



Designs in Finite Metric Spaces: A Probabilistic Approach

Minjia Shi¹ · Olivier Rioul² · Patrick Solé³

Received: 2 July 2020 / Revised: 20 May 2021 / Accepted: 22 May 2021
© The Author(s), under exclusive licence to Springer Japan KK, part of Springer Nature 2021

Abstract

A finite metric space is called here distance degree regular if its distance degree sequence is the same for every vertex. A notion of designs in such spaces is introduced that generalizes that of designs in Q -polynomial distance-regular graphs. An approximation of their cumulative distribution function, based on the notion of Christoffel function in approximation theory is given. As an application we derive limit laws on the weight distributions of binary orthogonal arrays of strength going to infinity. An analogous result for combinatorial designs of strength going to infinity is given.

Keywords Distance-regular graphs · Designs · Orthogonal polynomials · Christoffel function

Mathematics Subject Classification Primary 05E35 · Secondary 05E20 · 05E24

1 Introduction

In a celebrated paper [18] Sidelnikov proved that the weight distribution of binary codes of dual distance d^\perp going to infinity with the length is close deviates from the normal law up to a term in inverse square root of the dual distance [16, Chap. 9, §10]. Since the times of Delsarte [5], it is known that the quantity $d^\perp - 1$ is the strength of the code viewed as an orthogonal array. Further, this is a special case of designs in so-called Q -polynomial association schemes [2, 4, 5, 16]. The name

✉ Patrick Solé
sole@enst.fr

¹ Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Mathematical Sciences, Anhui University, Hefei 230601, China

² Telecom ParisTech, Palaiseau, France

³ I2M, (Aix-Marseille Univ., Centrale Marseille, CNRS), Marseille, France

designs comes from the Johnson scheme where this notion coincides with that of classical combinatorial designs [5]. Later a similar connection was found between designs in association schemes and designs in lattices [6, 20]. These kinds of generalized designs are popular now in view of the applications in random network coding [8]. In view of this deep connection it is natural to seek to extend Sidelnikov's theorem to designs in other Q -polynomial association schemes than the Hamming scheme. This is a vast research program which might take several years to accomplish.

In the present paper our contribution is twofold.

Firstly, we develop a theory of designs in finite metric spaces that replaces the concept of designs in Q -polynomial association schemes, when the considered metric space does not afford that structure. We observe that, in contrast with Delsarte definition of a design in a Q -polynomial association scheme, our definition (Definition 1 below), has an immediate combinatorial meaning in terms of distribution of distances. To wit, the combinatorial meaning of designs in certain Q -polynomial association schemes was only derived in [20] in 1986, thirteen years after Delsarte introduced designs in Q -polynomial association schemes in [5]. In particular, the example of permutations with distance the Hamming metric cannot be handled in the context of Q -polynomial association schemes, but can be treated within our framework. This space had been studied extensively in the context of permutations codes [3, 22]. The notion of t -design in that space is related to t -transitive permutation groups (Theorem 8).

Secondly, we use the technique of Chebyshev–Markov–Stieltjes inequalities in conjunction with orthogonal polynomials to control the difference between the cumulative distribution function of weights in designs with that of weights in the whole space. While this technique has been applied by Bannai to the weight distribution of spherical designs [1], it has not appeared in the literature of algebraic combinatorics so far. The bounding quantity in that setting is the Christoffel function, the inverse of the confluent Christoffel–Darboux kernel. While it is easy to make this quantity explicit in low strength cases, it is difficult to find asymptotic bounds. In the present paper, we will use the bounds of Krasikov on the Christoffel–Darboux kernel of binary Krawtchouk polynomials [13, 14] to derive an alternative proof of the Sidelnikov Theorem. We will give a proof of an analogous result for combinatorial designs by using the limiting behavior of Hahn polynomials.

The material is organized as follows. The next section contains background material on metric spaces, distance-regular graphs and Q -polynomial association schemes. Section 3 introduces the definitions that are essential to our approach. Section 4 contains the main equivalences of our notion of designs in finite metric spaces. Section 5 develops the bounds on the cumulative distribution functions of designs. Section 6 contains some asymptotic results. Section 7 recapitulates the results obtained and gives some significant open problems.

2 Background Material

2.1 Metric Spaces

Throughout the paper we write X for a finite set equipped with a metric d , that is to say a map $X \times X \rightarrow \mathbb{N}$ verifying the following three axioms

1. $\forall x, y \in X, d(x, y) = d(y, x)$
2. $\forall x, y \in X, d(x, y) = 0$ iff $x = y$
3. $\forall x, y, z \in X, d(x, y) \leq d(x, z) + d(z, y)$.

