COMPUTATIONAL ASPECTS OF CUI-FREEDEN STATISTICS FOR EQUIDISTRIBUTION ON THE SPHERE

CHRISTINE CHOIRAT AND RAFFAELLO SERI

ABSTRACT. In this paper, we derive the asymptotic statistical properties of a class of generalized discrepancies introduced by Cui and Freeden (SIAM J. Sci. Comput., 1997) to test equidistribution on the sphere. We show that they have highly desirable properties and encompass several statistics already proposed in the literature. In particular, it turns out that the limiting distribution is an (infinite) weighted sum of chi-squared random variables. Issues concerning the approximation of this distribution are considered in detail and explicit bounds for the approximation error are given. The statistics are then applied to assess the equidistribution of Hammersley low discrepancy sequences on the sphere and the uniformity of a dataset concerning magnetic orientations.

1. INTRODUCTION

The aim of this paper is to derive the properties of a new class of statistics for testing equidistribution on the sphere and to investigate their computational aspects. They have been introduced in a seminal paper by Cui and Freeden (see [10]) under the name of *generalized discrepancies*[1] in order to extend the famous Koksma-Hlawka inequality of Numerical Analysis from the unit hypercube to the sphere, and to evaluate the uniformity properties of some deterministic sequences on the sphere. Some further properties of these discrepancies with direct interest for Numerical Analysis are developed in [43]. These discrepancies can be generalized to hyperspheres of arbitrary dimension: for the numerical properties in this case, the reader may consult [9].

These discrepancies are indicated in the following as $D(\mathcal{P}_N; \mathbf{A})$ where $\mathcal{P}_N$ is a sample of points and $\mathbf{A}$ is a pseudodifferential operator acting on functions defined on the sphere $\Omega \subset \mathbb{R}^3$, that enters D only through its *symbol*, that is a sequence of constants $\{A_n; n = 0, 1, \dots\}$. In the following we will need no property of $\mathbf{A}$ apart from its symbol. We address the reader interested in the mathematical details underlying the formulas to the original paper [10], in which the authors do not however derive the statistical properties of this class of discrepancies. This is the object of the present paper. A short exposition of the mathematical background necessary to understand the statistical results is given in Section 2. The test statistics encompass those of Beran ([4]), Giné ([21]), Prentice ([33], see also [34]),

[1] The same name was used by Hickernell ([25]) to denote a different class of discrepancies on the unit hypercube, whose statistical properties were studied in [8].

and Pycke ([36, 37]). It is important to remark that some of these test statistics have particular optimality properties (e.g., against certain hypotheses) that are not shared by the general version $D(\mathcal{P}_N; \mathbf{A})$. The definition of $D(\mathcal{P}_N; \mathbf{A})$ is expressed in terms of a (generally infinite) sum and its computation may be unstable when no closed formula exists. For this reason, in Section 3 we review a set of cases in which closed formulas are available. The main statistical results are stated in Section 4: in particular, the asymptotic distribution of generalized discrepancies is shown to be an infinite weighted sum of chi-square random variables. The computation of this distribution passes customarily through the truncation of the infinite sum and its approximation through a finite weighted sum. Several methods are available to compute finite weighted sums of chi-squares (see [27, 14, 41, 15, 38], just to name a few), but the error caused by the truncation of the infinite sequence of weights is not sufficiently investigated in the literature. This is particularly relevant in this case, since the number of degrees of freedom of the chi-squares in the sum increases with the summation index. Therefore we provide some easily computable and effective bounds for the truncation error in Section 5. In Section 6, we consider a particular version of the statistic proposed in [10]; we evaluate through simulations the finite-sample performance of the statistic and the error in the approximation of the asymptotic distribution, and we use it to evaluate the uniformity of a deterministic sequence; at last, we compare several statistics on a dataset on Ω. Proofs are gathered in Section 7.

2. Mathematical preliminaries

In order to introduce the formula of $D(\mathcal{P}_n; \mathbf{A})$, we need some concepts from Functional Analysis on the sphere. Let $L^2(\Omega)$ be the space of Lebesgue square integrable scalar functions on Ω. The spherical harmonics

$$\{Y_{n,j}; n = 0, 1, \ldots; j = -n, -n+1, \ldots, n-1, n\}$$

are a class of double indexed functions defined on Ω, where n is called the *degree* and j is called the *order* of the spherical harmonics. Spherical harmonics constitute an orthonormal basis of $L^2(\Omega)$, when Ω is endowed with the Lebesgue measure.[2] The Legendre polynomials $\{P_n; n = 0, 1, \ldots\}$ are the infinitely differentiable eigenfunctions of the Legendre operator satisfying the normalization condition $|P_n(1)| = 1$. The addition theorem of spherical harmonics links these functions to the Legendre polynomials through the relation:

$$\sum_{j=-n}^{n} Y_{n,j}(\xi) \cdot Y_{n,j}(\eta) = \frac{2n+1}{4\pi} \cdot P_n(\xi \cdot \eta), \qquad (\xi, \eta) \in \Omega^2,$$

where $\cdot$ denotes the inner product on the sphere. In particular, taking $\xi = \eta$, we have that $\sum_{j=-n}^{n} |Y_{n,j}(\xi)|^2 = \frac{2n+1}{4\pi}$.

