



ACADÉMIE  
DES SCIENCES  
INSTITUT DE FRANCE

# *Comptes Rendus*

---

# *Mathématique*


Jean-Renaud Pycke

**New Karhunen-Loève expansions based on Hahn polynomials with application to a Sobolev test for uniformity on Johnson graphs**

Volume 362 (2024), p. 701-711

Online since: 17 September 2024

<https://doi.org/10.5802/crmath.470>

 This article is licensed under the  
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.  
<http://creativecommons.org/licenses/by/4.0/>



*The Comptes Rendus. Mathématique* are a member of the  
Mersenne Center for open scientific publishing  
[www.centre-mersenne.org](http://www.centre-mersenne.org) — e-ISSN : 1778-3569



Research article / *Article de recherche*  
Statistics / *Statistiques*

# New Karhunen-Loève expansions based on Hahn polynomials with application to a Sobolev test for uniformity on Johnson graphs

*De nouveaux développements de Karhunen-Loève basés sur les polynômes de Hahn avec application à un test d'uniformité de Sobolev sur les graphes de Johnson*

Jean-Renaud Pycke <sup>a</sup>

<sup>a</sup> LaMME CNRS (UMR 8071), University of Paris-Saclay (Evry), France  
E-mail: [jeanrenaud.pycke@univ-evry.fr](mailto:jeanrenaud.pycke@univ-evry.fr)

**Abstract.** We give a new family of Karhunen-Loève expansions involving Hahn polynomials. This enables us to introduce discrete analogues of Watson statistics, and a test for uniformity on Johnson's graphs. We use the fact that the zonal spherical functions on these graphs are Hahn polynomials. Our test is consistent against all alternatives and locally most powerful against some alternative.

**Résumé.** Nous proposons un nouveau développement de Karhunen-Loève avec les polynômes de Hahn. Cela nous permet d'introduire une analogue discrète de la statistique de Watson pour tester l'uniformité d'une distribution sur les graphes de Johnson, dont les fonctions sphériques zonales sont les polynômes de Hahn. Ce test peut être vu comme un test de Sobolev dans le cas discret, nous en déduisons certaines de ses propriétés asymptotiques.

**Keywords.** Karhunen-Loève expansions, classical orthogonal polynomials, distance regular graphs, directional statistics.

**Mots-clés.** développements de Karhunen-Loève, polynômes orthogonaux classiques, Graphes distance-réguliers, statistiques directionnelles.

**2020 Mathematics Subject Classification.** 62F05, 33C45.

*Manuscript received 6 December 2022, accepted 29 January 2023.*

## 1. Introduction

The statistics of Cramér-von Mises (CVM) and Watson, denoted in the literature by  $W_n^2$  and  $U_n^2$  respectively, remain two of the most widely used tools for testing the uniformity of a sample from a distribution over  $[0, 1]$  or the unit circle (see, e.g., [7, §4-5, in particular formulas (4.1.7), (5.4.2) p. 27 and 36], or [6, §4.2.2-3, formulas (4.2) p. 100-101]).

Recall that in order to test uniformity on the unit circle, Watson [21] introduced  $U_n^2$ , by adapting  $W_n^2$ . The circle is a homogeneous space and the simplest example of a compact continuous

two-point homogeneous space. The field of directional statistics deals with such sample spaces, less conventional than the real line. The fundamental reference for an introduction to this quite recent domain of statistics is [13]. The latter book should be supplemented by [15], which provides a detailed survey of recent advances in the field. Characteristically, no reference dealing with data from sample spaces similar to Johnson's graphs does occur in this survey.

Beran [1] and Giné [8] introduced a wide class of tests for uniformity on homogeneous spaces and compact Riemannian manifolds, but the discrete cases have received less attention than the continuous. As a result, directional statistics dealing with discrete sample spaces, remains a widely unexplored field.

The aim of the present paper is to contribute to bridge this gap, by building a counterpart of Watson's statistic, by means of Karhunen-Loève (K-L) expansions involving classical orthogonal polynomials, so that our results can be easily extended, *mutatis mutandis* to other classical discrete orthogonal polynomials. Our test for uniformity on Johnson graphs can be considered as a Sobolev test, in the discrete case, introduced by Giné's [8].

Our paper is organized as follows. In Section 2 we show that the two Karhunen-Loève (K-L) expansions (3), associated with CVM and Watson's statistics, via their definitions (2) as V-statistics, and from which most of their properties can be derived (principal components, limiting distribution, Bahadur efficiency), are associated with Chebyshev polynomials of the first kind.

This interpretation enables us, in Theorem 1 of Section 3, to give for kernel (18) the K-L development (19), which is a discrete analogue, valid for Hahn polynomials, of development (3).

Hahn polynomials can be seen as the zonal spherical functions of Johnson graphs. Using this fact, we build in Section 4 a test for uniformity on these graphs in the same way as Watson's statistic was built for testing uniformity on a circle from CVM's statistic. Theorem 3 gives a correspondence between some K-L expansions over Johnson graphs, and a one-variable Fourier series with Hahn polynomials. Theorem 4 introduces our new statistic and gives the asymptotic distribution under the hypothesis of uniformity. Theorem 4 also provides an alternative hypothesis under which the new test is locally most powerful.