In particular if X is the vertex set of a graph the shortest path distance on X is a metric.

The *diameter* of a finite metric space is the largest value the distance may take. A finite metric space is *Distance Degree Regular* (DDR) if for every integer i less than the diameter the number $|\{y \in X \mid d(x, y) = i\}|$ is a constant v_i that does not depend of the choice of $x \in X$.

Example Consider the symmetric group on n letters S_n with metric

$$d_S(\sigma, \theta) = n - F(\sigma\theta^{-1}),$$

where $F(v)$ denotes the number of fixed points of v . The space (S_n, d_S) is a DDR metric space. If $D_m = m! \sum_{j=1}^m \frac{(-1)^j}{j!}$ denotes the number of fixed-point-free permutations of S_m , (the so-called *dérangement* number), then

$$v_i = \binom{n}{i} D_i.$$

Note that d_S is not a shortest path distance since $d_S(\sigma, \theta) = 1$ is impossible. Codes in (S_n, d_S) were studied in [22] by using the conjugacy scheme of the group S_n . However, in contrast with the next two subsections, this scheme is neither induced by a graph nor Q -polynomial.

2.2 Distance-Regular Graphs

All graphs in this article are finite, undirected, connected, without multiple edges. The neighborhood $\Gamma(x)$ is the set of vertices connected to x . The *degree* of a vertex x is the size of $\Gamma(x)$. A graph is *regular* if every vertex has the same degree. The i -neighborhood $\Gamma_i(x)$ is the set of vertices at geodetic distance i to x . The *diameter* of the graph, denoted by d is the maximum i such that for some vertex x the set $\Gamma_i(x)$ is nonempty. A graph is *distance degree regular* (DDR for short) if all graphs Γ_i , for $i = 1, \dots, d$ are regular. A graph is *distance regular* (DR for short) if for every two vertices u and v at distance i from each other the values $b_i = |\Gamma_{i+1}(u) \cap \Gamma(v)|$, $c_i = |\Gamma_{i-1}(u) \cap \Gamma(v)|$ depend only on i and do not depend on the choice of u and v . In this case, the graphs Γ_i are regular of degree v_i and we will refer to the v_i s as the

valencies of Γ ; the sequence $\{b_0, \dots, b_{\text{diam}-1}; c_1, \dots, c_{\text{diam}}\}$ is usually called the *intersection array* of Γ . Thus every DR graph is DDR but not conversely.

Example For background material on the following two examples we refer to [4, 5, 16].

1. The *Hamming graph* $H(n, q)$ is a graph on \mathbb{F}_q^n two vertices being connected if they differ in exactly one coordinate. This graph is DR with valencies

$$v_i = \binom{n}{i}(q-1)^i.$$

2. The *Johnson graph* $J(v, d)$ is a graph on the subsets of cardinality d of a set of cardinality v . (Assume $2d \leq v$). Two subsets are connected iff they intersect in exactly $d - 1$ elements. This graph is DR with valencies

$$v_i = \binom{d}{i} \binom{v-d}{i}.$$

Note that $J(v, d)$ can be embedded in $H(v, 2)$ by identifying subsets and characteristic vectors.

2.3 Q-Polynomial Association Schemes

An *association scheme* on a set X with s classes is a partition of the cartesian product $X \times X = \cup_{i=0}^s R_i$ with the following properties

1. $R_0 = \{(x, x) \mid x \in X\}$
2. $(x, y) \in R_k$, iff $(y, x) \in R_k$,
3. if $(x, y) \in R_k$, the number of $z \in X$ such that $(x, z) \in R_i$, and $(z, y) \in R_j$, is an integer p_{ij}^k that depends on i, j, k but not on the special choice of x and y

A consequence of axiom 3 is that each graph R_i is regular of degree v_i , say. It can be shown that the adjacency matrices D_k of the relations R_k span a commutative algebra over the complex with idempotents J_j [16, Chap. 21]. Let $\mu_j = \text{rank}(J_j)$. The *first eigenvalues* $p_k(i)$ of the scheme are defined by $D_k J_i = p_k(i) J_i$. Considering the matrix $P = (p_k(i))$ and writing $PQ = |X|I$, with I an identity matrix defines the *second eigenvalues* $q_k(i)$ of the scheme by the relation $Q_{ik} = q_k(i)$. A scheme is said to be *Q-polynomial* if there are numbers z_0, z_1, \dots, z_s such that $q_k(i) = \Phi_k(z_i)$ for some polynomials Φ_k of degree k . In view of the orthogonality relation [16, Chap 21, (17)]

$$\sum_{k=0}^s v_k q_i(k) q_j(k) = |X| \mu_i \delta_{ij},$$

we see that the $\frac{\Phi_i(z)}{\sqrt{\mu_i}}$ form a system of orthonormal polynomials for the scalar product

$$\langle f, g \rangle = \sum_{k=0}^s \frac{v_k}{|X|} f(z_k)g(z_k).$$

Example In both Hamming and Johnson schemes we have $z_k = k$.