The only property of a pseudodifferential operator $\mathbf{A}$ that we will need is the *spherical symbol*, that is a sequence $\{A_n; n = 0, 1, \ldots\}$ of real numbers characterizing the behavior of the operator. If $\lim_{n\to\infty} \frac{|A_n|}{(n+\frac{1}{2})^t} = \text{const} \neq 0$ for some $t \in \mathbb{R}$, then $\mathbf{A}$ is a pseudodifferential operator of order t. If, on the other hand, $\lim_{n\to\infty} \frac{|A_n|}{(n+\frac{1}{2})^t} = 0$ for any $t \geq 0$, $\mathbf{A}$ is of order $-\infty$. Since we are not interested in the way in which $\mathbf{A}$ acts on functions defined on Ω, but only in $\{A_n; n = 0, 1, \ldots\}$,

---

[2]Note that here we use the real-valued spherical harmonics and not the complex ones.

for us the fact that $\mathbf{A}$ is of order t reduces to the rate of decrease of A_n to 0 (however, see [10, p. 597], for more information).

Let $\mathcal{P}_N = \{\xi_1, \xi_2, \ldots, \xi_N\}$ be a set of points on Ω. The *generalized discrepancy* associated with a pseudodifferential operator $\mathbf{A}$ with symbol $\{A_n\}$, $A_n \neq 0$ for $n \geq 1$, is defined by

$$(2.1) \qquad D\left(\mathcal{P}_N; \mathbf{A}\right) = \frac{1}{N} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(\xi_i \cdot \xi_j\right) \right]^{1/2}.$$

A difference with respect to the case considered in [10] is the fact that these authors suppose that A_n is finite for every $n \geq 1$, while in statistical applications there is some interest in allowing for $A_n^{-1} = 0$ for some n, as we will show in the examples below.

In Section 3 we consider the computation of the test statistic. In Section 4, we will consider the asymptotic properties of $D^2\left(\mathcal{P}_N; \mathbf{A}\right)$. The computation of the asymptotic distribution is addressed in Section 5, while Section 6 shows simulations providing support for the previous results.

3. Computation of the test statistics

The computation of $D^2\left(\mathcal{P}_N; \mathbf{A}\right)$ for a specific point set $\mathcal{P}_N$ is much more difficult than it seems. Indeed, the computation of the Legendre polynomials appearing in (2.1) can be overly complicated if a nonstable numerical formula for P_n is used (see the introductory treatment in [35, Section 5.4]). This fact is noted also in [10, p. 602]. Therefore, it is always better to have a closed form expression for $\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(\xi \cdot \eta\right)$.

Some tests used in Statistics are characterized by explicit forms. Closed formulas trivially exist for the Rayleigh statistic $\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} P_1\left(\xi_i \cdot \xi_j\right)$ obtained taking $A_1^2 = \frac{3}{4\pi}$ and $A_n^{-1} = 0$ for $n \geq 2$ (see [21, p. 1263]) and for the Bingham statistic $\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} P_2\left(\xi_i \cdot \xi_j\right)$ obtained taking $A_2^2 = \frac{5}{4\pi}$ and $A_n^{-1} = 0$ for $n \neq 2$ (see [21, p. 1264]). Other, more complicated cases are:

- Beran's form of Ajne's statistic (see [33, p. 172]; note that [21, p. 1262] uses a different normalization):

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(t\right) = \frac{1}{4} - (2\pi)^{-1} \cos^{-1} t,$$

  obtained taking $A_n^2 = +\infty$ for n even and $A_n^2 = n^2 \cdot \left(\frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)} \right)^2$ for n odd (see, e.g., equation 8.925.1 in [22]),

- Giné's statistic (see [21, p. 1262] and, in particular, Proposition 6.4):

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(t\right) = \frac{1}{2} - 2\pi^{-1} \sin \cos^{-1} t,$$

  obtained taking $A_n^2 = +\infty$ for n odd and $A_n^2 = \frac{n-1}{n+2} \cdot \left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \right)^2$ for n even (see, e.g., equation 8.925.2 in [22]),

- and Pycke's statistic (see [36, 37]):

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(t) = -\frac{1}{4\pi} \cdot \ln \frac{e}{2}(1-t),$$

obtained taking $A_n = \sqrt{n(n+1)}$ (see, e.g., equations 8.926.1 and 2 in [22]; however, note that this statistic does not respect the condition $\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2} < \infty$, and has to be analyzed as in Theorem 4.1 (iv)).