## 2. Cramér - von Mises and Watson statistics revisited

Let  $(T, \mu)$  be a probability measure space and  $L_2(T, \mu)$  the associated Hilbert space of real, square-integrable functions endowed with the inner product

$$\langle f | g \rangle_\mu = \int_T f(t)g(t)d\mu(t).$$

By a K-L expansion (with respect to  $\mu$ ) of the bivariate symmetric kernel  $K : T \times T \rightarrow \mathbb{R}$ , we mean a pointwise convergent series of the form

$$K(s, t) = \sum_{k \geq 1} \mu_k \cdot \frac{f_k(s)f_k(t)}{c_k} \quad (s, t \in T), \quad (1)$$

where the sequences of functions  $(f_k)$  and real positive numbers  $(\mu_k)$  and  $(c_k)$  satisfy the integral and orthogonality relations

$$\int_T K(s, t) f_k(s) d\mu(s) = \mu_k f_k(t), \quad \langle f_k | f_l \rangle_\mu = c_k \delta_{k, \ell} \quad (t \in T, k, \ell \geq 1),$$

with the Kronecker delta, or Dirac function, defined by

$$\delta_{i, j} = \delta_i(j) = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases} \quad (i, j \in \mathbb{Z}).$$

Concerning these expansions and statistical applications, see [18, Chapter 5]. For general facts about degenerate V-statistics, the reader is referred to [12, §4.3], [20, §12.3 (in particular

Examples 12.9 and 12.13)], [17, Chapter 5]. Whereas it is not the most standard way, CVM and Watson statistics can be defined in the form of the V-statistics (see [16, Introduction] and references therein), computed from samples  $X_1, \dots, X_n \in [0, 1]$  and  $\theta_1, \dots, \theta_n \in [0, 2\pi]$ ,

$$W_n^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n h_W(X_i, X_j), \quad U_n^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n h_U(\theta_i, \theta_j), \tag{2}$$

where the symmetric kernels and their K-L expansions (w.r.t. the uniform distribution) are given, for  $0 \leq t_1 \leq t_2 \leq 1$  or  $0 \leq \theta_1, \theta_2 \leq 2\pi$ , by

$$h_W(t_1, t_2) = \frac{t_1^2 + (1 - t_2)^2}{2} - \frac{1}{6} = \sum_{k=1}^{\infty} \frac{1}{k^2 \pi^2} \cdot 2 \cos(k\pi t_1) \cos(k\pi t_2), \tag{3}$$

$$h_U(\theta_1, \theta_2) = \frac{1}{4} h_W(0, \pi^{-1} d[P_1(\theta_1), P_2(\theta_2)]) \tag{4}$$

$$= \sum_{k=1}^{\infty} \frac{1}{4k^2 \pi^2} \cdot 2 \{ \cos(k\theta_1) \cos(k\theta_2) + \sin(k\theta_1) \sin(k\theta_2) \}. \tag{5}$$

where  $d[P_1(\theta_1), P_1(\theta_2)] = \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|) \in [0, \pi]$  denotes the distance between the circle points  $P_1$  and  $P_2$  associated with angles  $\theta_1$  and  $\theta_2$ . Therefore  $U_n^2$  does not depend on the origin or the orientation chosen arbitrarily on the circle. In other words, this statistic is well-defined for circular data, because its kernel is invariant under the action of  $O(2)$ , the group of isometries of the unit Euclidean circle. We will define an invariant kernel on Johnson graphs in a similar way by (27).

Kernels  $h_W$  and  $h_U$ , and by extension the associated  $V$ -statistics  $W_n^2$  and  $U_n^2$ , are said to be degenerate (w.r.t. the uniform distribution) in view of equalities

$$\int_0^1 h_W(t_1, t_2) dt_1 = \int_0^{2\pi} h_U(\theta_1, \theta_2) d\theta_1 = 0.$$

Let us express these relations in terms of Chebyshev polynomials of the first kind, whose sequence  $(T_k)_{k \geq 0}$  is defined by  $T_k(\cos \theta) = \cos(k\theta)$ . They satisfy, (see [11, §9.8.2]) for  $0 \leq k, \ell$ , the differential equation and the orthogonality relations

$$(\sigma(x)\omega(x)T'_k(x))' = -\lambda_k \omega(x)T_k(x), \quad \int_{-1}^1 T_k(x)T_\ell(x)\omega(x)dx = c_k \delta_{k,\ell}, \tag{6}$$

with  $\sigma(x) = 1 - x^2$ , eigenvalues  $\lambda_k = k^2$  and constants  $c_k = 1/2$  for  $k \geq 1$ , the weight function  $\omega(x) = \pi^{-1}(1 - x^2)^{-1/2} = \Omega'(x)$  being the arcsine probability density function with distribution function  $\Omega$  and the associated tail  $\bar{\Omega}$  given by

$$\Omega(x) = \int_{-1}^x \omega(y)dy = \pi^{-1} \arccos(-x) = 1 - \bar{\Omega}(x) \quad (-1 \leq x \leq 1).$$

It is easily checked that (3) can be rewritten, for  $-1 \leq x_1 \leq x_2 \leq 1$ , as

$$\begin{aligned} \pi^2 h_W(\Omega(x_1), \Omega(x_2)) &= \sum_{k=1}^{\infty} \frac{1}{\lambda_k} \frac{T_k(x_1)T_k(x_2)}{c_k} \\ &= \int_{-1}^{x_1} \frac{\Omega(u)}{\sigma(u)\omega(u)} du + \int_{x_2}^1 \frac{\bar{\Omega}(u)}{\sigma(u)\omega(u)} du - \int_{-1}^1 \frac{\Omega(u)\bar{\Omega}(u)}{\sigma(u)\omega(u)} du. \end{aligned} \tag{7}$$

Now, since expressions appearing on both sides of the second equality in (7) involve functions and constants associated with Chebyshev polynomials, having their analogues for Hahn polynomials, we are in a position to state the corresponding results.

