb{B}))\leq 2^{-n}S(n,s) \tag{12}$$

but combining this with the inequality above clearly contradicts our assumption that $\mathbf{P}_d(X_V\in\mathbb{B})\geq 2^{-n}S(n,s)$.

2. The case in which $|\varphi(\mathbf{v}_i)|\leq\Delta/s$ is a little more involved. We look at the smaller multiset of vectors $V_{n-k}=\{\mathbf{v}_1,\mathbf{v}_2,\dots\mathbf{v}_{n-k}\}$, for which we consider $\mathbf{P}_d(X_{V_{n-k}}+\xi_n\mathbf{v}_n\in\mathbb{B}_0)$ for an arbitrary unit ball $\mathbb{B}_0$. Since for this set of linear combinations we can freely condition the remnant i.i.d Bernoulli variables ξ_j where $(n-k+1)\leq j\leq(n-1)$, we must have the existence of a ball $\mathbb{B}_0$

for which the small ball probability is greater for $\mathbf{P}_d(X_{V_{n-k}}+\xi_n)$ with $\mathbb{B}_0$ than that for the original X_V with $\mathbb{B}$. That is,

$$\mathbf{P}_d(X_{V_{n-k}}+\xi_n\mathbf{v}_n\in\mathbb{B}_0)\geq\mathbf{P}_d(X_V\in\mathbb{B}) \tag{13}$$

Now let $\mathbf{c}(\mathbb{B}_0)$ be the center of the ball $\mathbb{B}_0$. Then if we were to have $X_{V_{n-k}}+\xi_n\mathbf{v}_n\in\mathbb{B}_0$, for any ξ_n, or equivalently $X_{V_{n-k}}\in\mathbb{B}_0-\xi_n\mathbf{v}_n$, it follows that after the orthogonal projection to the line,

$$|X_{\varphi(V_{n-k})}-\pi(\mathbf{c}(\mathbb{B}_0))|\leq\text{ radius }\mathbb{B}+\|\xi_n\mathbf{v}_n\|\leq\Delta+\frac{\Delta}{s} \tag{14}$$

Now suppose we also have $|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|>\sqrt{\Delta^2-1}$, then in this case only one of $X_{(V_{n-k})}+\mathbf{v}_n$ and $X_{(V_{n-k})}-\mathbf{v}_n$ can be in $\mathbb{B}_0$. For a more detailed explanation of this using the parallelogram law, see [TV12]. So then under the condition that $|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|>\sqrt{\Delta^2-1}$, we see that the probability is reduced by a factor of 1/2. So then we get, the total probability as

$$\mathbf{P}_d(X_{V_{n-k}}+\xi_n\mathbf{v}_n\in\mathbb{B}_0)$$

$$\leq\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq\sqrt{\Delta^2-1}\Big)+\frac{1}{2}\mathbf{P}_d\Big(\sqrt{\Delta^2-1}<|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq\Delta+\frac{\Delta}{s}\Big) \tag{15}$$

$$\leq\frac{1}{2}\Big[\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq\sqrt{\Delta^2-1}\Big)+\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq\Delta+\frac{\Delta}{s}\Big)\Big] \tag{16}$$

Now suppose Δ in fact satisfies the required chain of inequalities

$$\sqrt{\Delta^2-1}<s-1\leq\Delta<\Delta+\frac{\Delta}{s}<s \tag{17}$$

Then by scaling Theorem 2 by a factor of $|\varphi(v_i)|$, we can conclude that

$$\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq\sqrt{\Delta^2-1}\Big)\leq\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq s-1\Big)\leq 2^{-(n-k)}S(n-k,s-1) \tag{18}$$

and for the other half that

$$\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq\Delta+\frac{\Delta}{s}\Big)\leq\mathbf{P}_d\Big(|X_{\varphi(V_{n-k})}-\varphi(\mathbf{c}(\mathbb{B}_0))|\leq s\Big)\leq 2^{-(n-k)}S(n-k,s) \tag{19}$$

Substituting (18) and (19) into (16),

$$\mathbf{P}_d(X_{V_{n-k}}+\xi_n\mathbf{v}_n\in\mathbb{B}_0)\leq\frac{1}{2}\Big[2^{-(n-k)}S(n-k,s-1)+2^{-(n-k)}S(n-k,s)\Big] \tag{20}$$