In Numerical Analysis, the most prominent case arises when

$$A_n = \sqrt{(2n+1)\, n\, (n+1)},$$

yielding (see equations 8.926.1 and 2 in [22]):

$$D^2(\mathcal{P}_N; \mathbf{A}) = \frac{1}{N^2} \cdot \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{1}{4\pi} \cdot \left[1 - 2\ln\left(1 + \sqrt{\frac{1 - \eta_i \cdot \eta_j}{2}}\right)\right].$$

This is the case considered in [10, p. 598], [43] and [26] (where it is called **D**−discrepancy, see p. 6).

A closed form arises also when $A_n = -n(n+1)$. In this case, described in detail in [20, Section 4.7], we have

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(t) \;=\; \frac{1}{4\pi} \cdot \left\{1 - \frac{\pi^2}{6} + \mathcal{L}_2\left(\frac{1+t}{2}\right)\right\}$$

where $\mathcal{L}_2(\cdot)$ is the so-called *dilogarithm*:

$$\mathcal{L}_2(u) = -\int_0^u \frac{\ln(1-v)}{v}\mathrm{d}v = \sum_{k=1}^{\infty} \frac{u^k}{k^2}.$$

The formula appearing in [20] is slightly more complicated than the one above, which is taken from [24, p. 142]. We remark, in passing, that this is equal, up to multiplication by $\frac{1}{4\pi}$, to (0.8) in [16], even if the formula in the last reference is even more complicated than those quoted above.

A case that has attracted a lot of interest arises if we consider the Riesz kernels (see, e.g., [11]), appearing in the analysis of s−energies ([30, 12]). For $\xi_1, \xi_2 \in \Omega$, we define

$$K(\xi_1, \xi_2) = \begin{cases} \mathrm{sign}(s) \cdot \|\xi_1 - \xi_2\|_2^{-s}, & s \neq 0, \\ -\ln\|\xi_1 - \xi_2\|_2^2, & s = 0, \end{cases}$$

where $\|\cdot\|_2$ is the Euclidean distance. The choices $s = 0$ and $s = 1$ lead respectively to the logarithmic and Coulombic potential. We can write this kernel as $K(\xi_1, \xi_2) = k_s(\xi_1 \cdot \xi_2)$, where

$$k_s(t) = \begin{cases} \mathrm{sign}(s) \cdot |2(1-t)|^{-\frac{s}{2}}, & s \neq 0, \\ -\ln 2(1-t), & s = 0, \end{cases}$$

for $t \in [-1, 1)$. When $s \neq 0$, the choice

$$A_n^2 = \frac{2^{s-2}\Gamma\left(\frac{s}{2}\right)\Gamma\left(-\frac{s}{2}+n+2\right)}{\pi\Gamma\left(\frac{s}{2}+n\right)\Gamma\left(1-\frac{s}{2}\right)}$$

for $s < 2$ yields

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(t) = |2(1-t)|^{-\frac{s}{2}} - \frac{2^{-s}}{1 - \frac{s}{2}} = \text{sign}(s) \cdot k_s(t) - \frac{2^{-s}}{1 - \frac{s}{2}}.$$

When $s = 0$, the choice

$$A_n^2 = \frac{n(n+1)}{4\pi}$$

yields

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(t) = -\ln 2(1-t) - \ln \frac{e}{4} = k_0(t) - \ln \frac{e}{4}.$$

This is simply a version of Pycke's statistic.

Note that $A_n^2 \sim n^{-s+2}$, so that $\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \sim \sum_{n=1}^{\infty} n^{s-1}$, and this converges if $s < 0$. Therefore the $D^2(\mathcal{P}_N; \mathbf{A})$ form of the discrepancy can be used only when $s < 0$. On the other hand, $d^2(\mathcal{P}_N; \mathbf{A})$ (see Theorem 4.1(iv) below) can be used if $\sum_{n=1}^{\infty} \left(\frac{2n+1}{4\pi A_n^2}\right)^2 \sim \sum_{n=1}^{\infty} n^{2s-2}$ converges: this holds if $s < \frac{1}{2}$.

The requirement that $s < 2$ is quite unfortunate. Generalized Riesz kernels remove it (see [13, p. 817]). In [5, Appendix B], some computations are available that allow one to recover A_n^2 for $s > 0$ in the case of generalized Riesz kernels.