### 3. K-L expansions with Hahn polynomials

Assume  $N$  is a positive integer and  $\mu, \nu \in (-1, \infty)$ . In the present paper, Hahn polynomials, denoted by  $H_k^{(\mu, \nu, N)}(x)$  for  $0 \leq k \leq N$ , will be those denoted by  $Q_k(x, \alpha, \beta, N)$  in [11, §9.5] (see formulas (1.4.1) p. 5 and (9.5.1) p. 204) with  $\alpha = -N - \nu - 1, \beta = -N - \mu - 1$ , or also  $\tilde{h}_k^{(\mu, \nu)}(x, N + 1) / \tilde{h}_k^{(\mu, \nu)}(0, N + 1)$  in [14] (see the first equality p. 53 and Table 2.4 p. 54), so that

$$H_k^{(\mu, \nu, N)}(x) = \sum_{j=0}^k \frac{(-k)_j (-2N - \mu - \nu + k - 1)_j (-x)_j}{(-N - \nu)_j (-N)_j j!} \quad (0 \leq k \leq N) \tag{8}$$

where the Pochhammer symbol  $(a)_k$  is defined to be

$$(a)_0 = 1, \quad (a)_k = a(a + 1) \cdots (a + k - 1) \quad (k = 1, 2, 3, \dots).$$

Following [9, (6.31) p. 259], consider the hypergeometric probability mass function (p.m.f.), supported by the set  $\llbracket 0, N \rrbracket = \{0, 1, \dots, N\}$ , given by

$$\omega^{(\mu, \nu, N)}(i) = \begin{cases} \binom{N+\mu}{N-i} \binom{N+\nu}{i} / \binom{2N+\mu+\nu}{N}, & \text{for } i \in \{0, 1, \dots, N\}, \\ 0 & \text{for } i \in \mathbb{Z} \setminus \{0, 1, \dots, N\}. \end{cases} \tag{9}$$

The associated cumulative distribution function (c.d.f.) and its tail are given by

$$\Omega^{(\mu, \nu, N)}(i) := \sum_{j \leq i} \omega^{(\mu, \nu, N)}(j) = 1 - \bar{\Omega}^{(\mu, \nu, N)}(i) \quad (i \in \mathbb{Z}). \tag{10}$$

Hahn polynomials satisfy, for  $0 \leq k, \ell \leq N$ , the orthogonality relations

$$\sum_{i=0}^N \binom{N+\mu}{N-i} \binom{N+\nu}{i} H_k^{(\mu, \nu, N)}(i) H_\ell^{(\mu, \nu, N)}(i) = \delta_{k, \ell} d_k^{(\mu, \nu, N)}, \tag{11}$$

$$d_k^{(\mu, \nu, N)} = \frac{(2N + \mu + \nu + 1 - k)!(N + \mu)!(N + \nu - k)!(N - k)!}{(2N + \mu + \nu + 1 - 2k)(N + \mu + \nu - k)!(N + \nu)!(N + \mu - k)!N!^2}. \tag{12}$$

The latter formula is most of the time given with parameters  $\alpha, \beta$ , so that we shall prove ours, for  $\mu, \nu$ , at the beginning of Section 5, with a reminder of some notations for factorials and binomial coefficients with non integer parameters.

Hahn polynomials also satisfy the discrete analogue of (6) in the form of the difference equations

$$\mathcal{L}^{(\mu, \nu, N)} H_k^{(\mu, \nu, N)} := \Delta[\sigma^{(\mu, \nu)} \omega^{(\mu, \nu, N)} \nabla H_k^{(\mu, \nu, N)}] = -\lambda_k^{(\mu, \nu, N)} \omega^{(\mu, \nu, N)} H_k^{(\mu, \nu, N)}, \tag{13}$$

for  $0 \leq k \leq N$ , with

$$\lambda_k^{(\mu, \nu, N)} = k(2N + \mu + \nu - k + 1), \quad \sigma^{(\mu, \nu)}(i) = i(i + \mu) \quad (0 \leq i, k \leq N) \tag{14}$$

(see [14, (2.1.18) p. 21]), and where for any function  $f : \mathbb{Z} \rightarrow \mathbb{R}$ , the forward and backward shift operators are defined by

$$\Delta f(i) = f(i + 1) - f(i), \quad \nabla f(i) = f(i) - f(i - 1) = \Delta f(i - 1) \quad (i \in \mathbb{Z}).$$

In order to avoid problems at the endpoints 0 and  $N$  when dealing with difference operators, any function  $f : \llbracket 0, N \rrbracket \rightarrow \mathbb{R}$  (including Hahn polynomials) will be identified with its extension to  $\mathbb{Z}$  obtained by setting  $f(n) = 0$  for  $n \notin \llbracket 0, N \rrbracket$ . Note for such a function  $f$  the fundamental property, valid for  $k = 0, \dots, N$ ,

$$\mathcal{L}^{(\mu, \nu, N)} f = -\lambda_k^{(\mu, \nu, N)} \omega^{(\mu, \nu, N)} f \iff \exists c \in \mathbb{R}, f = c H_k^{(\mu, \nu, N)}. \tag{15}$$

Let us introduce the scalar product

$$\langle f | g \rangle_\omega = \sum_{i=0}^N f(i) g(i) \omega^{(\mu, \nu, N)}(i). \tag{16}$$

In this setting, noticing that  $d_0^{(\mu, \nu, N)} = \binom{2N + \mu + \nu}{N}$ , the orthogonality relations (11) take the form

$$\langle H_k^{(\mu, \nu, N)} | H_\ell^{(\mu, \nu, N)} \rangle_\omega = \frac{d_k^{(\mu, \nu, N)}}{d_0^{(\mu, \nu, N)}} \delta_{k, \ell} \quad (0 \leq k, \ell \leq N). \tag{17}$$

Finally, by analogy with (7), consider the kernel

$$h^{(\mu, \nu, N)}(i, j) := \sum_{\ell=1}^i \frac{\Omega^{(\mu, \nu, N)}(\ell-1)}{\sigma^{(\nu)}(\ell)\omega^{(\mu, \nu, N)}(\ell)} + \sum_{\ell=j+1}^N \frac{\bar{\Omega}^{(\mu, \nu, N)}(\ell-1)}{\sigma^{(\nu)}(\ell)\omega^{(\mu, \nu, N)}(\ell)} - \sum_{\ell=1}^N \frac{\Omega^{(\mu, \nu, N)}(\ell-1)\bar{\Omega}^{(\mu, \nu, N)}(\ell-1)}{\sigma^{(\nu)}(\ell)\omega^{(\mu, \nu, N)}(\ell)}, \tag{18}$$

