Irregularities of Distribution, Derandomization, and Complexity Theory*

Bernard Chazelle†
chazelle@cs.princeton.edu

Abstract

In 1935, van der Corput asked the following question: Given an infinite sequence of reals in $[0,1]$, define

$$D(n) = \sup_{0 \le x \le 1} \Big| \, |S_n \cap [0,x]\,| - nx \Big|,$$

where S_n consists of the first n elements in the sequence. Is it possible for $D(n)$ to stay in $O(1)$? Many years later, Schmidt proved that $D(n)$ can never be in $o(\log n)$. In other words, there are limitations on how well the discrete distribution, $x \mapsto |S_n \cap [0,x]\,|$, can simulate the continuous one, $x \mapsto nx$. The study of this intriguing phenomenon and its numerous variants related to the irregularities of distributions has given rise to *discrepancy theory.* The relevance of the subject to complexity theory is most evident in the study of probabilistic algorithms. Suppose that we feed a probabilistic algorithm not with a perfectly random sequence of bits (as is usually required) but one that is only pseudorandom or even deterministic. Should performance necessarily suffer? In particular, suppose that one could trade an exponential-size probability space for one of polynomial size without letting the algorithm realize the change. This form of derandomization can be expressed by saying that a very large distribution can be simulated by a small one for the purpose of the algorithm. Put differently, there exists a measure with respect to which the two distributions have low discrepancy. The study of discrepancy theory predates complexity theory and a wealth of mathematical techniques can be brought to bear to prove nontrivial derandomization results. The pipeline of ideas that flows from discrepancy theory to complexity theory constitutes the *discrepancy method.* We give a few examples in this survey. A more thorough treatment is given in our book [15]. We also briefly discuss the relevance of the discrepancy method to complexity lower bounds.

1 Facts from Discrepancy Theory

Let $(V, \mathcal{S})$ be a set system, where $V = \{v_1, \ldots, v_n\}$ is the ground set and $\mathcal{S} = \{S_1, \ldots, S_m\}$, with $S_i \subseteq V$. We wish to color the elements of V red and blue so that, within each S_i, no color greatly outnumbers the other one. To do that, we choose a function χ that maps each $v_j \in V$ to an element in $\{-1, 1\}$, and we define the *discrepancy* of the set S_i to be

$$\chi(S_i) = \sum_{v_j \in S_i} \chi(v_j).$$

*Proceedings of FSTTCS-2000. This work was supported in part by NSF Grant CCR-96-23768, ARO Grant DAAH04-96-1-0181, and NEC Research Institute.

†Department of Computer Science, Princeton University, and NEC Research Institute

The maximum value of $|\chi(S_i)|$, over all $S_i \in \mathcal{S}$, is the discrepancy of the set system under the given coloring. The *discrepancy* of the set system itself, denoted by $D_\infty(\mathcal{S})$, refers to its minimum discrepancy under all possible colorings. The L^2 norm creates an easier environment to work with, and so we also define

$$D_2(\mathcal{S}) \stackrel{\text{def}}{=} \min_\chi \sqrt{\chi(S_1)^2 + \cdots + \chi(S_m)^2}\,,$$

where the minimum is taken over all colorings $\chi : V \mapsto \{-1, 1\}$. The discrepancy can be characterized by using matrices, which is sometimes more convenient. Let A be the *incidence matrix* of the set system $(V, \mathcal{S})$: the n columns are indexed by the elements of V and the m rows are the characteristic vectors of the sets S_i, so that A_{ij} is 1 if $v_j \in S_i$ and 0 otherwise. The discrepancy of the set system, also denoted by $D_\infty(A)$, can be expressed as the L^∞ norm of a column vector. Generally, for any $p \in \{1, 2, \ldots, \infty\}$, we have $D_p(A) = \min_{x \in \{-1,1\}^n} \|Ax\|_p$. The following result of Spencer [44] is tight.

Theorem 1.1 *Any set system $(V, \mathcal{S})$ such that $|V| = |\mathcal{S}| = n$ has $O(\sqrt{n}\,)$ discrepancy.*

For general set systems with m sets, the bound becomes $O(\sqrt{n \ln(2m/n)}\,)$. A simple, elegant result concerns the case of small-degree set systems. The degree refers to the maximum number of sets containing a given element. The classical Beck-Fiala theorem [7] states that:

Theorem 1.2 *The discrepancy of a set system of degree at most t is less than $2t$.*

Techniques for proving lower bounds often involve spectral arguments and, in particular, harmonic analysis. The latter comes from the fact that set systems are often defined by using a convolution operator, which the Fourier transform diagonalizes. Bounding the eigenvalues gives us a handle on the L^2-norm discrepancy. Perhaps the simplest result obtained in this manner is Roth's $\frac{1}{4}$-Theorem [40].