1. If $\Gamma = H(n, q)$ then $\Phi_k(z) = \frac{K_k(z)}{\sqrt{v_k}}$, where K_k is the Krawtchouk polynomial of degree k given by the generating function

$$\sum_{k=0}^n K_k(x)z^k = (1 + (q - 1)x)^{n-x}(1 - z)^x.$$

2. If $\Gamma = J(v, n)$ then $\Phi_k(z) = \frac{H_k(z)}{\sqrt{v_k}}$, where H_k is the Hahn polynomial of degree k given, as per [7, (19) p. 2481], by the formula

$$H_k(z) = m_k \sum_{j=0}^k (-1)^j \frac{\binom{k}{j} \binom{v+1-k}{j}}{\binom{n}{j} \binom{v-n}{j}} \binom{z}{j},$$

where $m_k = \binom{v}{k} - \binom{v}{k-1}$.

3 Preliminaries

For any integer $\mathcal{N} > 0$, denote by $[0.. \mathcal{N}]$ the set of integers in the range $[0, \mathcal{N}]$. A finite metric space (X, d) is *distance degree regular* (DDR) if its distance degree sequence is the same for every point. Assume (X, d) to be of diameter n . In that case (X, d) is DDR iff for each $0 \leq i \leq n$ the graph $\Gamma_i = (X, E_i)$ which connects vertices at distance i in (X, d) is regular of degree v_i . Thus $E_0 = \{(x, x) \mid x \in X\}$ is the diagonal of X^2 . Note that the E_i 's form a partition of X^2 .

If D is any non void subset of X we define its *frequencies* as

$$\forall i \in [0..n], f_i = \frac{|D^2 \cap E_i|}{|D|^2}.$$

Thus $f_0 = \frac{1}{|D|}$, and $\sum_{i=0}^n f_i = 1$. Note also that if $D = X$, then $f_i = \frac{v_i}{|X|}$. Consider the random variable a_D defined on D^2 with values in $[0..n]$ which affects to an equiprobably chosen $(x, y) \in D^2$ the only i such that $(x, y) \in E_i$. Thus the frequencies $f_i = Prob(a_D = i)$. Denote by $\mathbb{E}()$ mathematical expectation. Thus

$$\mathbb{E}(a_D^i) = \sum_{j=0}^n f_j j^i, \quad \mathbb{E}(a_X^i) = \sum_{j=0}^n \frac{v_j}{|X|} j^i.$$

Definition 1 The set $D \subseteq X$ is a *t-design* for some integer t if

$$\mathbb{E}(a_D^i) = \mathbb{E}(a_X^i)$$

for $i = 1, \dots, t$.

(Note that trivially $\mathbb{E}(1) = 1$ so that we do not consider $i = 0$.) Thus, distances in t -designs are very regularly distributed. For a 2-design, for instance, the average and variance of the distance coincide with that of the whole space. We will see in the next section that in the case of Hamming and Johnson graphs, we obtain classical combinatorial objects: block designs, orthogonal arrays.

Lemma 1 *We define a scalar product on $\mathbb{R}[x]$ attached to D by the relation*

$$\langle f, g \rangle_D = \sum_{i=0}^n f(i)g(i).$$

Thus, in the special case of $D = X$ we have

$$\langle f, g \rangle_X = \frac{1}{|X|} \sum_{i=0}^n v_i f(i)g(i).$$

We shall say that a sequence $\Phi_i(x)$ of polynomials of degree i is *orthonormal* of size $N + 1$ if it satisfies

$$\forall i, j \in [0..N], \langle \Phi_i, \Phi_j \rangle_X = \delta_{ij},$$

where $N \leq n$, the letter δ denotes the Kronecker symbol. That sequence is uniquely defined if we assume the leading coefficient of all $\Phi_i(x)$ for $i = 0, 1, \dots, N$ to be positive.