The discrete analogue of (7) is given by

**Theorem 1.** *The K-L expansion*

$$h^{(\mu, \nu, N)}(i, j) = \sum_{k=1}^N \frac{1}{\lambda_k^{(\mu, \nu, N)}} \cdot \frac{d_0^{(\mu, \nu, N)}}{d_k^{(\mu, \nu, N)}} H_k^{(\mu, \nu, N)}(i) H_k^{(\mu, \nu, N)}(j) \tag{19}$$

holds, for  $0 \leq i, j \leq N$ , with respect to the p.m.f.  $\omega^{(\mu, \nu, N)}$ .

**Proof.** From Lemma 7, Lemma 8, and (15) with  $k = 0$ , we infer that given  $j$ , the two functions of the variable  $i$  appearing on the two sides of the equal sign in (19) are equal up to an additive constant. The latter is zero in view of Lemma 5, and the equalities

$$\langle H_k^{(\mu, \nu, N)} | H_0^{(\mu, \nu, N)} \rangle_\omega = 0 \quad (1 \leq k \leq N). \quad \square$$

#### 4. A Sobolev test for uniformity on Johnson's graph

Let us introduce some notations and basic facts about Johnson's graphs. These graphs provide an example of distance-transitive graphs, discrete analogues of the continuous compact two-point homogeneous (also called rank-one symmetric) spaces. In [4, see §4.1.F p. 136-138 for the definition of distance transitive graphs and §9.1 p. 255-261 about Johnson graphs]; we will also use [19, II.3]. In Section 5, our Lemma 9 will restate and give some elements of proof for the spectrum features (22).

Assume  $S$  is a finite set with cardinality  $m \geq 2N$ . The graph of Johnson  $\Gamma = J(m, N)$  has vertex set, denoted by  $\mathcal{V} = \mathcal{V}(m, N)$ , the collection of subsets of  $S$  whose cardinality is  $N$ . The volume, or size of the graph, is therefore  $|\mathcal{V}(m, N)| = \binom{m}{N}$ . Two vertices  $v_1, v_2$  are adjacent if  $v_1 \cap v_2$  has cardinality  $N - 1$ . The distance between two vertices  $v_1, v_2$  is  $d(v_1, v_2) = m - |v_1 \cap v_2|$  and can take any value in  $\llbracket 0, N \rrbracket$ ; consequently  $J(m, N)$  has diameter  $N$ . The volume of any sphere with radius  $i$  is

$$|S^{m, N}(i)| = \binom{N}{N-i} \binom{m-N}{i} = |\mathcal{V}(m, N)| \omega^{(0, m-2N, N)}(i) \quad (0 \leq i \leq N). \tag{20}$$

Johnson's graph is  $r$ -regular with  $r = |S^{m, N}(1)| = N(m - N)$ . If  $\mathbf{A}$  denotes the adjacency matrix of an  $r$ -regular graph, then  $\mathbf{A}$  has a decreasing sequence of, say  $s$ , distinct eigenvalues denoted by  $(\lambda'_k)_{0 \leq k \leq s-1}$ ,  $\lambda'_k$  having multiplicity  $m_k$ . The sequence  $(\lambda'_k, m_k)_{0 \leq k \leq s-1}$  is the *ordinary spectrum* of  $\Gamma$  (see [2, Definition 2.2], [5, p. 23]).

The combinatorial Laplacian matrix (also called Kirchoff matrix in [3, p. 54], admittance matrix in [5, p. 27]) of  $\Gamma$  is  $r\mathbf{I} - \mathbf{A}$ , ( $\mathbf{I}$  being the identity matrix of size the same as  $\mathbf{A}$ ) with eigenvalues given by the increasing sequence

$$\lambda_k = r - \lambda'_k \quad (0 \leq k \leq s - 1) \tag{21}$$

and the same multiplicities (see [3, p. 268], [5, (1.33') p. 30]). The sequence  $(\lambda_k, m_k)_{0 \leq k \leq s-1}$  is the *Laplacian spectrum* of the graph. For a connected graph,  $\lambda_0 = 0$  has multiplicity one, the corresponding eigenspace  $E_0$  consisting of all constant functions. All other eigenvalues are therefore positive.

Johnson graph  $J(m, N)$  has  $s = N + 1$  distinct eigenvalues and the Laplacian spectrum is associated with the eigenvalues of Hahn polynomials via relations

$$(\lambda_k, m_k)_{0 \leq k \leq N} = (\lambda_k^{(0, m-2N, N)}, \dim E_k^{(m, N)})_{0 \leq k \leq N} \tag{22}$$

$$= \left( k(m-k+1), \binom{m}{k} - \binom{m}{k-1} \right)_{0 \leq k \leq N} \tag{23}$$

where  $E_k^{(m, N)}$  denotes the eigenspace associated with  $\lambda_k = \lambda_k^{(0, m-2N, N)}$ .

Let us now, using the framework of [8] in the discrete case, introduce the following notations. Let  $V_1, \dots, V_n$  be a sample of vertices drawn from a distribution over  $\mathcal{V}$ . The uniform distribution assigns the measure  $|\mathcal{V}|^{-1}$  to every vertex  $v \in \mathcal{V}$ .