Theorem 1.3 *Any two-coloring of the integers $\{1, \ldots, n\}$ contains an arithmetic progression whose discrepancy is $\Omega(n^{1/4})$.*

There exists a wealth of techniques and results for geometric set systems. In such cases, it is useful to define the notion of volume discrepancy. Consider the problem of placing a set P of n points in the unit cube $[0,1]^d$ to minimize the discrepancy with respect to axis-parallel boxes. The (volume) discrepancy of a box $B = \prod_{k=1}^d [p_k, q_k)$ is defined as

$$D(B) \stackrel{\text{def}}{=} n \cdot \text{vol}\,(B) - |P \cap B|.$$

Theorem 1.4 *There is a set of n points in $[0,1]^d$ such that the volume discrepancy of any box in $[0,1]^d$ is $O(\log n)^{d-1}$ in absolute value.*

Here is a construction in two dimensions [46, 47]. Given a nonnegative integer m, let $\sum_{i \geq 0} b_1(i)\, 2^i$ be its binary decomposition, and let

$$x_1(m) = \sum_{i \geq 0} \frac{b_1(i)}{2^{i+1}} \in [0, 1).$$

The numbers $x_1(m)$, for $0 \le m < n$, form the classical *van der Corput* sequence. We can use it to define the *bit-reversal* point set:

$$\left\{ (x_1(m), m/n) \,\Big|\, 0 \le m < n \right\}.$$

This easily generalizes to d dimensions. Choose $d-1$ relatively prime numbers: $2 = p_1, p_2, \ldots, p_{d-1}$. The integer m has a unique decomposition in base p_k, $m = \sum_{i \ge 0} b_k(i) p_k^i$, so we can define

$$x_k(m) = \sum_{i \ge 0} \frac{b_k(i)}{p_k^{i+1}}.$$

The point set

$$P = \left\{ \left(x_1(m), \ldots, x_{d-1}(m), \frac{m}{n} \right) \,:\, 0 \le m < n \right\}$$

is called *Halton-Hammersley* [25] and satisfies Theorem 1.4.

What about the L^2 norm? Let P be a set of n points in the unit square. Given a box B_q of the form $[0, q_1) \times [0, q_2)$, where $q = (q_1, q_2)$, the discrepancy of B_q is

$$D(B_q) = n \cdot \text{area}\,(B_q) - |P \cap B_q|.$$

We define the L^2-norm discrepancy of P as

$$D_2(P) \stackrel{\text{def}}{=} \sqrt{\int_{[0,1]^2} D(B_q)^2 \, dq}.$$

The following result is by Davenport [22].

Theorem 1.5 *It is possible to find a set P of n points in $[0,1]^2$ such that $D_2(P) = O(\sqrt{\log n}\,)$.*

We forsake the Halton-Hammersley construction and, instead, turn to a construction based on irrational lattices. Take the set of $n = 2k - 1$ points of the form

$$\left(\{ j\varphi \}, \frac{|j|}{n} \right),$$

for all j ($|j| < k$), where $\{x\} \stackrel{\text{def}}{=} x \pmod 1$ is the fractional part of x and $\varphi = \frac{1}{2}(\sqrt{5} + 1)$ is the golden ratio. The only property we use about the golden ratio is the size of the partial quotients of its continued fraction expansion, so many other choices exist for Theorem 1.5.

We generalize the discrepancy to $\mathbf{R}^d$ in the obvious manner. Given a point $q = (q_1, \ldots, q_d)$ in the unit cube $[0,1]^d$, let B_q denote the box $[0, q_1) \times \cdots \times [0, q_d)$. Fix a set P of n points in $[0,1]^d$, and as usual define the volume discrepancy $D(B_q)$ at a point $q \in [0,1]^d$ as $D(B_q) = n q_1 \cdots q_d - |P \cap B_q|$. We write $D_2(P) = \sqrt{\int_{[0,1]^d} D(B_q)^2 \, dq}$. The following bound is due to Roth [39], and shows the optimality of Theorem 1.5.