Now we consider the asymptotic variant of $2^{-m}S(m,s)$ using Stirling's Formula for large values of m and use it in place of (20). Assuming without loss of any generality that m and s are even, we see that

$$2^{-m}S(m,s)=2^{-m}\sum_{j=-s/2}^{s/2}\binom{m}{m/2+j}=2^{-m}\sum_{j=-s/2}^{s/2}\frac{m!}{(m/2+j)!(m/2-j)!}$$

$$=2^{-m}\sum_{j=-s/2}^{s/2}(1+o(1))\frac{\sqrt{2\pi m}\left(\frac{m}{e}\right)^m}{2\pi\sqrt{\frac{m^2}{4}-j^2}\cdot(m/2+j)^{m/2+j}\cdot(m/2-j)^{m/2-j}\cdot e^{-m}}$$

$$=2^{-m}\sum_{j=-s/2}^{s/2}(1+o(1))\frac{\sqrt{2\pi m}\cdot m^m}{\pi\sqrt{m^2-4j^2}\cdot(m+2j)^{m/2+j}\cdot(m-2j)^{m/2-j}}\cdot 2^m$$

$$=\sum_{j=-s/2}^{s/2}(1+o(1))\sqrt{\frac{2}{\pi}}\cdot\left(\frac{m^2}{m^2-4j^2}\right)^{m/2}\cdot\sqrt{\frac{m}{m^2-4j^2}}\cdot\left(\frac{m-2j}{m+2j}\right)^j=\sum_{j=-s/2}^{s/2}(1+o(1))\sqrt{\frac{2}{\pi m}}$$

$$= (s + o(1))\sqrt{\frac{2}{\pi m}}$$

And so for the probability, we get

$$\frac{1}{2}\Big[2^{-(n-k)}S(n-k, s-1) + 2^{-(n-k)}S(n-k, s)\Big] = \frac{1}{2}\Big([(s-1)+o(1)]\sqrt{\frac{2}{\pi n}} + (s+o(1))\sqrt{\frac{2}{\pi m}}\Big) \tag{21}$$

$$= [s - 1/2 + o(1)] \cdot \sqrt{\frac{2}{\pi m}}$$

But because of this notice that $\mathbf{P}_d(X_{V_{n-k}} + \xi_n \mathbf{v}_n \in \mathbb{B}_0) \leq 2^{-m}S(n, s)$, from which it follows that $\mathbf{P}_d(X_V \in \mathbb{B}) \leq 2^{-n}S(n, s)$, which clearly raises a contradiction.

Our calculation of the probabilities is based on the premise that (17) holds. On treating the first inequality, we see that $\Delta < \sqrt{(s-1)^2 + 1}$. The second inequality enforces the restriction that $s - 1 \leq \Delta$. Now note that the fourth inequality is actually unnecessary since the constraint imposed by it is already covered by the first two: squaring both sides, we see that $\Delta^2(1+1/s)^2 \leq s^2$, simplifying further and substituting $s - 1 < \Delta$,

$$(s-1)^2 \cdot \Big(\frac{s^2 + 2s + 1}{s^2}\Big) \leq s^2, \text{ or } (s-1)^2(s+1)^2 \leq s^4$$

which holds true for $s > 2$. This gives us the restrictions from the conjecture, and also finishes our proof of the result from the conjecture.

□

4. Fourier Analytic Techniques

An essential generalization of the Littlewood-Offord problem that we have saved for this last section is about some form of variation in the constraints imposed upon $\mathbf{v}_i$. While the combinatorial method, as we have seen, provides sharp and exact estimates and bounds for the probability, it is often difficult to generalize to be made applicable to other cases.