Assume that for  $0 \leq k \leq s-1$ ,  $\{g_{k, \ell} : 1 \leq \ell \leq \dim E_k^{(m, N)}\}$  is an orthonormal basis of  $E_k^{(m, N)}$ , thus providing an orthonormal basis of  $L_2(J(m, N))$  with respect to the scalar product

$$\langle f | g \rangle = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} f(v)g(v).$$

Given the positive real numbers  $\alpha_1, \dots, \alpha_{s-1}$ , a consistent, invariant Sobolev test for uniformity, is provided by the degenerate  $V$ -statistic defined by

$$T_n(\{\alpha_k\}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n h(V_i, V_j), \quad h(v_1, v_2) = \sum_{k=1}^{s-1} \alpha_k^2 \left\{ \sum_{\ell=1}^{\dim E_k^{(m, N)}} g_{k, \ell}(v_1)g_{k, \ell}(v_2) \right\}$$

(see Theorem 4.4, formula (2.7), (5.5) in [8]), the null hypothesis of uniformity being rejected for large values of  $T_n(\{\alpha_k\})$ .

Recall that the function  $\tilde{f} : \mathcal{V} \rightarrow \mathbb{R}$  is called *zonal spherical* (with respect to  $v \in \mathcal{V}$ ) if  $\tilde{f}$  is an eigenfunction of the Laplacian and can be written as  $\tilde{f}(w) = f(d[v, w])$  for some one-variable function  $f : [0, N] \rightarrow \mathbb{R}$ . The function  $f$  is also referred to as a zonal spherical function, giving rise to  $\tilde{f}$  after the choice of any arbitrary  $v \in \mathcal{V}$ . Following [8, (5.1)], consider the functions defined, for  $0 \leq k \leq N$ , by

$$f_k^{(m, N)}(i) = (\dim E_k^{(m, N)})^{1/2} H_k^{(0, m-2N, N)}(i) \quad (0 \leq i \leq N). \tag{24}$$

**Proposition 2.** Assume that for  $0 \leq k \leq s-1$ ,  $\{g_{k, \ell} : 1 \leq \ell \leq \dim E_k^{(m, N)}\}$  is an orthonormal base of  $E_k^{(m, N)}$ . Hahn polynomials satisfy

$$\frac{1}{|\mathcal{V}(m, N)|} \sum_{i=0}^N H_k^{(0, m-2N, N)}(i) H_\ell^{(0, m-2N, N)}(i) |S^{(m, N)}(i)| = \frac{\delta_{k, \ell}}{\dim E_k^{(m, N)}}, \tag{25}$$

and for functions defined by (24) satisfy, for  $0 \leq k, \ell \leq N$ , Giné's addition formula

$$\sum_{\ell=1}^{\dim E_k^{(m, N)}} g_{k, \ell}(v_1)g_{k, \ell}(v_2) = (\dim E_k^{(m, N)})^{1/2} f_k^{(m, N)}(i) \tag{26}$$

holds, whenever  $d(v_1, v_2) = i$ , for  $v_1, v_2$  vertices of  $J(m, N)$ .

**Proof.** In view of (17), (20) and relation (51) from Lemma 9 of Section 5, (25) is a restatement of (11). Equality in (26) is the discrete version of Giné's addition formula ([8, Theorem 3.2]) for the following reason. From (4) in [10, §13.1.5, p. 492], we can assert, due to (25), that  $H_k^{(0, m-2N, N)}(i)$  is the  $k$ -th zonal spherical function of our graph, normalized by  $H_k^{(0, m-2N, N)}(0) = 1$ . Then we notice that (a) in Theorem 3.2 of [8] agrees with our normalization  $f_k^{(m, N)}(0) = (\dim E_k^{(m, N)})^{1/2}$ , in view of definition (24). □

As a corollary of this proposition we obtain

**Theorem 3.** Assume  $\alpha_1, \dots, \alpha_N \in \mathbb{R}$  and let  $v_1, v_2$  be two vertices of the graph  $J(m, N)$ . The equality

$$\sum_{k=1}^N \left\{ \alpha_k^2 \cdot \sum_{\ell=1}^{\dim E_k^{(m,N)}} g_{k,\ell}(v_1) g_{k,\ell}(v_2) \right\} = \sum_{k=1}^N \alpha_k^2 \cdot \frac{d_0^{(0,m-2N,N)} H_k^{(0,m-2N,N)}(d[v_1, v_2])}{d_k^{(0,m-2N,N)}}$$

holds. In particular

$$H^{(m,N)}(v_1, v_2) := \sum_{k=1}^N \left\{ \frac{1}{k(m-k+1)} \cdot \sum_{\ell=1}^{\dim E_k^{(m,N)}} g_{k,\ell}(v_1) g_{k,\ell}(v_2) \right\} \tag{27}$$

$$= h^{(0,m-2N,N)}(0, d[v_1, v_2]) \tag{28}$$

**Proof.** The first equality is a direct consequence of (26) and (51). Then the particular case  $\alpha_k^2 = 1/\lambda_k^{(0,m-2N,N)}$  for  $1 \leq k \leq N$ , in view of (19), yields (27).  $\square$

A kernel associated with  $H^{(m,N)}$  for statistical purposes in Giné’s framework (see [8, (5.3)’–(5.4)’]) is

$$G^{(m,N)}(v_1, v_2) := \sum_{k=1}^N \frac{1}{\sqrt{k(m-k+1)}} \left\{ \sum_{\ell=1}^{\dim E_k^{(m,N)}} g_{k,\ell}(v_1) g_{k,\ell}(v_2) \right\}. \tag{29}$$

We will denote by  $\|G^{(m,N)}\|_\infty$  the number  $\max_{v_2} G^{(m,N)}(v_1, v_2)$ , which is independent upon  $v_1$ , since our graph is homogeneous.

If  $d$  is a positive integer let  $\chi(d)$  denote a chi-square random variable with  $d$  degrees of freedom. The following result combines standard properties of degenerate  $V$ –statistics with [8, Theorem 5.3].