Theorem 1.6 *Given a set P of n points in $[0,1]^d$, the mean-square discrepancy for axis-parallel boxes satisfies*

$$D_2(P) > c (\log n)^{(d-1)/2},$$

for some constant $c = c(d) > 0$.

In two dimensions, we have this interesting lower bound by Schmidt [41], which shows a rare divergence between L^2 and L^∞ behaviors.

Theorem 1.7 *Given n points in $[0,1]^2$, there exists a box B such that $|D(B)| = \Omega(\log n)$.*

We now consider rotated boxes. Given a set P of n points in $[0,1]^2$, the discrepancy of a (rotated) box R is defined naturally as $D(R) = n \cdot \mathrm{area}\,(R \cap [0,1]^2) - |P \cap R|$. By rotated box, we mean any rectangle not necessarily parallel to the axes. The following upper bound was established by Beck; see Beck and Chen's book [6].

Theorem 1.8 *It is possible to place n points in the unit square $[0,1]^2$, so that any (rotated) box R satisfies $|D(R)| = O(n^{1/4}\sqrt{\log n}\,)$.*

A quasi-matching lower bound was first proven by Beck [5], using his beautiful *Fourier transform method* (other proof techniques exist).

Theorem 1.9 *Given n points in the unit square $[0,1]^2$, there exists a rotated box R such that $|D(R)| = \Omega(n^{1/4})$.*

The same bound holds for disks as well. The proof, by Montgomery [34, 35], also uses harmonic analysis.

Theorem 1.10 *Given n points in the unit square $[0,1]^2$, there exists a disk K such that $|D(K)| = \Omega(n^{1/4})$.*

2 Sampling

The red-blue discrepancy of a set system tells us how well we can sample its ground set by choosing about half of its elements. What about different sample sizes? For example, given a collection of n points in the plane, is it possible to choose a subset of constant size, such that any disk that encloses at least one percent of the points also includes at least one sample point? Surprisingly, the answer is yes. The surprise is that the sample size can be kept independent of n. The magic lies in the notion of VC dimension.

Let $(V, \mathcal{S})$ be a (finite or infinite) set system. Given $Y \subseteq V$, let $(Y, \mathcal{S}|_Y)$ denote the set system *induced* by Y, ie, $\{\, Y \cap S \mid S \in \mathcal{S}\,\}$. A subset Y of V is said to be *shattered* (by $\mathcal{S}$) if $\mathcal{S}|_Y = 2^Y$, ie, every subset of Y (including the empty set) is of the form $Y \cap S$, for some $S \in \mathcal{S}$. The supremum of all sizes of finite shattered subsets of X is called the *Vapnik-Chervonenkis dimension* (or *VC-dimension* for short) of the set system.

Let $(V, \mathcal{S})$ be a finite set system, where $|V| = n$ and $|\mathcal{S}| = m$. Given any $0 < \varepsilon < 1$, a set $N \subseteq V$ is called an *ε-net for* $(V, \mathcal{S})$ if $N \cap S \neq \emptyset$, for any $S \in \mathcal{S}$ with $|S|/|V| > \varepsilon$. A set $A \subseteq V$ is called an *ε-approximation* for $(V, \mathcal{S})$ if, for any $S \in \mathcal{S}$,

$$\left| \frac{|S|}{|V|} - \frac{|A \cap S|}{|A|} \right| \le \varepsilon.$$

Equivalently, given a random v uniformly distributed in V, for each $S \in \mathcal{S}$,

$$\Big| \mathrm{Prob}[v \in S] - \mathrm{Prob}[\, v \in S \mid v \in A\,] \Big| \le \varepsilon.$$

The following was proven by Chazelle and Matoušek [18], building on the foundational work in [16, 26, 29, 48].

Theorem 2.1 *Let $(V, \mathcal{S})$ be a set system of VC-dimension d. Given any $r \geq 2$, a $(1/r)$-approximation for $(V, \mathcal{S})$ of size $O(dr^2 \log dr)$ can be computed in time $O(d)^{3d}(r^2 \log dr)^d |V|$.*

Theorem 2.2 *Let $(V, \mathcal{S})$ be a set system of VC-dimension d. Given any $r \geq 2$, a $(1/r)$-net for $(V, \mathcal{S})$ of size $O(dr \log dr)$ can be computed in time $O(d)^{3d}(r^2 \log dr)^d |V|$.*

Note that the set systems are usually understood as members of an infinite family; for example the set of all points in $\mathbf{R}^2$ and the set of all disks. The term *range space* is often used in the literature to refer to such a family.