However, a Fourier-analytic approach suggested by Halasz in [Hal77] provided a new viewpoint that gave bounds even though were not as sharp as due to the combinatorial strategies, could be generalized across varying structures of $\mathbf{v}_i$. In many situations, we obtain sharper bounds by imposing restrictions:

Example 4.1. Consider the closely related probabilistic notion $\sup_{x\in\mathbb{R}} \mathbf{P}(X_V = x)$, that is the least upper bound on the probability that the linear combination X_V is equal to a certain real x. Then if V is a multiset in $\mathbb{R}$ such that all v_i are distinct, then

$$\sup_{x\in\mathbb{R}} \mathbf{P}(X_V = x) = O(n^{-3/2})$$

In fact, we can further show by Halasz methods a chain of bounds by increasing the restrictions

$$\begin{array}{ccc} v_i \text{ distinct} & \Longrightarrow & \sup_{x\in\mathbb{R}^d} \mathbf{P}(X_V = x) = O(n^{-3/2}) \\ & & \downarrow \\ v_i + v_j \text{ distinct} & \Longrightarrow & \sup_{x\in\mathbb{R}^d} \mathbf{P}(X_V = x) = O(n^{-5/2}) \\ & & \downarrow \\ v_i + v_j + v_k \text{ distinct} & \Longrightarrow & \sup_{x\in\mathbb{R}^d} \mathbf{P}(X_V = x) = O(n^{-7/2}) \end{array}$$

Using the same Fourier anlaytic techniques, a broad generalization is also possible:

Theorem 9. *For a set of vectors V, if $|s_m|$ is the number of m-sums that are equal, then*

$$\sup_{x\in\mathbb{R}} \mathbf{P}(X_V = x) = O(n^{-2m-\frac{1}{2}}|s_m|)$$

Theorem 9 reflects an important relationship between the structure of V and the the probability itself: if more chains of m-sums are equal to one another, then we can get a larger bound on $\sup \mathbf{P}(X_V = x)$, which does make some intuitive sense. A proof of this theorem and that of other theorems like it can be found in [TV06] and [Hal77].

On the broader scale, Halasz result deeper insight into the structures of $\mathbf{v}_i$ that allowed certain types of small-ball probabilities to exist paving the way for the formulation of the Inverse-Littlewood Offord theory. As for the Fourier analytic techniques, our focus in this paper will primarily be on developing the following important concentration inequality along with the requisite background in Fourier analysis.

Theorem 10. *There exists a constant C such that for any unit ball $\mathbb{B} \in \mathbb{R}^d$ and linear combination of vectors X,*

$$\mathbf{P}_d(X \in \mathbb{B}) \leq C \int_{\|t\|\leq 1} |\phi_X(t)|\, dt = C \int_{\|t\|\leq 1} \big|\mathbb{E}[\exp(i\langle t, X\rangle)]\big|\, dt,$$

where $\phi_X(t)$ is the characteristic function of the probability distribution of X and $\|t\|$ denotes the standard Euclidean norm.

Definition 4.1. The **characteristic function** $\phi_X(t)$ of a random variable X with cumulative distribution function $F_X(x)$ and density function $f_X(x)$ is defined by the following integral:

$$\phi_X(t) = \mathbb{E}[\exp(itX)] = \int_{-\infty}^{\infty} e^{itx} dF_X(x) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx \tag{22}$$

The characteristic function can be interpreted as the Fourier transform of the probability density function $f_X(x)$. Conventionally, the Fourier transform is implemented as $\hat{f}(t) = \int_{-\infty}^{\infty} f(x)e^{-2\pi itx}dx$, but our definition of the characteristic function is just a rescaling of the aforementioned definition and works better in probabilistic scenarios.

The inequality in Theorem 10 converts the problem of finding the small ball probability into the calculation of an integral, which in some cases makes the problem easier. Another big advantage is that the introduction of ideas from probability and Fourier analysis allows us to use powerful tools from these fields, which once again will prove instrumental in solving for small ball probabilities.

Notice that X is actually forms a multivariate random probability distribution as it is a linear combination of vectors with coefficients as i.i.d Bernoulli random variables, and so we will need a higher dimensional analog to (22).

Definition 4.2. The characteristic function $\phi_X(t)$ of a multivariate random variable X is defined for a probability density function $f_X(x)$ as

$$\phi_X(t) = \mathbb{E}[\exp(i\langle t, X\rangle)] = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} f_X(x)dx \tag{23}$$

It is also important to note that given a characteristic function, we can recover the probability density function by an inversion formula

Theorem 11 (Inversion theorem)**.** *For a multivariate random variable X, if $\phi_X(t)$ is an integrable characteristic function for the probability density function $f_X(t)$, then*

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(t) e^{-i\langle t,x\rangle} dt \tag{24}$$

We do not provide a proof as it is somewhat engaged in probability theory. However, with this theorem, we can move on to proving the concentration inequality.