**Theorem 4.** Let  $J(m, N)$  be a Johnson graph, with notations (22) for the Laplacian spectrum and definitions (18), (27). Let  $V_1, V_2, \dots$  be a sequence of vertices from the uniform distribution over  $J(m, N)$ . Giné’s statistic

$$T_n \left( \left\{ \sqrt{\lambda_k^{(0,m-2N,N)}} \right\} \right) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n H^{(m,N)}(V_i, V_j) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n h^{(0,m-2N,N)}(0, d[V_i, V_j]) \tag{30}$$

converges in law toward the random variable

$$\sum_{k=1}^N \frac{\chi_k(\dim E_k^{(m,N)})}{\lambda_k^{(0,m-2N,N)}}$$

where the random variables  $\chi_k(\dim E_k^{(m,N)})$ ,  $1 \leq k \leq N$  are independent.

Furthermore, the test based on the rejection of uniformity for large values of  $T_n$  is most powerful invariant except for terms of order  $O(\alpha^3)$ , against the alternative p.m.f.

$$w \in \mathcal{V} \mapsto \frac{1}{|\mathcal{V}(m, N)|} \left( 1 + \alpha \frac{G^{(m,N)}(v, w)}{\|G^{(m,N)}\|_\infty} \right) \quad (0 < |\alpha| < 1)$$

where  $v$  is any fixed vertex of  $J(m, N)$ .

The lack of data or models from real-life, associated with Johnson graphs, deprives us from any privileged alternative distribution, under which the power of our new test for uniformity should be discussed through simulations. Therefore numerical simulations involving some specific alternatives will only be the object of some forthcoming papers.

To our knowledge no specific probabilistic or statistical models have been discussed in the literature, so that we just outline directions for further research. Giné’s alternative (29) arises



from Fourier analysis and might not be of particular interest. Other alternative distributions are of the form

$$w \in \mathcal{V} \mapsto \frac{1}{|\mathcal{V}(m, N)|} (1 + \alpha H_1^{(0, m-2N)}(d(v, w))) \tag{31}$$

with  $|\alpha|$  small enough and where  $v$  is any fixed vertex of  $J(m, N)$ . They are perturbations of the uniform distribution by a zonal spherical function associated with the first non-constant Hahn polynomial, in other words the analogues, for  $J(m, N)$ , of the cardioid distributions for the circle (see [13, §3.5.5]).

Other alternatives can be associated with the subgroups of the full isometry group of  $J(m, N)$  (given, e.g., in [4, Theorem 9.1.2]). The uniform distribution being characterised by its invariance with respect to the full group, some proper subgroup may determine a wider family of distributions, in the same way as, for instance, on the Euclidean sphere, rotations about a given axis determine a family of rotationally invariant distributions.

All these alternatives arise from the so-called transformation models, see [13, 3.5.1], and references therein about this topic.

### 5. Useful technical results

Throughout this section we will sometimes drop the superscripts and write, e.g.  $\lambda_k$  instead of  $\lambda_k^{(\mu, \nu, N)}$ .

Following [9, §1.1] and [11, §1.3], we use the generalized factorial and binomial coefficients defined by

$$r! = \Gamma(r + 1), \binom{r}{k} = (-1)^k \frac{(-r)_k}{k!} \quad (r > -1, k \in \mathbb{N}),$$

where  $\Gamma$  is the gamma function, and set

$$\binom{-r}{k} = (-1)^k \binom{r + k - 1}{k}, \quad (r \geq 1, k \in \mathbb{N}).$$

Note the identity

$$(-r)_k = (-1)^k \frac{r!}{(r - k)!} \quad (r \geq k)$$

The traditional orthogonality relation for Hahn polynomials given by [11, §9.5], reads

$$\sum_{i=0}^N \binom{\alpha + i}{i} \binom{\beta + N - i}{N - i} Q_k(i; \alpha, \beta, N) Q_\ell(i; \alpha, \beta, N) = \frac{(-1)^k (k + \alpha + \beta + 1)_{N+1} (\beta + 1)_k k!}{(2k + \alpha + \beta + 1)(\alpha + 1)_k (-N)_k} =: d'_k \tag{32}$$

With parameters  $\nu = -\alpha - N - 1, \mu = -\beta - N - 1$ , the above identities give

$$\binom{\alpha + k}{k} \binom{\beta + N - k}{N - k} = \binom{-N - \nu - 1 + k}{k} \binom{-N - \mu - 1 + N - k}{N - k} = (-1)^k \binom{N + \nu}{k} (-1)^{N-k} \binom{\mu + N}{N - k} \tag{33}$$

On the other hand,

$$(k + \alpha + \beta + 1)_{N+1} = (k - 2N - \mu - \nu - 1)_{N+1} = (-1)^{N+1} \frac{(2N + \mu + \nu + 1 - k)!}{(N + \mu + \nu - k)!}, \tag{34}$$

$$(\beta + 1)_k = (-N - \mu)_k = (-1)^k \frac{(N + \mu)!}{(N + \mu - k)!}, \quad (\alpha + 1)_k = (-N - \nu)_k = (-1)^k \frac{(N + \nu)!}{(N + \nu - k)!}, \tag{35}$$

$$(-N)_k = (-1)^k \frac{N!}{(N - k)!}, \quad (2k + \alpha + \beta + 1) = -(2N + \mu + \nu + 1 - 2k), \tag{36}$$

so that

$$d'_k = \frac{(-1)^N (2N + \mu + \nu + 1 - k)! (N + \mu)! (N + \nu - k)! (N - k)! k!}{(2N + \mu + \nu + 1 - 2k) (N + \mu + \nu - k)! (N + \nu)! (N + \mu - k)! N! 2^2},$$

and (11) readily follows.