3 Geometric Algorithms

Suppose that we are given a set H of n hyperplanes in $\mathbf{R}^d$. We wish to subdivide $\mathbf{R}^d$ into a small number of simplices, so that none of them is cut by too many hyperplanes. Given a parameter $\varepsilon > 0$, a collection $\mathcal{C}$ of closed full-dimensional simplices is called an *ε-cutting* if: (i) their interiors are pairwise disjoint, and together they cover $\mathbf{R}^d$; and (ii) the interior of any simplex of $\mathcal{C}$ is intersected by at most εn hyperplanes of H.

Cuttings are among the most useful, versatile tools in computational geometry, as they lay the grounds for efficient divide-and-conquer [1, 2, 20, 26, 27]. Using some of the sampling technology for finite VC dimension discussed earlier, Chazelle [11] proved the following:

Theorem 3.1 *Given a collection H of n hyperplanes in $\mathbf{R}^d$, for any $r > 0$, there exists a $(1/r)$-cutting for H of optimal size $O(r^d)$. A full description of the cutting, including the list of hyperplanes intersecting the interior of each simplex, can be found deterministically in $O(nr^{d-1})$ time.*

Here are some direct applications of cuttings: Point location is understood here as the problem of preprocessing an arrangement of n hyperplanes in $\mathbf{R}^d$ so that, given a query point, the face of the arrangement that contains the point can be found quickly. Simplex range searching is the problem of preprocessing n points in $\mathbf{R}^d$ so that given a query simplex the points inside it can be counted quickly.

Theorem 3.2 *Point location among n hyperplanes in $\mathbf{R}^d$ can be done in $O(\log n)$ query time, using $O(n^d)$ preprocessing.*

Theorem 3.3 *To decide whether n points and n lines in the plane are free of any incidence can be done in $n^{4/3} \cdot O(\log n)^{1/3}$ time.*

Theorem 3.4 *Given n points in $\mathbf{R}^d$, there exists a data structure of size m (for any $n \leq m \leq n^d$), which allows simplex range searching to be done in time $O(n^{1+\varepsilon}/m^{1/d})$ per query, for any fixed $\varepsilon > 0$.*

A far more involved application of cuttings and the discrepancy method gives the following result (and its corollary), which was proven by Chazelle [12]. The complexity is tight in the worst case.

Theorem 3.5 *The convex hull of a set of n points in $\mathbf{R}^d$ can be computed deterministically in $O(n \log n + n^{\lfloor d/2 \rfloor})$ time, for any fixed $d > 1$.*

Theorem 3.6 *The Voronoi diagram of a set of n points in $\mathbf{E}^d$ can be computed deterministically in $O(n \log n + n^{\lceil d/2 \rceil})$ time, for any fixed $d > 1$.*

Applications to linear and quadratic programming include the following results by Chazelle and Matoušek [18].

Theorem 3.7 *The ellipsoid of minimum volume that encloses a set of n points in $\mathbf{R}^d$ can be computed in time $d^{O(d^2)}n$.*

Theorem 3.8 *Linear programming with n constraints and d variables can be solved in $d^{O(d)}n$ time.*

These last two results build on important previous work. In particular, we mention the general formalism for linear programming developed by Sharir and Welzl [43], known as *LP-type*. The first algorithm for linear programming with a running time linear in the number of constraints was found by Megiddo [32, 33]. Subsequent improvements were found in [19, 21, 23, 24, 42].

4 Linear Circuit Complexity

Let A be an n-by-n matrix with 0/1 elements. Consider the task of assembling A by forming a sequence of column vectors $U_1, \ldots, U_s \in \mathbf{Z}^n$, where $s \geq n$ and (i) $(U_1, \ldots, U_n)$ is the n-by-n identity matrix; (ii) $A = (U_{s-n+1}, \ldots, U_s)$; and (iii) for any $i = n+1, \ldots, s$, there exist $j, k < i$ and $\alpha_i, \beta_i \in \mathbf{Z}$, such that $U_i = \alpha_i U_j + \beta_i U_k$. The minimum length s of any sequence that satisfies these three conditions is called the *complexity* of A. It is easy to see that all 0/1 matrices have complexity $O(n^2)$ and that a random one has complexity $\Omega(n^2/\log n)$.