Proof of Concentration inequality. To prove Theorem 10, we first observe that our small ball probability $\mathbf{P}_d(X \in \mathbb{B})$ can be expressed in terms of its density function $f_X(x)$ as follows:

$$\mathbf{P}_d(X \in \mathbb{B}) = \int_{\|x - C_0\| \leq 1} f_X(x)dx \tag{25}$$

where C_0 is the center of $\mathbb{B}$. Our approach will be to define a new function $k(t)$ and its Fourier transform $K(x)$, which we can embed with the necessary properties to bound the RHS of (25). Since $\phi_X(t) = \mathbb{E}[\exp(i\langle t, X\rangle)]$ is also the Fourier transform of $f_X(x)$, we can use Plancherel's theorem stated below to relate the two functions.

Theorem 12 (Plancherel's Theorem)**.** *For two integrable functions $f(t)$ and $g(t)$ with sufficient additional properties, let $F(x)$ and $G(x)$ be their Fourier transforms respectively. Then*

$$\int_{\mathbb{R}^d} f(t)\overline{g(t)}dt = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} F(x)\overline{G(x)}dx,$$

where $\overline{\cdot}$ represents the complex conjugate.

Theorem 12 is often referred to as Parseval's identity or equality, and in many cases is stated without the $1/(2\pi)^d$ term; the presence or absence of this term depends on the way we have defined our fourier transform, and if we adhere to the conventional definition $\hat{f}(\omega) = \int_{\mathbb{R}} f(t) \cdot \exp(-2\pi i \omega t)dt$, the $1/(2\pi)^d$ factor goes away. See Theorem 25.9 of [How16] for more details and a proof.

Based on the definition of $K(x)$, we have

$$K(x) = \int_{\mathbb{R}^d} k(t) \cdot e^{i\langle x,t\rangle}dt \tag{26}$$

It is important to note that while $\phi_X(t)$ is a Fourier transform of $f_X(x)$, it represents a different change of variables, i.e from x to t, and as a result Theorem 13 cannot yet be applied. To state the relationship $\phi_X(t) = \int_{\mathbb{R}^d} \exp(i\langle t, X\rangle) f_X(x)dx$ in a similar format, we use the inversion theorem to get

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_X(t) e^{-i\langle t,x\rangle}dt \tag{27}$$

Now, if we take the complex conjugate on both sides,

$$\overline{f_X(x)} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \overline{\phi_X(t) e^{-i\langle t,x\rangle}}dt = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \overline{\phi_X(t)} e^{i\langle t,x\rangle}dt = \int_{\mathbb{R}^d} \overline{\phi'_X(t)} e^{i\langle t,x\rangle}dt \tag{28}$$

and so $\overline{f_X(x)}$ is the Fourier transform of $\overline{\phi'_X(t)} := \frac{1}{(2\pi)^d}\overline{\phi_X(t)}$. Now we can use Plancherel's Theorem to relate the two functions as follows:

$$\int_{\mathbb{R}^d} k(t) \overline{\overline{\phi'_X(t)}}\, dt = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} K(x) \overline{\overline{f_X(x)}}\, dx,$$

or equivalently,

$$\int_{\mathbb{R}^d} k(t) \frac{\phi_X(t)}{(2\pi)^d} dt = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} K(x) f_X(x)dx$$

And so we finally get

$$\int_{\mathbb{R}^d} K(x) f_X(x)dx = \int_{\mathbb{R}^d} k(t)\phi_X(t)dt \tag{29}$$

Now we define $k(t)$ as

$$k(t) = \begin{cases} |k(t)| \leq c_1 & \text{for } \|t\| \leq 1 \\ k(t) = 0 & \text{for } \|t\| \geq 1 \end{cases},$$

so that in combination with (29), we get

$$\int_{\mathbb{R}^d} K(x) f_X(x) dx = \int_{\mathbb{R}^d} k(t)\phi_X(t) dt \leq c_1 \int_{\|t\| \leq 1} \phi_X(t) dt \leq C \int_{\|t\| \leq 1} |\phi_X(t)| dt \tag{30}$$

Our next step will be to bound the small ball probability by defining $K(x)$ appropriately. We define $K(x)$ as

$$K(x) = \begin{cases} K(x) \geq 1 & \text{for } \|x\| \leq c_2, \text{ where } c_2 \text{ is a constant} \\ K(x) \geq 0 & \text{for } \|x\| \geq c_2 \end{cases}$$