Some new cases can be obtained from the mathematical literature on special functions. When $A_n^2 = (2n+1) \cdot \rho^{-n}$ for $\rho \in (-1, 1)$, the generating function of Legendre polynomials yields (see [39, p. 533]; see [28] for a more general formula):

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(t) = \frac{1}{4\pi} \frac{1}{(1 - 2\rho t + \rho^2)^{1/2}} - \frac{1}{4\pi},$$

while when $A_n^2 = \rho^{-n}$ for $\rho \in (0, 1)$ we get the Abel-Poisson kernel (see [39, p. 534]):

$$\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(t) = \frac{1}{4\pi} \frac{1 - \rho^2}{(1 - 2\rho t + \rho^2)^{3/2}} - \frac{1}{4\pi}.$$

Other situations in which closed form expressions exist are given in [16, pp. 41-42] and [24], but not all of them respect the condition $\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2} < \infty$ (see Theorem 4.1).

4. Asymptotic properties

The following theorem provides the most relevant asymptotic properties of the generalized discrepancies under the null hypothesis of uniform distribution of the sample of points $\mathcal{P}_N$.

Theorem 4.1. *Let $\mathcal{P}_N$ be a sample of iid uniform random variables on Ω, and suppose that the spherical symbol $\{A_n\}$ of the pseudodifferential operator $\mathbf{A}$ is such that $A_n \neq 0$ for any n. Then, if $\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2} < \infty$ the first three properties hold.*

(i) $D^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\text{as}} 0$.

(ii)

$$N \cdot D^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\mathcal{D}} \sum_{n=1}^{\infty} \frac{1}{4\pi A_n^2} \cdot \chi_n^2(2n+1),$$

where $\chi^2(k)$ denotes a χ^2 random variable with k degrees of freedom, and $\left\{\chi_n^2(2n+1); n \in \mathbb{N}\right\}$ is a sequence of independent χ^2 random variables.

(iii) *The following uniform bound holds:*

$$\left| \mathbb{P}\left\{N \cdot D^2 \leq y\right\} - \mathbb{P}\left\{\sum_{n=1}^{\infty} \frac{\chi_n^2(2n+1)}{4\pi A_n^2} \leq y\right\}\right|$$

$$\leq \frac{e^{c \cdot \lambda_{13}^{-1} \sqrt{\sum_{n=1}^{\infty} \frac{2n+1}{A_n^4}}}}{N} \cdot \left(1 + \frac{\left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2}\right)^{\frac{3}{2}}}{\left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^4}\right)^{\frac{3}{4}}}\right)$$

where λ_{13} is defined in the proof of the theorem.

(iv) *If* $\sum_{n=1}^{\infty}\left(\frac{2n+1}{A_n^2}\right)^2 < \infty$, *the modified statistic*

$$d^2(\mathcal{P}_N; \mathbf{A}) = \frac{1}{N^2} \sum_{i \neq j=1}^{N} \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n(\xi_i \cdot \xi_j)$$

obtained removing the diagonal elements is such that $d^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\text{as}} 0$ *and*

$$N \cdot d^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\mathcal{D}} \sum_{n=1}^{\infty} \frac{1}{4\pi A_n^2} \cdot \left[\chi_n^2(2n+1) - (2n+1)\right].$$

Remark 1. (i) When $A_n = \sqrt{(2n+1)n(n+1)}$, $\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2} = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1 < \infty$ and

$$N \cdot D^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\mathcal{D}} \sum_{n=1}^{\infty} \frac{\chi_n^2(2n+1)}{4\pi(2n+1)n(n+1)}.$$

When $A_n^2 = \rho^{-n}$, then $N \cdot D^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\mathcal{D}} \frac{1}{4\pi} \sum_{n=1}^{\infty} \rho^n \cdot \chi_n^2(2n+1)$. When $A_n = \sqrt{n(n+1)}$ (i.e., Pycke's statistic), we have to use the statistic d^2 and we get

$$N \cdot d^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\mathcal{D}} \sum_{n=1}^{\infty} \frac{\left[\chi_n^2(2n+1) - (2n+1)\right]}{4\pi n(n+1)}.$$

(ii) If $\mathbf{A}$ is such that $k_1 \cdot n^{1+\alpha} \leq A_n \leq k_2 \cdot n^{1+\alpha}$ for $\alpha > 0$ and $0 < k_1 \leq k_2 < +\infty$, then the bound has the following behavior in terms of N and α:

$$\left| \mathbb{P}\left\{N \cdot D^2 \leq y\right\} - \mathbb{P}\left\{\sum_{n=1}^{\infty} \frac{\chi_n^2(2n+1)}{4\pi A_n^2} \leq y\right\}\right|$$

$$\leq \frac{e^{c \cdot 9^\alpha \sqrt{\zeta(3+4\alpha)}}}{N} \cdot \left(1 + \frac{\zeta^{\frac{3}{2}}(1+2\alpha)}{\zeta^{\frac{3}{4}}(3+4\alpha)}\right)$$

where ζ is the Riemann zeta function. The only practical implication of α is in the exponent; this shows that a large value of α implies a worse convergence rate.