For a given DDR metric space (X, d) , we shall denote by $N(X)$ the largest possible such N . For instance if X is an n -class P - and Q -polynomial association scheme, it is well-known that $N(X) = n$. This fact is extended to DDR graphs in the next Proposition.

Proposition 1 *If none of the v_i 's are zero, then $\langle \cdot, \cdot \rangle_X$ admits an orthonormal system of polynomials of size $n + 1$. In particular, the metric space of a DDR graph admits an orthonormal system of polynomials of size $n + 1$.*

Proof By Lagrange interpolation we see that the functions $1, x, \dots, x^n$ are linearly independent on $[0..n]$. The sequence of the Φ_i 's for $i = 0, 1, \dots, n$ is then constructed by the usual Gram-Schmidt orthogonalization process. Note that this is possible because the bilinear form $\langle \cdot, \cdot \rangle_X$ is then nondegenerate: $\langle f, f \rangle_X = 0 \Rightarrow f = 0$. By properties of the shortest path distance, the property of non vanishing of the v_i 's holds in particular for the metric space of a DDR graph. \square

Definition 2 For a given $D \subseteq X$ the *dual frequencies* are defined for $i = 0, 1, \dots, N(X)$ as

$$\widehat{f}_i = \sum_{k=0}^n \Phi_i(k) f_k.$$

Definition 3 For a given $D \subseteq X$ the *cumulative distribution function* (c.d.f.) is defined as

$$F_D(x) = \text{Prob}(a_D \leq x) = \sum_{i \leq x} f_i.$$

Example

1. If D is a linear code of $H(n, q)$, with weight distribution

$$A_i = |\{x \in D \mid w_H(x) = i\}|,$$

then $F_D(x) = \frac{\sum_{i \leq x} A_i}{|D|}$.

2. If D is a set of points in $J(v, k)$, with Hamming distance distribution B_{2i} in $H(v, 2)$, then $F_D(x) = \frac{\sum_{i \leq x} B_{2i}}{|D|}$.

4 Structure Theorems

First, we give a characterization of t -designs in terms of dual frequencies.

Proposition 2 *Let t be an integer $\in [1..N(X)]$. The set $D \subseteq X$ is a t -design iff $\widehat{f}_i = 0$ for $i = 1, \dots, t$.*

Proof Note first that

$$\mathbb{E}(a_D^i) = \langle x^i, 1 \rangle_D, \mathbb{E}(a_X^i) = \langle x^i, 1 \rangle_X.$$

Moving the basis of $\mathbb{R}[x]$ from the Φ_i 's to the basis of monomials we see that D is a t -design if and only if for $i = 1, 2, \dots, t$, we have

$$\langle \Phi_i, 1 \rangle_D = \langle \Phi_i, 1 \rangle_X.$$

Now, by definition, the dual frequency $\widehat{f}_i = \langle \Phi_i, 1 \rangle_D$. By orthogonality of the Φ_i 's for the scalar product $\langle \cdot, \cdot \rangle_X$, we see that $\langle \Phi_i, 1 \rangle_X = 0$ for $i = 1, 2, \dots, t$. Thus, the condition $\widehat{f}_i = 0$ for $i = 1, 2, \dots, t$, is equivalent to the fact that D is a t -design. □

Next, we connect the notion of designs in Q -polynomial association schemes with our notion of designs in metric spaces.

Theorem 1 *If (X, d) is the metric space induced by a Q -polynomial DR graph Γ , with $z_k = k$ for $k = 0, 1, \dots, n$, then a t -design in (X, d) is exactly a t -design in the underlying association scheme of Γ .*

Proof In that situation the frequencies are proportional to the inner distribution (see [4, p. 54]) of D in the scheme of the graph, and the dual frequencies are proportional to the dual inner distribution since the second eigenvalues of the scheme, by the Q -polynomiality condition, are orthogonal polynomials w.r.t. the distribution $\frac{v_i}{|X|}$. The result follows. □

Example The following two examples of interpretation of t -designs as classical combinatorial objects were observed first in [5] and can be read about in [16, chap. 21].

1. If Γ is the Hamming graph $H(n, q)$ then a t -design is an orthogonal array of strength t . That means that every row induced by a t -uple columns of D sees the q^t possible values a constant number of times.
2. If Γ is the Johnson graph $J(v, n)$ then a t -design D is a combinatorial design of strength t . This means the following. Consider D as a collection of subsets of size n , traditionally called blocks. That means that every t -uple of elements of the groundset is contained in the same number η of blocks. One says that D is a $t - (v, n, \eta)$ design.