From this and (30), the following chain of inequalities arise:

$$\int_{\|x\| \leq c_2} f_X(x) dx \leq \int_{\mathbb{R}^d} K(x) f_X(x) dx \leq C \int_{\mathbb{R}^d} |\phi_X(t)| dt \tag{31}$$

As a final step, notice that we can translate $K(x)$ by a factor of C_0 by multiplying $k(t)$ by the phase $\exp(i \langle C_0, x \rangle)$. So the following translated version of (31) also holds true:

$$\int_{\|x - C_0\| \leq c_2} f_X(x) dx \leq C \int_{\mathbb{R}^d} |\phi_X(t)| dt \tag{32}$$

Now for an appropriate radius c_2, the definition of the small ball probability in (25) combined with (32) finishes the proof

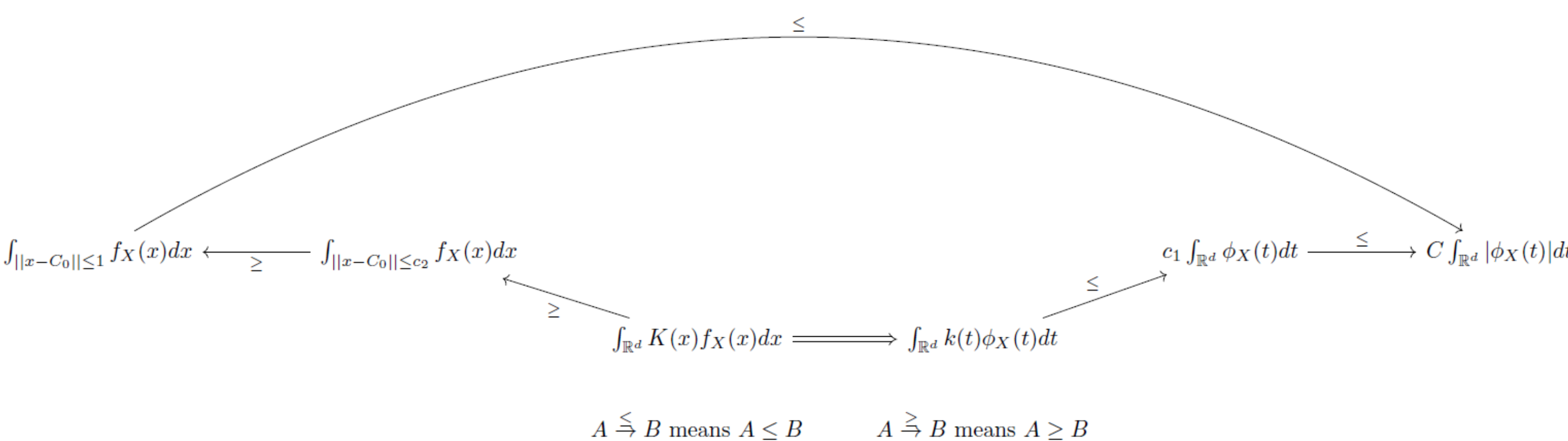
$$\mathbf{P}_d(X \in \mathbb{B}_0) = \int_{\|x - C_0\| \leq 1} f_X(x) dx \leq C \int_{\mathbb{R}^d} |\phi_X(t)| dt \tag{33}$$

Note that even if $c_2 < 1$, we can add the density function of a series of balls that can cover $\mathbb{B}$, still giving a finite constant C for which the above relation holds. □

Lastly, all we need to verify is the existence of such a function $k(t)$. Notice that if we take

$$k(y) := \int_{y \in \mathbb{R}^d} \kappa_0(y) \kappa_0(t - y) dy \tag{34}$$

where $\kappa_0(x) = 1$ for $\|x\| \leq 1/2$ and is zero everywhere, $k(t)$ suits our definition. As a reference, the following illustrates our approach and the various inequalities considered:



Example 4.2. To demonstrate the prowess of this concentration inequality, and Halasz Fourier analytic method in general, we will give a proof of Erdos Result in Theorem 2, which states that

$$\mathbf{P}_1(X \in \mathbb{B}) = \mathbf{P}_1(n, 1) = O(1/\sqrt{n})$$

Proof of Theorem 2. Let $V = \{v_1, v_2, v_3, \dots v_n\}$ be our set of real numbers with $|v_i| \geq 1$. Then equipped with Theorem 10, all we need to show is that

$$\int_{|t|\leq 1} \Big|\mathbb{E}\Big[\exp(it\sum_{j=1}^{n}\xi_j v_j)\Big]\Big| dt = O(1/\sqrt{n}) \tag{35}$$

Now since ξ_j are i.i.d Bernoulli random variables,

$$\Big|\mathbb{E}\Big[\exp(it\sum_{j=1}^{n}\xi_j v_j)\Big]\Big| = \Big|\mathbb{E}\Big[\prod_{j=1}^{n}\exp(it\xi_j v_j)\Big]\Big| = \prod_{j=1}^{n}\big|\mathbb{E}[\exp(it\xi_j v_j)]\big| = \prod_{j=1}^{n}|\cos(tv_j)|$$

On applying Holder's inequality to the integral,

$$\int_{|t|\leq 1} \Big|\mathbb{E}\Big[\exp(it\sum_{j=1}^{n}\xi_i v_i)\Big]\Big| dt = \int_{|t|\leq 1}\prod_{j=1}^{n}|\cos(tv_j)|\, dt \leq \prod_{j=1}^{n}\Big(\int_{|t|\leq 1}|\cos(tv_j)|^n\, dt\Big)^{1/n} \tag{36}$$

And as a result, it is sufficient to show the bound stated in the following lemma

Lemma 3. *For a constant c, sufficiently large n, and* $|v_j| \geq 1$,

$$\mathcal{I} = \int_{-1}^{1}|\cos(tv_j)|^n\, dt \leq \frac{c}{\sqrt{n}}$$

Proof of Lemma 3. First, without loss of generality note that we can assume $v_j > 1$ since for negative v_j we can just use the identity $\cos(-x) = \cos(x)$. Additionally, because of the same property of $\cos(x)$, we also have that $\int_{-1}^{1}|\cos(tv_j)|^n\, dt = 2\int_0^1|\cos(tv_j)|^n\, dt = 2\mathcal{J}$. Our proof of Lemma 3 will rely on the inequality $|\cos(x)| \leq e^{-x^2/2}$ for $|x| \leq \pi/2$, which we can prove by looking at the Maclaurin expansions of the two functions:

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots, \text{ and so } \cos(x) \leq 1 - \frac{x^2}{2} + \frac{x^4}{24}$$

$$e^{-x^2/2} = 1 - \frac{x^2}{2} + \frac{x^4}{8} - \frac{x^6}{48} + \dots, \text{ and so } e^{-x^2/2} \leq 1 - \frac{x^2}{2} + \frac{x^4}{8} - \frac{x^6}{48}$$

Consequently, we now need to show that

$$\frac{x^4}{8} - \frac{x^6}{48} \geq \frac{x^4}{24}, \text{ or equivalently } \frac{x^4}{12} \geq \frac{x^6}{48}$$

which holds as long as $x^4(2+x)(2-x) \leq 0$, and so also over the interval $-\pi/2 \leq x \leq \pi/2$. We can now rescale our inequality to fit our integrand and get

$$|\cos(v_j t)|^n \leq e^{-nv_j^2t^2/2} \text{ for } |t| \leq \frac{\pi}{2v_j} \tag{37}$$

This last inequality provides insight into our general approach towards obtaining the mentioned bound. Since the bound holds over one period of the $f(t) = |\cos(v_j t)|^n$, we can bound each the function over each period $\big[\frac{(2k-1)\pi}{2v_j}, \frac{(2k+1)\pi}{2v_j}\big]$ by an appropriately translated gaussian function. Note that since the function peaks at $k\pi/v_j$ over the aforementioned interval, the translated gaussian must also have its maximum value translated $k\pi/v_j$ units, and so we get the gaussian associated with the interval $\big[\frac{(2k-1)\pi}{2v_j}, \frac{(2k+1)\pi}{2v_j}\big]$ as $g_k = \exp(-n(k\pi - v_j t)^2/2)$. Over $[0, 1]$, we will have to account for a total of $(m+1)$ peaks, where $m = \lceil \frac{v_j}{\pi} \rceil$. and so a corresponding $(m+1)$ gaussians. Since some peaks may lie outside $[0,1]$, we extend the interval to $\left[-\frac{\pi}{2v_j}, (m+1/2)\frac{\pi}{v_j}\right]$. Now since