The complexity of A is the same as the linear circuit complexity of computing $A^T x$. (A circuit consists of gates that can add linear forms.) For the case where $|\alpha_i|, |\beta_i| = O(1)$ (which is to be understood from now on), Chazelle's *spectral lemma* [14] gives us a line of attack:

Lemma 4.1 *The complexity of an n-by-n 0/1 matrix A is $\Omega(\max_k k \log \lambda_k)$, where λ_k is the k-th largest eigenvalue of $A^T A$.*

Of course, the same lemma applies to the circuit complexity as well. A recent variant by Chazelle and Lvov [17] gives us another powerful tool which bypasses the need to bound individual eigenvalues.

Lemma 4.2 *The complexity of an n-by-n 0/1 matrix A is*

$$\Omega_\varepsilon\Big(n \log\Big(\operatorname{tr} M/n - \varepsilon\sqrt{\operatorname{tr} M^2/n}\,\Big)\Big),$$

where $M = A^T A$ and $\varepsilon > 0$ is an arbitrarily small constant.

The complexity of range searching relates to the complexity of certain geometric matrices. A box matrix refers to a set system formed by points and axis-parallel boxes. Simplex matrices, on the other hand, denote the incidence matrices of set systems formed by points and simplices in $\mathbf{R}^d$. The following results, by Chazelle [9, 10, 13, 14], make heavy use of the discrepancy method.

Theorem 4.3 *There are n-by-n box matrices of circuit complexity* $\Omega(n \log\log n)$ *in* $\mathbf{R}^2$ *and monotone circuit complexity* $\Omega(n(\log n/\log\log n)^{d-1})$ *in* $\mathbf{R}^d$.

Theorem 4.4 *There are n-by-n simplex matrices of circuit complexity* $\Omega(n \log n)$ *and monotone circuit complexity* $\Omega(n^{4/3})$ *in* $\mathbf{R}^2$.

Recall that the monotone circuit model disallows the use of subtraction. While the monotone complexity of these problems is essentially resolved (there are quasi-matching upper bounds), the nonmonotone case is still wide open.

References

[1] Agarwal, P.K. Partitioning arrangements of lines II: Applications, *Disc. Comput. Geom.* 5 (1990), 533–573.

[2] Agarwal, P.K. Geometric partitioning and its applications, in *Computational Geometry: Papers from the DIMACS Special Year,* eds., Goodman, J.E., Pollack, R., Steiger, W., Amer. Math. Soc., 1991.

[3] Agarwal, P.K., Erickson, J. Geometric range searching and its relatives, in *Advances in Discrete and Computational Geometry,* eds. Chazelle, B., Goodman, J.E., Pollack, R., *Contemporary Mathematics* 223, Amer. Math. Soc., 1999, pp. 1–56.

[4] Alon, N., Spencer, J.H. *The Probabilistic Method*, Wiley-Interscience, 1992.

[5] Beck, J. Irregularities of distribution, I, *Acta Math.* 159 (1987), 1–49.

[6] Beck, J., Chen, W.W.L. *Irregularities of Distribution*, Cambridge Tracts in Mathematics, 89, Cambridge University Press, 1987.

[7] Beck, J., Fiala, T. "Integer-making" theorems, *Discrete Applied Mathematics* 3 (1981), 1–8.

[8] Beck, J., Sós, V.T. Discrepancy theory, in *Handbook of Combinatorics,* Chap. 26, eds., Graham, R.L., Grötschel, M., Lovász, L., North-Holland, 1995, pp. 1405–1446.

[9] Chazelle, B. Lower bounds on the complexity of polytope range searching, *J. Amer. Math. Soc.* 2 (1989), 637–666.

[10] Chazelle, B. Lower bounds for orthogonal range searching: II. The arithmetic model, *J. ACM* 37 (1990), 439–463.

[11] Chazelle, B. Cutting hyperplanes for divide-and-conquer, *Disc. Comput. Geom.* 9 (1993), 145–158.

[12] Chazelle, B. An optimal convex hull algorithm in any fixed dimension, *Disc. Comput. Geom.* 10 (1993), 377–409.

[13] Chazelle, B. Lower bounds for off-line range searching, *Disc. Comput. Geom.* 17 (1997), 53–65.

[14] Chazelle, B. A spectral approach to lower bounds with applications to geometric searching, *SIAM J. Comput.* 27 (1998), 545–556.

[15] Chazelle, B. *The Discrepancy Method: Randomness and Complexity*, Cambridge University Press, 2000.

[16] Chazelle, B., Friedman, J. A deterministic view of random sampling and its use in geometry, *Combinatorica* 10 (1990), 229–249.