Now, we give an example of t -design in a metric space that is not a DR graph, or even a DDR graph.

Theorem 2 *If $D \subseteq S_n$ is a t -transitive permutation group then it is a t -design in (S_n, d_S) .*

Proof The moments of order $i \leq t$ of the number of fixed points of the permutations in D coincide with those of a Poisson law of parameter one. This is a result of Frobenius [24]. A modern exposition is in [12, Chap. 5.5]. The result follows by the definition of d_S . □

5 Distribution Functions

In this section we show that the distribution function of designs are close to that of the whole space. The proof of the following result follows the philosophy of [1].

Theorem 3 *Let D be a t -design in (X, d) , with $t \leq N(X)$. Put $\kappa = \lfloor \frac{t}{2} \rfloor$. Denote by $\lambda(x)$ the Christoffel function given by $\lambda(x) = (\sum_{i=0}^{\kappa} \Phi_i(x)^2)^{-1}$. Then we have the bound*

$$|F_D(x) - F_X(x)| \leq \lambda(x).$$

Proof By Definition 1, we have

$$\langle x^i, 1 \rangle_X = \langle x^i, 1 \rangle_D \text{ for } i = 0, 1, 2, \dots, t.$$

The orthonormal polynomials for $\langle \cdot, \cdot \rangle_X$ exist for degrees $\leq t$ by the hypothesis $t \leq N(X)$. A coincidence of moments up to order t entails a coincidence of orthonormal polynomials up to degree κ by Chebyshev determinant for orthonormal polynomials [11, Lemma 2.1] (see also [21, (2.2.6), p. 27]). By the same formula, the orthonormal polynomials for $\langle \cdot, \cdot \rangle_D$ are well-defined for degrees $\leq \kappa$, since the orthonormal polynomials up to degree κ attached to X exist. By [11, Th. 4.1] or [18, Th. 7.2] we have the Markov–Stieltjes inequalities

$$\sum_{x_i < x} \lambda(x_i) \leq F_D(x) \leq \sum_{x_i < x} \lambda(x_i) + \lambda(x), \tag{1}$$

where the x_i 's are the κ zeros of $\Phi_\kappa(t)$. Similarly we have

$$\sum_{x_i < x} \lambda(x_i) \leq F_X(x) \leq \sum_{x_i < x} \lambda(x_i) + \lambda(x). \tag{2}$$

The result follows upon combining equations (1) and (2). □

As a bound uniform in x , we have the following result.

Corollary 1 *If D is a binary orthogonal array of strength at least five, then its c.d.f. is close to that of the binomial distribution as*

$$|F_D(x) - F_X(x)| < \frac{2(n-1)}{3n-2}$$

Proof We compute explicit lower bounds on $1/\lambda$ by using the first three Krawtchouk polynomials [16, Chap. 5 §7] given by

$$K_0 = 1, \quad K_1(x) = n - 2x, \quad K_2(x) = 2x^2 - 2nx + \binom{n}{2}.$$

We are seeking a lower bound for

$$1 + \frac{K_1(x)^2}{n} + \frac{K_2(x)^2}{\binom{n}{2}},$$

when $x \in [0, n]$. Making the change of variable $y = n - 2x \in [-n, n]$, we obtain

$$K_2(x) = \frac{y^2 - n}{2},$$

and, therefore

$$1/\lambda = 1 + \frac{y^2}{n} + \frac{(y^2 - n)^2}{2n(n-1)} = \frac{3n^2 - 2n + (y^2 - 1)^2 - 1}{2n(n-1)},$$

an increasing function of y^2 that takes its minimum over $[0, n^2]$ at $y = 0$. □

Example If D is the extended Hamming code of length $n = 16$, dual distance 8, the weight distribution is, in Magma notation [15], equal to

$$[\langle 0, 1 \rangle , \langle 4, 140 \rangle , \langle 6, 448 \rangle , \langle 8, 870 \rangle , \langle 10, 448 \rangle , \langle 12, 140 \rangle , \langle 16, 1 \rangle] .$$

For $x = 8$, we get $F_D(x) = \frac{1+140+448+870}{2^{11}} \approx 0.712$, and $F_X(x) = \frac{\sum_{j=0}^8 \binom{n}{j}}{2^{11}} \approx 0.598$. The difference is $\approx 0.112 < \frac{3 \times 15}{46} \approx 0.652$.

We give three bounds that are not uniform in x . First, for orthogonal arrays.