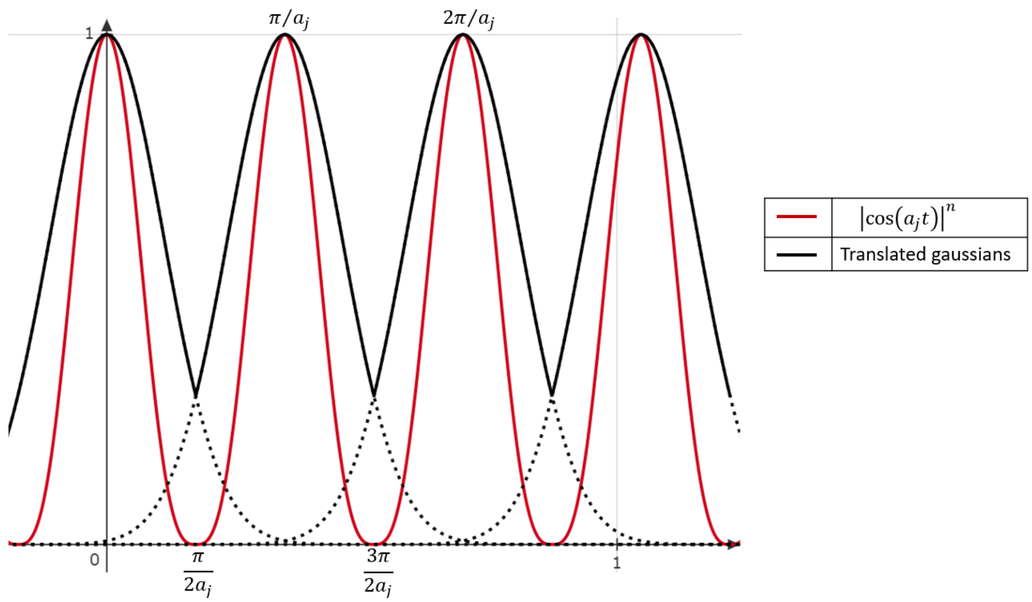
$$e^{-n(k\pi - v_j t)^2/2} = e^{-n((k+1)\pi - v_j t)^2/2} \text{ when } t = \frac{(2k-1)\pi}{2v_j}, \frac{(2k+1)\pi}{2v_j},$$

f is completely bounded over the interval since the gaussians fill the interval completely. So then we can replace the integral $\mathcal{J}$ by a finite sum of a chain of $(m+1)$ gaussians g_k integrated over the bound $\left[\frac{(2k-1)\pi}{2v_j}, \frac{(2k+1)\pi}{2v_j}\right]$ respectively.

$$\mathcal{J} < \int_{-\pi/2v_j}^{\pi/2v_j} e^{-nv_j^2 t^2/2} dt + \cdots + \int_{\frac{(2k-1)\pi}{2v_j}}^{\frac{(2k+1)\pi}{2v_j}} e^{-n(k\pi - v_j t)^2/2} dt + \cdots + \ell e^{-n(m\pi - v_j t)^2/2} dt$$

$$\text{So, } \mathcal{J} < \sum_{\ell=0}^{m} \int_{\frac{(2\ell-1)\pi}{2v_j}}^{\frac{(2\ell+1)\pi}{2v_j}} g_\ell(t)\, dt \tag{38}$$

The figure below illustrates this process:



Now to evaluate the integral of the translated gaussian $g_\ell(t) = \exp(-n(\ell\pi - v_j t)^2/2)$, we first make the substitution $u = \ell\pi - v_j t$, for which $dt = \frac{-dt}{v_j}$.

$$\int_{\frac{(2\ell-1)\pi}{2v_j}}^{\frac{(2\ell+1)\pi}{2v_j}} g_\ell(t)\, dt = \frac{1}{v_j} \int_{-\ell\pi/2}^{\ell\pi/2} e^{-nu^2} du \tag{39}$$

Furthermore, since $e^{-nu^2/2} > 0$ for all $\mathbb{R}$, we can replace the limits of integration to span over $\mathbb{R}$.