(iii) Adapting Theorem 4.4 in [21], it is possible to show the following fact: The condition $A_n^2 < +\infty$ for any n is equivalent to the fact that $D^2(\mathcal{P}_N; \mathbf{A}) \xrightarrow{\mathbb{P}} 0$ iff the sample $\mathcal{P}_N$ comes from a uniform distribution. This implies that a test based on a generalized discrepancy may converge to 0 also for nonuniform distributions if $A_n^2 = +\infty$ for some n. As an example, Giné's statistic (which has $A_n^2 = +\infty$ for n odd, see above) converges to 0 for every probability measure uniform on the projective space and not only on the sphere (see [21, p. 1262]).

As concerns the asymptotic properties under the alternative hypothesis that the points are independent realizations of a nonuniform distribution, we just quote them, together with references to results in the literature that allow for deriving them. Consider an iid sample $\mathcal{P}_N^\star$ drawn from a distribution $\mathbb{P}^\star$ on the sphere different from the uniform distribution. Note that, according to Remark 1 (iii), if $A_n^2 = +\infty$ for some n, then $D^2\left(\mathcal{P}_N^\star; \mathbf{A}\right)$ can converge in probability to 0 even if the sample $\mathcal{P}_N^\star$ does not come from a uniform distribution. The following properties hold if $D^2\left(\mathcal{P}_N^\star; \mathbf{A}\right)$ does not converge in probability to 0. Using Corollary 4 of [1, p. 5], it is possible to show that when $D^2\left(\mathcal{P}_N^\star; \mathbf{A}\right)$ is suitably centered and scaled, it converges in distribution to a standard normal random variable. A uniform bound on the distance between the finite sample and the asymptotic distribution can be obtained through [1, Corollary 5, p. 5] and decreases as $1/\sqrt{N}$. The asymptotic distribution of $N \cdot D^2\left(\mathcal{P}_N^\star; \mathbf{A}\right)$ under a local alternative converging to the null, i.e., under a Pitman drift, is given by a weighted sum of noncentral χ^2 random variables (see [23, Theorem 2.1]).

5. Computation of the asymptotic distribution

In this section, we consider approximations of the asymptotic distribution of $N \cdot D^2\left(\mathcal{P}_N; \mathbf{A}\right)$, denoted, in the following as $X = \sum_{n=1}^{\infty} \frac{\chi_n^2 (2n+1)}{4\pi A_n^2}$, through three different methods. This topic is relevant for any distribution given by a weighted sum of χ^2 random variables, but is particularly critical here. Indeed, contributions to X corresponding to large values of n (that is $\frac{\chi_n^2 (2n+1)}{4\pi A_n^2}$) can be quite relevant since the degrees of freedom of the χ^2 random variable depend on n (this problem does not arise with the most common test statistics on the real line, such as Cramér-von Mises and Anderson-Darling, and is a facet of the curse of dimensionality problem).

The first approximation amounts to considering a scaled and translated Γ distributed random variable X_Γ that matches the first three moments of X:

$$X_\Gamma = \mu + \frac{\Gamma\left(\frac{\nu}{2}, 2\right) - \nu}{\sqrt{2\nu}} \cdot \sigma$$

where $\Gamma(\alpha, \theta)$ denotes a Γ random variable with shape parameter α and scale parameter θ,

$$\mu = \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2}, \quad \sigma^2 = 2\sum_{n=1}^{\infty} \frac{2n+1}{\left(4\pi A_n^2\right)^2}$$

and

$$\nu = \left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^4}\right)^3 \Bigg/ \left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^6}\right)^2.$$

This approximation for linear combinations of χ^2 random variables has been considered in [6], in [31] and in [46], where uniform bounds are also derived.

We propose also the following truncated version of X:

$$X_\nu = \sum_{n=1}^{\nu} \frac{\chi_n^2 (2n+1)}{4\pi A_n^2}.$$

We will show that a better alternative is to use a centered version of X_ν:

$$X_\nu^\star = \sum_{n=1}^{\nu} \frac{\chi_n^2 (2n+1)}{4\pi A_n^2} + \sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}.$$

The computation of X_ν can then be easily performed using the algorithm described in [14] and linked in R from C in [7]. We majorize the errors implied by the replacement of X with X_ν or $X_\nu^\star$ using the following bounds.