Corollary 2 *If D is a q -ary orthogonal array of strength at least two, then its c.d.f. is as close to that of the binomial distribution as*

$$|F_D(x) - F_X(x)| < \frac{n}{n + (n(q - 1) - qx)^2} .$$

Proof Immediate from the data of the first two Krawtchouk polynomials:

$$K_0 = 1, \quad K_1(x) = n(q - 1) - qx .$$

□

Example If D is a binary Simplex code of length $n = 2^m - 1$, there is a unique nonzero weight, namely $\frac{n+1}{2}$ that appears $|D| - 1 = n$ times. If we compute the bound for $x = \frac{n+1}{2}$, its right hand side is $\frac{n}{n+1}$, which is also the value of $F_D(x)$ while $F_X(x) > 0.5$.

Next, we consider combinatorial designs.

Corollary 3 *If D is a $2 - (v, n, \lambda)$ design, then its c.d.f. is as close to that of the hypergeometric distribution as*

$$|F_D(x) - F_X(x)| < \frac{(v - n)^3}{(v - n)^3 + n(v - 1)^2} .$$

Proof From the data of the first two Hahn polynomials:

$$H_0 = 1, \quad H_1(x) = (v - 1) \left(1 - \frac{vx}{n(v - n)} \right) ,$$

we obtain

$$1/\lambda = 1 + \frac{H_1(x)^2}{n(v - n)} = 1 + (v - 1)^2 \frac{(n^2 - nv + vx)^2}{n^3(v - n)^2} ,$$

a monotonic function of x .

□

Eventually, we consider permutation groups. Exceptionally, we do not consider the distance but the codistance $n - d_s$.

Corollary 4 *If D is a 2-transitive permutation group on n letters, then the c.d.f. of its fixed points $G_D(x)$ is as close to that of the Poisson law of parameter one $P(x) =$*

$$\sum_{1 \leq i \leq x} \frac{1}{i!} \text{ as}$$

$$|G_D(x) - P(x)| < \frac{n}{n + (1 - x)^2}.$$

Proof Immediate from the data of the first two Charlier polynomials $C_0 = 1, C_1(x) = 1 - x$, obtained from the generating series

$$e^t(1 - t)^x = \sum_{n=0}^{\infty} C_n(x) \frac{t^n}{n!}$$

of [17, (1.12.11)]. □

6 Asymptotic Results

6.1 Orthogonal Arrays

In this section we give an alternative proof of a result of Sidelnikov on the weight enumerator of long codes [18]. We prepare for the proof by a form of the Central Limit Theorem for the binomial distribution. Denote by $\Psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{t^2}{2}) dt$ the cumulative distribution function of the centered normal law of variance unity.

Let $B_n(x) = \sum_{i \leq x} \frac{\binom{n}{i}}{2^n}$ denote the cumulative distribution function of the binomial law (sum of n Bernoulli trials).

Theorem 4 *For some absolute constant $C > 0$, we have*

$$|B_n(x) - \Psi(x)| \leq \frac{C}{\sqrt{n}}.$$

Proof Immediate by Berry–Essen theorem [9]. □

Recall the binary entropy function [16] defined as

$$H(x) = -x \log_2 x - (1 - x) \log_2(1 - x).$$

A tedious but straightforward consequence of Stirling formula is

$$\binom{N}{\alpha N} \sim \frac{2^{NH(x)}}{\sqrt{2\pi\alpha(1-\alpha)}} \tag{3}$$

for $N \rightarrow \infty$ and $0 < \alpha < 1$. See (1) in [10].

Theorem 5 *Let $n \rightarrow \infty$, and let k be an integer such that $k \sim \theta n$, with $0 < \theta < 1$ a real constant. Assume $x = \frac{n}{2} + O(\sqrt{n})$. Then any binary orthogonal array D with n columns, of strength $\geq 2k + 1$, satisfies*

$$|F_D(x) - B_n(x)| = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof (sketch) We use [13, Th. 1.1] or [14, Lemma 4] to claim the lower bound

$$\frac{1}{\lambda(x)} = \Omega\left(\frac{\binom{k+1}{2} G_k(x)}{\binom{n}{k}}\right),$$

where

$$G_k(x) = \frac{2k(2x + p_k - n)(n - k)^2 \Gamma(x) \Gamma(n - x)}{n^3(p_k + 2) \Gamma(\frac{n}{2} + 1) \Gamma(\frac{n}{2} - 1)} \left(\frac{n/2}{k/2}\right)^2,$$

where $p_k = 2\sqrt{k(n - k)}$ (note that μ_k tends to a constant in n).