Consider the three-variable kernel, defined over  $\llbracket 0, N \rrbracket^3$  by

$$\tilde{h}^{(\mu, \nu, N)}(i, j, z) = \frac{\mathbf{1}_{\{i \leq z-1\}} - \Omega^{(\mu, \nu, N)}(z-1)}{\sigma^{(\nu)}(z)\omega^{(\mu, \nu, N)}(z)^2} \cdot \frac{\mathbf{1}_{\{j \leq z-1\}} - \Omega^{(\mu, \nu, N)}(z-1)}{\sigma^{(\nu)}(z)\omega^{(\mu, \nu, N)}(z)^2} \quad (z = 1, \dots, N), \quad (37)$$

$$\tilde{h}^{(\mu, \nu, N)}(i, j, 0) = 0, \quad (38)$$

and  $\tilde{h}^{(\mu, \nu, N)}(i, j, z) = 0$  for  $i, j$  or  $z$  belonging to  $\mathbb{Z} \setminus \llbracket 0, N \rrbracket$ . Throughout this subsection,  $I$  and  $Z$  will denote two independent random variables with p.m.f.  $\omega^{(\mu, \nu, N)}$ .

Elementary computations give, for  $0 \leq i \leq j \leq N$ ,

$$\begin{aligned} E_Z[\tilde{h}(i, j, Z)] &= \sum_{\ell=1}^i \frac{\Omega(\ell-1)^2}{\sigma(\ell)\omega(\ell)} - \sum_{\ell=i+1}^j \frac{\bar{\Omega}(\ell-1)\Omega(\ell-1)}{\sigma(\ell)\omega(\ell)} + \sum_{\ell=j+1}^N \frac{\bar{\Omega}(\ell-1)^2}{\sigma(\ell)\omega(\ell)} \\ &= \sum_{\ell=1}^i \frac{\Omega(\ell-1)}{\sigma(\ell)\omega(\ell)} + \sum_{\ell=j+1}^N \frac{\bar{\Omega}(\ell-1)}{\sigma(\ell)\omega(\ell)} - \sum_{\ell=1}^N \frac{\Omega(\ell-1)\bar{\Omega}(\ell-1)}{\sigma(\ell)\omega(\ell)} = h(i, j) \end{aligned} \quad (39)$$

(for the second equality use relations  $\Omega^2 + \bar{\Omega}\Omega = \Omega$  and  $\bar{\Omega}^2 + \bar{\Omega}\Omega = \bar{\Omega}$ ). Then from  $E_I[\mathbf{1}_{\{I \leq z-1\}} - \Omega(z-1)] = 0$ , we infer relations  $E_I[\tilde{h}(I, j, z)] = 0$  for every  $j, z \in \mathbb{Z}$ , and  $E_Z\{E_I[\tilde{h}(I, j, Z)]\} = 0$ . Then by the discrete Fubini theorem we obtain  $E_I\{E_Z[\tilde{h}(I, j, Z)]\} = 0$ , which means  $E_I[h(I, j)] = 0$ . We have proved

**Lemma 5.** *The kernel  $(i, j) \mapsto h^{(\mu, \nu, N)}(i, j)$  is degenerate with respect to  $\omega^{(\mu, \nu, N)}$ . In other words*

$$\forall j \in \{0, 1, \dots, N\} : \sum_{i=0}^N h^{(\mu, \nu, N)}(i, j)\omega^{(\mu, \nu, N)}(i) = 0.$$

In view of (17), an expansion of Dirac distribution is given by the dual relations

$$\omega^{(\mu, \nu, N)}(\cdot) \sum_{k=0}^N \frac{d_0^{(\mu, \nu, N)} H_k^{(\mu, \nu, N)}(\cdot) H_k^{(\mu, \nu, N)}(j)}{d_k^{(\mu, \nu, N)}} = \delta_j(\cdot) \quad (0 \leq j \leq N) \quad (40)$$

The fact that  $H_0^{(\mu, \nu, N)} = 1$  then implies

**Lemma 6.** *Assume  $0 \leq j \leq N$ . Then*

$$\omega^{(\mu, \nu, N)}(\cdot) \sum_{k=1}^N \frac{d_0^{(\mu, \nu, N)} H_k^{(\mu, \nu, N)}(\cdot) H_k^{(\mu, \nu, N)}(j)}{d_k^{(\mu, \nu, N)}} = \delta_j(\cdot) - \omega^{(\mu, \nu, N)}(\cdot). \quad (41)$$

By linearity and in view of (13), we obtain as a corollary

**Lemma 7.** *Assume  $0 \leq j \leq N$ . Then*

$$\mathcal{L}^{(\mu, \nu, N)} \left\{ \sum_{k=1}^N \frac{d_0^{(\mu, \nu, N)} H_k^{(\mu, \nu, N)}(\cdot) H_k^{(\mu, \nu, N)}(j)}{\lambda_k^{(\mu, \nu, N)} d_k^{(\mu, \nu, N)}} \right\} = \omega^{(\mu, \nu, N)}(\cdot) - \delta_j(\cdot) \quad (42)$$

**Lemma 8.** *Assume  $0 \leq j \leq N$ . Then*

$$\mathcal{L}^{(\mu, \nu, N)}[h^{(\mu, \nu, N)}(\cdot, j)] = \omega^{(\mu, \nu, N)}(\cdot) - \delta_j(\cdot). \quad (43)$$