Theorem 5.1. *We have:*

$$|\mathbb{P}\{X \le x\} - \mathbb{P}\{X_\nu \le x\}| \le \sqrt{\frac{2\sum_{n=\nu+1}^{\infty}\frac{2n+1}{A_n^4} + \left(\sum_{n=\nu+1}^{\infty}\frac{2n+1}{A_n^2}\right)^2}{\sum_{n=1}^{\nu}\frac{2n+1}{A_n^4}}}$$
$$\cdot \frac{2}{\pi\sqrt{3}} \cdot B\left(\frac{1}{2},\frac{1}{4}\right)$$

and

$$|\mathbb{P}\{X \le x\} - \mathbb{P}\{X_\nu^\star \le x\}| \le \frac{\sum_{n=\nu+1}^{\infty}\frac{2n+1}{A_n^4}}{\sum_{n=2}^{\nu}\frac{2n+1}{A_n^4}} \cdot \frac{5}{\pi}.$$

If $\mathbf{A}$ is such that $A_n^{-1} \sim k \cdot n^{-1-\alpha}$ for $\alpha > 0$ (so that $\mathbf{A}$ is a pseudodifferential operator of order $t = 1 + \alpha$ with const $= \frac{1}{k}$), we have:

$$|\mathbb{P}\{X \le x\} - \mathbb{P}\{X_\nu \le x\}| \le \frac{k^2}{\alpha\sigma} \cdot \frac{\sqrt{3}}{16\pi^2} \cdot B\left(\frac{1}{2},\frac{1}{4}\right) \cdot \nu^{-2\alpha} \cdot (1 + o(1)),$$

$$|\mathbb{P}\{X \le x\} - \mathbb{P}\{X_\nu^\star \le x\}| \le \frac{k^4}{(1+2\alpha) \cdot \left\{\sigma^2 - \frac{3}{8\pi^2 A_1^4}\right\}} \cdot \frac{5}{8\pi^3}$$
$$\cdot \nu^{-2-4\alpha} \cdot (1 + o(1))$$

where σ^2 has been defined above.

Remark 2. (i) Both bounds are tight, as will be shown in Section 6.

(ii) The previous bounds assert that the approximation of X through X_ν and $X_\nu^\star$ is better when the order of the pseudodifferential operator $\mathbf{A}$ is higher. This is in line with [10, p. 602], where it is stated that computation of $D^2(\mathcal{P}_N; \mathbf{A})$ is easier the larger t is.

(iii) For the interesting case $A_n = \sqrt{(2n+1)n(n+1)}$, we have $k = 1/\sqrt{2}$ and $\alpha = 1/2$, so that

$$|\mathbb{P}\{X \le x\} - \mathbb{P}\{X_\nu \le x\}| \le 6.41233 \cdot \nu^{-1} \cdot (1 + o(1)),$$
$$|\mathbb{P}\{X \le x\} - \mathbb{P}\{X_\nu^\star \le x\}| \le 28.33329 \cdot \nu^{-4} \cdot (1 + o(1)),$$

where $X_\nu^\star = X_\nu + \frac{1}{4\pi(\nu+1)}$. Even if the constant of the second bound is quite large, the bound shows that the approximation is very good even for small values of ν. When $A_n^2 = (2n+1) \cdot \rho^{-n}$, the first bound is $O(\rho^\nu)$ while the second is $O(\rho^{2\nu})$; when $A_n^2 = \rho^{-n}$, the former is $O(\nu\rho^\nu)$ while the latter is $O(\nu\rho^{2\nu})$.

(iv) When using the statistic $d^2(\mathcal{P}_N; \mathbf{A})$, centering is automatic and only the second bound applies.

6. SIMULATION RESULTS AND APPLICATIONS

In this section we deal with four distinct topics, namely the rate of convergence of the finite-sample distribution of $N \cdot D^2(\mathcal{P}_N; \mathbf{A})$ to the asymptotic one, the accuracy of the approximations X_Γ, X_ν and $X_\nu^\star$ to X, the use of our asymptotic results to

evaluate the quality of quasi-Monte Carlo sequences on the sphere, and at last a statistical application to magnetic orientations.

When not explicitly stated otherwise, we consider only the Cui-Freeden test statistic $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ for $A_n = \sqrt{(2n+1) \, n \, (n+1)}$, since in this case a closed form expression exists and the previously quoted topics can be investigated without further problems arising in the computation of the test statistic through Legendre polynomials.

6.1. **Rate of convergence to the asymptotic distribution.** In Figure 1, we have reproduced the finite-sample cdf (say F_N) for $N \in \{10, 30, 100\}$ and the asymptotic cdf (say F_∞). The finite-sample distributions are based on Monte Carlo sampling of $S = 10,000,000$ values of the test statistic $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$. In this case, the Laws of the Iterated Logarithm for the empirical cdf (see, e.g., [44, p. 268]) prescribe that (asymptotically) the maximal error in the Monte Carlo approximation of F_N through $10,000,000$ points is at most 0.0003728232 and at least 0.0002106624.
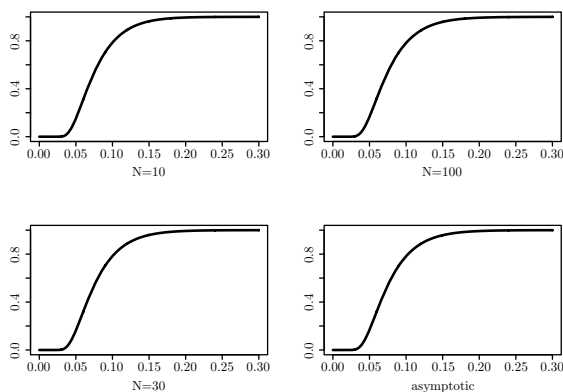

FIGURE 1. Distribution of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ for $N = 10$, 30, 100 and ∞.