To derive the said bound, divide numerator and denominator by n^4 and simplify. Observe that $p_k \sim 2n\sqrt{\theta(1 - \theta)}$. For the term $\binom{n/2}{k/2}$ we use the entropic estimate mentioned above. We write $\Gamma(x)\Gamma(n - x) = \frac{(n-2)!}{(x-1)}$. We use the Moivre-Laplace formula to get

$$\frac{\binom{n}{x}}{2^n} \sim \frac{\exp\left(-\frac{(x-n/2)^2}{n/2}\right)}{\sqrt{\pi n/2}} = O\left(\frac{1}{\sqrt{n}}\right),$$

where the constant implied by $O()$ is independent of x . The result follows after tedious but straightforward manipulations. □

We are now ready for the main result of this section.

Theorem 6 *Let $n \rightarrow \infty$, and let k be an integer such that $k \sim \theta n$, with $0 < \theta < 1$ a real constant. Assume $x = \frac{n}{2} + O(\sqrt{n})$. Then any binary orthogonal array D with n columns, of strength $\geq 2k + 1$, satisfies*

$$|F_D(x) - \Psi(x)| = O\left(\frac{1}{\sqrt{n}}\right).$$

Proof Immediate by combining Theorem 4 with Theorem 5. □

6.2 Designs

Note, before doing asymptotics on the strength of designs, that t designs exist for all t [23].

Let $H_k(x)$ denote the Hahn polynomial of degree k of the variable x , as defined in [7]. Let $v_k = \binom{n}{k} \binom{v-n}{k}$ be the valency of order k of the Johnson graph $J(v, n)$. We normalize $\widehat{H_k(x)} = \frac{H_k(x)}{\sqrt{v_k}}$.

Theorem 7 Assume both v and n go to infinity with $n/v \rightarrow p \in (0, 1)$. Put $q = 1 - p$. Let $z = nx$ with $x \in (0, 1)$. Then, we have for fixed k , and $n \rightarrow \infty$ the limit

$$\widehat{H_k(z)} \rightarrow \frac{(1 - x/q)^k}{\sqrt{p^k q^k}}.$$

Proof First, note that $m_k \sim \frac{v^k}{k!} \sim \frac{v^k}{p^k k!}$. Next, observe that $v_k \sim \frac{n^k (v-n)^k}{k!^2} \sim \frac{n^{2k}}{(k!)^2} (q/p)^k$. This yields $\sqrt{v_k} \sim \frac{n^k}{k!} (\sqrt{q/p})^k$. Combining we obtain $\frac{m_k}{\sqrt{v_k}} \sim 1/\sqrt{p^k q^k}$. Similar calculations give the term of order j of $H_k(xn)$ to have the limit

$$(-1)^j \binom{k}{j} \frac{(v^j) n^j x^j}{n^j (v-n)^j} \rightarrow (-1)^j \binom{k}{j} (x/q)^j,$$

and, summing on j yield

$$H_k(xn) \rightarrow \sum_{j=0}^k \binom{k}{j} (-x/q)^j = (1 - x/q)^k.$$

The result follows upon writing $\widehat{H_k(xn)} = \frac{m_k}{\sqrt{v_k}} H_k(xn)$. □

We can now derive the main result of this section.

Theorem 8 Let D be a $t - (v, n, \eta)$ design with $v, n, t \rightarrow \infty$, and t fixed and $n \sim pv$ with $0 < p < 1$ real constants. Put $q = 1 - p$, and $k = \lfloor t/2 \rfloor$. Let $J(v, n; x) = \sum_{i \leq x} \frac{v_i}{\binom{v}{n}}$. Then

$$|F_D(x) - J(v, n; x)| \leq \lambda_k(n),$$

where

$$\lim_{n \rightarrow \infty} \lambda_k(n) = \frac{1 - a(x)}{1 - a(x)^{k+1}}.$$

and $a(x) = \frac{(1-x/q)^2}{\sqrt{pq}}$.