[17] Chazelle, B., Lvov, A. A trace bound for the hereditary discrepancy, *Proc. 16th Annual ACM Symp. Comput. Geom.* (2000), 64–69. To appear in Disc. Comput. Geom.

[18] Chazelle, B., Matoušek, J. On linear-time deterministic algorithms for optimization problems in fixed dimension, *J. Algorithms* 21 (1996), 579–597.

[19] Clarkson, K.L. Linear programming in $O(n \times 3^{d^2})$ time, *Inform. Process. Lett.* 22 (1986), 21–24.

[20] Clarkson, K.L. New applications of random sampling in computational geometry, *Disc. Comput. Geom.* 2 (1987), 195–222.

[21] Clarkson, K.L. Las Vegas algorithms for linear and integer programming when the dimension is small, *J. ACM* 42 (1995), 488–499.

[22] Davenport, H. Note on irregularities of distribution, *Mathematika* 3 (1956), 131–135.

[23] Dyer, M.E. On a multidimensional search technique and its application to the Euclidean one-centre problem, *SIAM J. Comput.* 15 (1986), 725–738.

[24] Dyer, M.E., Frieze, A.M. A randomized algorithm for fixed-dimensional linear programming, *Mathematical Programming* 44 (1989), 203–212.

[25] Hammersley, J.M. Monte Carlo methods for solving multivariable problems, *Ann. New York Acad. Sci.* 86 (1960), 844-874.

[26] Haussler, D., Welzl, E. ε-nets and simplex range queries, *Disc. Comput. Geom.* 2 (1987), 127–151.

[27] Matoušek, J. Construction of ε-nets, *Disc. Comput. Geom.* 5 (1990), 427–448.

[28] Matoušek, J. Geometric range searching, *ACM Comput. Surv.* 26 (1994), 421–461.

[29] Matoušek, J. Approximations and optimal geometric divide-and-conquer, *J. Comput. Syst. Sci.* 50 (1995), 203–208.

[30] Matoušek, J. Derandomization in computational geometry, *J. Algorithms* 20 (1996), 545–580.

[31] Matoušek, J. *Geometric Discrepancy: An Illustrated Guide*, Algorithms and Combinatorics, 18, Springer, 1999.

[32] Megiddo, N. Linear-time algorithms for linear programming in R^3 and related problems, *SIAM J. Comput.* 12 (1983), 759–776.

[33] Megiddo, N. Linear programming in linear time when the dimension is fixed, *J. ACM* 31 (1984), 114–127.

[34] Montgomery, H.L. On irregularities of distribution, in *Congress of Number Theory* (Zarautz, 1984), Universidad del País Vasco, Bilbao, 1989, pp. 11–27.

[35] Montgomery, H.L. *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics, No. 84, Amer. Math. Soc., Providence, 1994.

[36] Motwani, R., Raghavan, P. *Randomized Algorithms*, Cambridge University Press, 1995.

[37] Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF, SIAM, Philadelphia, PA, 1992.

[38] Pach, J., Agarwal, P.K. *Combinatorial Geometry*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., 1995.

[39] Roth, K.F. On irregularities of distribution, *Mathematika* 1 (1954), 73–79.

[40] Roth, K.F. Remark concerning integer sequences, *Acta Arithmetica* 9 (1964), 257–260.

[41] Schmidt, W.M. Irregularities of distribution, VII, *Acta Arithmetica* 21 (1972), 45–50.

[42] Seidel, R. Small-dimensional linear programming and convex hulls made easy, *Disc. Comput. Geom.* 6 (1991), 423–434.

[43] Sharir, M., Welzl, E. A combinatorial bound for linear programming and related problems, *Proc. 9th Annual Symp. Theoret. Aspects Comput. Sci.,* LNCS, 577, Springer-Verlag, 1992, pp. 569–579.

[44] Spencer, J. Six standard deviations suffice, *Trans. Amer. Math. Soc.* 289 (1985), 679–706.

[45] Spencer, J. *Ten Lectures on the Probabilistic Method*, CBMS-NSF, SIAM, 1987.

[46] van der Corput, J.G. Verteilungsfunktionen I. *Proc. Nederl. Akad. Wetensch.* 38 (1935), 813–821.

[47] van der Corput, J.G. Verteilungsfunktionen II. *Proc. Nederl. Akad. Wetensch.* 38 (1935), 1058–1066.

[48] Vapnik, V.N., Chervonenkis, A.Ya. On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications* 16 (1971), 264–280.