From the graphs it is not possible to devise any deviation of the finite-sample distributions from the asymptotic one. Therefore, in Figure 2 we have reproduced also the deviation $F_N - F_\infty$ between the finite-sample and the asymptotic distributions. It turns out that the convergence is indeed very fast and the error for $N = 10$ is around 0.015. The decrease in $1/N$ in the figure is coherent with the bound in Theorem 4.1.

6.2. **Accuracy of approximations to the asymptotic distribution.** In Figure 3, we represent the deviation between the asymptotic distribution of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ and the proposed approximations. UNC-Real is the deviation $\sup_x |\mathbb{P} \{X \leq x\} - \mathbb{P} \{X_\nu \leq x\}|$, while CEN-Real is $\sup_x |\mathbb{P} \{X \leq x\} - \mathbb{P} \{X_\nu^\star \leq x\}|$. UNC-Comp and CEN-Comp are the bounds computed through Theorem 5.1. It is possible to see that both UNC-Comp and CEN-Comp are precise upper bounds respectively on UNC-Real and CEN-Real. The most striking feature arising from the graph is the dramatic improvement in the performance of the approximation
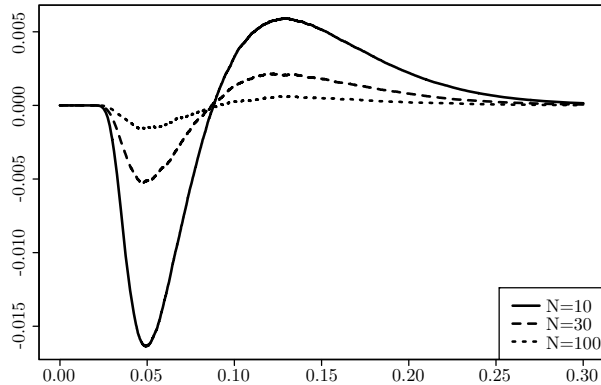
FIGURE 2. Deviation of the distribution of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ for $N = 10, 30, 100$ with respect to the asymptotic one.
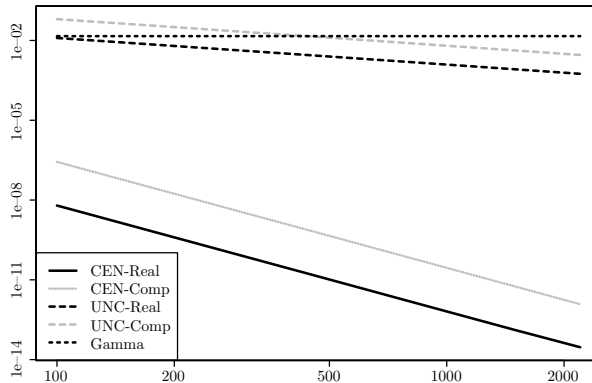

FIGURE 3. Accuracy of the approximations to the asymptotic distribution of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$.

method that can be obtained simply through centering, even if this operation involves a shift of the distribution through a small constant such as $\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2} = \frac{1}{4\pi(\nu+1)}$. Centering allows for a reduction of the error behavior from $\Theta(\nu^{-1})$ to $\Theta(\nu^{-4})$ as shown by Theorem 5.1.

From the previous figure, it seems that the Γ approximation cannot compete with the other methods, but the use of a uniform bound hides the fact that for most quantiles of interest (confined to the upper tail) the Γ approximation is extremely good. In Figure 4, the quantities $\mathbb{P}\{X_\nu^\star \leq x\} - \mathbb{P}\{X \leq x\}$, $\mathbb{P}\{X_\nu \leq x\} - \mathbb{P}\{X \leq x\}$ and $\mathbb{P}\{X_\Gamma \leq x\} - \mathbb{P}\{X \leq x\}$ are shown for $\nu = 100$, that is when the errors in the approximation of X through X_ν and X_Γ are almost equal: this comparison shows that, while $X_\nu^\star$ is by far the best method, X_Γ approximates X better than X_ν in the upper tail. The graph of the Gamma approximation is strikingly similar to the one in [6, p. 155].
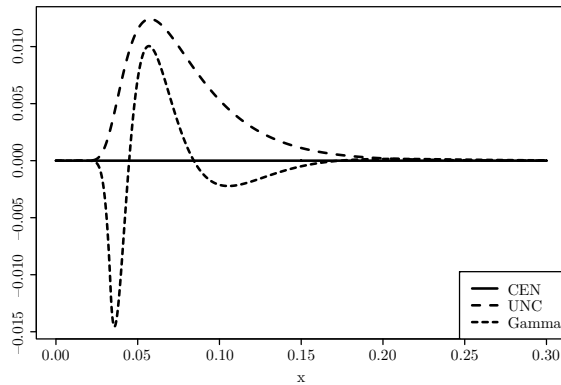
FIGURE 4. Deviation of X_{100}, $X_{100}^\star$ and X_Γ with respect to the asymptotic distribution of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$.