**Proof.** The difference operators acting on the variable  $i$ , we obtain successively

$$\nabla\{\mathbf{1}_{\{i \leq z-1\}} - \Omega(z-1)\} = \mathbf{1}_{\{i \leq z-1\}} - \mathbf{1}_{\{i-1 \leq z-1\}} = -\mathbf{1}_{\{i=z\}} \quad (i, z \in \mathbb{Z}), \quad (44)$$

$$\nabla\tilde{h}(i, j, z) = -\mathbf{1}_{\{i=z\}} \cdot \frac{\{\mathbf{1}_{\{j \leq z-1\}} - \Omega(z-1)\}}{\sigma(z)\omega(z)^2} \quad (i, z \in \llbracket 1, N \rrbracket), \quad (45)$$

$$\nabla h(i, j) = \nabla\{E_Z\tilde{h}(i, j, Z)\} = E_Z\{\nabla\tilde{h}(i, j, Z)\} \quad (46)$$

$$= -\omega(i) \cdot \frac{\{\mathbf{1}_{\{j \leq i-1\}} - \Omega(i-1)\}}{\sigma(i)\omega(i)^2} \quad (i \in \llbracket 1, N \rrbracket), \quad (47)$$

$$\sigma(i)\omega(i)\nabla h(i, j) = \Omega(i-1) - \mathbf{1}_{\{j \leq i-1\}} \quad (i \in \llbracket 1, N \rrbracket). \quad (48)$$

The latter is also true for  $i \leq 0$  or  $i > N$  (both sides being equal to zero), so that we may write, for  $i \in \mathbb{Z}$ ,

$$\Delta[\sigma(i)\omega(i)\nabla h(i, j)] = \Delta[\Omega(i-1) - 1_{\{j \leq i-1\}}] = \omega(i) - \delta_j(i), \quad (49)$$

and the result is proved.  $\square$

**Lemma 9.** *The graph  $J(m, N)$  has eigenvalues*

$$\lambda_k^{(0, m-2N, N)} = k(m-k+1) \quad (0 \leq k \leq N) \quad (50)$$

*with multiplicities*

$$\dim E_k^{(m, N)} = \binom{m}{k} - \binom{m}{k-1} = \frac{m!(m-2k+1)}{k!(m-k+1)!} = \frac{d_0^{(0, m-2N)}}{d_k^{(0, m-2N)}}. \quad (51)$$

**Proof.** Let us use Theorem 9.1.2 (with  $n = m, e = N$ ) in [4, p. 255], or [19, §II.3 p. 431] (with  $v = m$  and  $n = N$ ). The eigenvalues of the adjacency matrix of the graph are  $\lambda'_k = (N-k)(m-N-k) - k$ ,  $0 \leq k \leq N$ , and (50) follows from (21) with  $r = N(m-N)$  and (14). For the value of  $\dim E_k^{(m, N)}$ , see the last equality p. 431 in [19] and the last equality in [4, Theorem 9.1.2]. The second equality in (51) is simple calculus; for the third, use (11) yielding,

$$d_k^{(0, m-2N, N)} = \frac{(m-k+1)!k!}{(m-2k+1)(m-N)!N!}, \quad d_0^{(0, m-2N, N)} = \frac{m!}{(m-N)!N!}. \quad (52)$$

$\square$

## References

- [1] R. J. Beran, "Testing for uniformity on a compact homogeneous space", *J. Appl. Probab.* **5** (1968), p. 177-195.
- [2] N. Biggs, *Algebraic graph theory*, Cambridge Mathematical Library, Cambridge University Press, 1994.
- [3] B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer, 1998.
- [4] A. E. Brouwer, A. M. Cohen, A. Neumaier, *Distance-Regular Graphs*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge, vol. 18, Springer, 1989.
- [5] D. M. Cvetković, M. Doob, H. Sachs, *Spectra of Graphs: Theory and Applications*, 3rd rev. enl. ed., J. A. Barth Verlag, 1998.
- [6] R. B. D'Agostino, M. A. Stephens (eds.), *Goodness-of-Fit Techniques*, Statistics: Textbooks and Monographs, vol. 68, Marcel Dekker, 1986.
- [7] J. Durbin, *Distribution theory for tests based on the sample distribution function*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 9, Society for Industrial and Applied Mathematics, 1973.
- [8] E. Giné, "Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms", *Ann. Stat.* **3** (1975), no. 6, p. 1243-1266.
- [9] N. L. Johnson, S. Kotz, N. Balakrishnan, *Continuous univariate distributions. Volume 2*, 2nd ed., John Wiley & Sons, 1995.
- [10] A. U. Klimyk, N. Y. Vilenkin, *Representations of Lie groups and special functions. Volume II*, Kluwer Academic Publishers, 1995.
- [11] R. Koekoek, P. A. Lesky, R. F. Swarttouw, *Hypergeometric orthogonal polynomials and their q-analogues*, Springer Monographs in Mathematics, Springer, 2010.
- [12] V. S. Korolyuk, Y. V. Borovskikh, *Theory of U-statistics*, Springer, 2013.
- [13] K. V. Mardia, P. E. Jupp, *Directional statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, 2000.
- [14] A. F. Nikiforov, S. K. Suslov, V. B. Uvarov, *Classical orthogonal polynomials of a discrete variable*, Springer Series in Computational Physics, Springer, 1991.
- [15] A. Pewsey, E. García-Portugués, "Recent advances in directional statistics", *Test* **30** (2021), no. 1, p. 1-58.
- [16] J.-R. Pycke, "U-statistics based on the Green's function of the Laplacian on the circle and the sphere", *Stat. Probab. Lett.* **77** (2007), no. 9, p. 863-872.
- [17] R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, 2009.
- [18] G. R. Shorack, J. A. Wellner, *Empirical processes with applications to statistics*, Classics in Applied Mathematics, vol. 59, Society for Industrial and Applied Mathematics, 2009.

- [19] D. Stanton, "An introduction to group representations and orthogonal polynomials", in *Orthogonal polynomials: theory and practice*, NATO ASI Series. Series C. Mathematical and Physical Sciences, vol. 294, Kluwer Academic Publishers, 1990, p. 419-433.
- [20] A. W. van der Vaart, *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, 2000.
- [21] G. S. Watson, "Goodness-of-fit tests on a circle", *Biometrika* **48** (1961), no. 1/2, p. 109-114.