$$\frac{1}{v_j} \int_{-\ell\pi/2}^{\ell\pi/2} e^{-nu^2} du < \frac{1}{v_j} \int_{-\infty}^{\infty} e^{-nu^2} du = \frac{1}{v_j}\sqrt{\frac{2\pi}{n}} \tag{40}$$

where the last equality follows from our evaluation of RHS of (4). From this and (38), we get the desired

$$\mathcal{I} < \frac{2m}{v_j}\sqrt{\frac{2\pi}{n}}, \text{ and so } \int_{-1}^{1} |\cos(tv_j)|^n\, dt = O(1/\sqrt{n})$$

□

5. Further Research

The Fourier analytic techniques mentioned in the last section have had a profound impact on the development of the Littlewood-Offord Problem. Theorem 10 shows that decreasing the additive structure of V, we could decrease the concentration probability of X_V. Using similar methods, many such results can be shown that exhibit a clear relationship between the structure of V and the small ball probability. So in [TV09], the authors asked and partly answered the following question that became the premise of the Inverse Littlewood-Offord Theory:

Question 5.1. Assume that for some constant c,

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}^d} \mathbf{P}(X_V = x) \geq n^{-c}$$

what can we say about the structure of V and about the vectors $\mathbf{v}_i$?

In the same paper, Tao and Vu proved a series of **inverse** theorems that proposed the following possible generalization:

Theorem 13. *If* $\sup \mathbf{P}(X_V = x)$ *is large, then* V *must have a strong additive structure.*

We refer the reader to [TV09] for more on inverse Littlewood-Offord theory.
Another interesting possible generalization is to analyze the probability when ξ_i have an arbitrary distribution, which has recently been explored in [JK20].

References

[Con16] Keith Conrad. Stirling's formula. *Available in http://www. math. uconn. edu/˜ kconrad/blurbs/analysis/stirling. pdf*, 2016.

[Erd45] Paul Erdös. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

[FF88] Péter Frankl and Z Furedi. Solution of the littlewood-offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.

[Gri80] Jerrold R Griggs. The littlewood-offord problem: tightest packing and an m-part sperner theorem. *European Journal of Combinatorics*, 1(3):225–234, 1980.

[Hal77] Gábor Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8(3-4):197–211, 1977.

[How16] Kenneth B Howell. *Principles of Fourier analysis*. CRC Press, 2016.

[JK20] T Juškevičius and V Kurauskas. On littlewood–offord problem for arbitrary distributions. 2020.

[Kat66] Gy Katona. On a conjecture of erdős and a stronger form of sperner's theorem. *Studia Sci. Math. Hungar*, 1:59–63, 1966.

[Kha74] Rasul A Khan. A probabilistic proof of stirling's formula. *The American Mathematical Monthly*, 81(4):366–369, 1974.

[Kle65] Daniel J Kleitman. On a lemma of littlewood and offord on the distribution of certain sums. *Mathematische Zeitschrift*, 90(4):251–259, 1965.

[Kle70] Daniel J Kleitman. On a lemma of littlewood and offord on the distributions of linear combinations of vectors. *Advances in Mathematics*, 5(1):155–157, 1970.

[LO39] John Edensor Littlewood and Albert C Offord. On the number of real roots of a random algebraic equation. ii. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 35, pages 133–148. Cambridge University Press, 1939.

[NV13] Hoi H Nguyen and Van H Vu. Small ball probability, inverse theorems, and applications. In *Erdős centennial*, pages 409–463. Springer, 2013.

[RF10] HL Royden and Patrick Fitzpatrick. Real analysis (4th edtion). *New Jersey: Printice-Hall Inc*, 2010.

[Sal83] Attila Sali. Stronger form of an m-part sperner theorem. *European Journal of Combinatorics*, 4(2):179–183, 1983.

[TV06] Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[TV09] Terence Tao and Van H Vu. Inverse littlewood-offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, pages 595–632, 2009.

[TV12] Terence Tao and Van Vu. The littlewood-offord problem in high dimensions and a conjecture of frankl and füredi. *Combinatorica*, 32(3):363–372, 2012.

[Won77] Chi Song Wong. A note on the central limit theorem. *The American Mathematical Monthly*, 84(6):472–472, 1977.