Proof Immediate by taking the limit of the Christoffel–Darboux kernel of order k given by

$$\sum_{j=0}^k H_j(\widehat{xn})^2 = \sum_{j=0}^k \frac{H_j(xn)^2}{v_j}$$

and summing the geometric series of ratio $a(x)$ coming from Theorem 7. □

7 Conclusion

In this paper we have used a probabilistic approach to approximate the c.d.f. of designs in various finite metric spaces. The key tool is the Christoffel–Darboux kernel attached to the orthonormal polynomials w.r.t. the valencies of the space. We have used some strong analytic bounds on this quantity for binary Krawtchouk polynomials derived in [13, 14]. It would be desirable to extend these analytical results to other families of polynomials, beginning with q -ary Krawtchouk polynomials. This special case would yield an alternative proof of the q -ary version of Sidelnikov theorem proved by us in [19]. Further, it is a worthwhile project to derive analogous results for the polynomials relevant to the eight types of designs in [20]. A first step in that direction would be to extend or adapt Theorem 7 to Q -polynomial schemes where z_k is not always equal to k . Regarding more general DDR metric spaces, it would be nice to have examples of t -designs in the space of permutations that are not t -transitive permutation groups.

Acknowledgements This research is supported by National Natural Science Foundation of China (12071001, 61672036), Excellent Youth Foundation of Natural Science Foundation of Anhui Province (1808085J20), the Academic Fund for Outstanding Talents in Universities (gxbjZD03).

References

1. Bannai, E.: On the weight distribution of spherical t -designs. *Eur. J. Comb.* **1**, 19–26 (1980)
2. Brouwer, A.E., Haemers, W.H.: *Spectra of Graphs*. Springer, New York (2011)
3. Blake, I., Cohen, G.D., Deza, M.: Coding with permutations. *Inf. Control* **43**, 1–19 (1979)
4. Brouwer, A.E., Cohen, A.M., Neumaier, A.: *Distance-Regular Graphs*. Springer, Berlin (1989)
5. Delsarte, P.: An algebraic approach to the association schemes of Coding Theory. *Philips Research Report Suppl.* 10 (1973)
6. Delsarte, P.: Association schemes and t -designs in regular semi-lattices. *J. Comb. Theory A* **20**, 230–243 (1976)
7. Delsarte, P., Levenshtein, V.: Association schemes and coding theory. *IEEE Trans. Inf. Theory* **44**(6), 2477–2503 (1998)
8. Etzion, T.: Problems in q -analogs in coding theory. [arXiv:1305.6126](https://arxiv.org/abs/1305.6126)
9. Feller, W.: *Introduction to Probability theory and Its Applications*, vol. II. Wiley, New York (1968)
10. Galvin, D.: Three lectures on entropy and counting. [arXiv:1406.7872](https://arxiv.org/abs/1406.7872)
11. Hürliman, W.: An explicit version of the Chebyshev–Markov–Stieltjes inequalities and its applications. *J. Inequal. Appl.* (2015)
12. James, G., Kerber, A.: *The Representation Theory of the Symmetric Group*. Addison-Wesley, Reading (1981)
13. Krasikov, I.: Bounds for the Christoffel–Darboux kernel of the binary Krawtchouk polynomials. In: Barg, A., Litsyn, S. (eds.) *Codes and Association Schemes*, pp. 193–198. AMS (2001)

14. Krasikov, I., Litsyn, S.: On the distance distributions of BCH codes and their duals. *Des. Codes Cryptogr.* **23**, 223–231 (2001)
15. Magma.: <http://magma.maths.usyd.edu.au/magma/>
16. MacWilliams, F.J., Sloane, N.J.A.: *The Theory of Error-correcting Codes*. North-Holland, Amsterdam (1977)
17. Roelof, K., Swarttouw, R.F.: The Askey scheme of hypergeometric orthogonal polynomials and its q -analogue. [arXiv:math/9602214](https://arxiv.org/abs/math/9602214)
18. Sidelnikov, V.M.: Weight spectrum of binary Bose–Chaudhuri–Hocquenghem codes. *Problemy Pederachi Informatsii* **7**(1), 14–22 (1971)
19. Shi, M., Rioul, O., Solé, P.: On the asymptotic normality of Q -ary linear codes. *IEEE Commun. Lett.* **23**(11), 1895–1898 (2019)
20. Stanton, D.: t -designs in classical association schemes. *Graphs Comb.* **2**, 283–286 (1986)
21. Szegő, G.: *Orthogonal Polynomials*, vol. 23, 4th edn. AMS Colloquium Publications, Providence (1975)
22. Tarnanen, H.: Upper bounds on permutation codes via linear programming. *Eur. J. Comb.* **20**, 101–114 (1999)
23. Teirlinck, L.: Non-trivial t -designs without repeated blocks exist for all t . *Discrete Math.* **65**, 301–311 (1987)
24. Frobenius, G.: Über die charaktere der mehrfach transitive gruppen, *Berliner Bericht*, pp 558–571 (1904)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.