6.3. **Uniformity of quasi-Monte Carlo sequences.** In this section, we show how generalized discrepancies can be used in the assessment of low-discrepancy (often called quasi-Monte Carlo) sequences.[3]

We recall the definition of the Hammersley sequence in the two-dimensional hypercube and on the sphere (see [10, p. 605]). Let p be an integer greater than or equal to 2. Then any integer n can be expressed as $n = a_0 + \cdots + a_r p^r$, for an integer r and positive or null integer coefficients $a_0, \ldots, a_r$. The nth element of the p–adic van der Corput sequence is defined by $x_n^{(p)} = \frac{a_0}{p} + \cdots + \frac{a_r}{p^{r+1}}$. The Hammersley sequence of length N in the two-dimensional hypercube is defined by $(t_n, \phi_n)_{n=1,\ldots,N} = \left(\frac{2n-1}{2N}, x_n^{(p)} \right)_{n=1,\ldots,N}$. The Hammersley sequence on the sphere is defined by applying to each element of the sequence $(t_n, \phi_n)_{n=1,\ldots,N}$ the following transformation:

$$(t, \phi) \mapsto \left(\sqrt{1 - (2t-1)^2} \cdot \cos(2\pi\phi), \sqrt{1 - (2t-1)^2} \cdot \sin(2\pi\phi), 2t - 1 \right).$$

In [45], the Hammersley sequence has been proposed for the solution of some problems in computer graphics. In particular, the authors remark that when p (in their paper denoted as p_1) increases, the uniformity of the sequence seems to get worse:

> Figures 2 and 3 show the Hammersley points with different bases, on the plane and the sphere respectively. When $p_1 = 2$, it gives the best uniformly distributed pattern on the sphere. Cui et al. [5] measures the uniformity of Hammersley points with $p_1 = 2$ on the sphere using the generalized discrepancy. It gives the lowest generalized discrepancy (most uniformly distributed) among the methods tested. As p_1 increases (from Figures 3(b) to 3(f)), points start to line up and form regular lines on the sphere. The position of the pole (marked with an arrow) becomes distinguishable from the pattern. [...] The Hammersley point set with $p_1 = 2$ gives the

---

[3]For other computational studies concerning sequences on the sphere, the reader may consult [10, 19, 26, 2, 29, 43] and the references therein.

> most uniformly distributed sampling pattern. For higher p_1, the
> points tend to align and reduce its usefulness.

Here, we quantify the qualitative remark in [45]. Therefore, we have computed the value of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ when $\mathcal{P}_N$ is a Hammersley sequence on the sphere for N varying in the interval $[10, 10000]$ and for p prime and comprised between 2 and 23 (the restriction to prime numbers is not necessary, but customary; see, e.g., [17, p. 28]). The results are represented in Figure 5: from this, it is evident that the sequence with $p = 2$, represented by the continuous line, is uniformly the best one. When p increases, the generalized discrepancy $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$, represented by a dashed line, shows an oscillating behavior but remains consistently above the (black) curve with $p = 2$. By the way, this is coherent with the behavior of the star discrepancy of Hammersley sequences that can be shown to worsen when p increases (see [32, Theorem 3.8]).
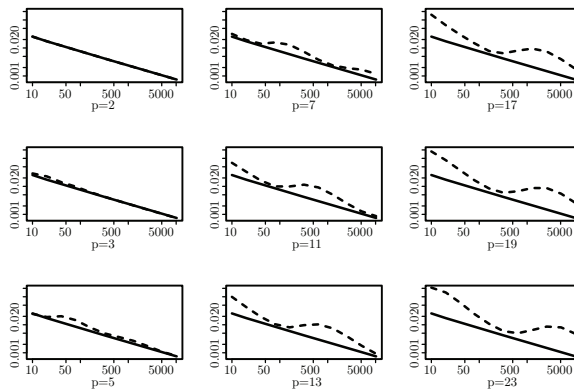

FIGURE 5. Values of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ when $\mathcal{P}_N$ is an Hammersley sequence with varying values of N (on the horizontal axis) and of p.

6.4. **A statistical example.** We have considered a classical dataset concerning 52 measurements of magnetic remanence from specimens of red beds from Bowen Basin, Queensland, after thermal demagnetisation to $670°C$ (see Table B5 in [18]). We have computed seven statistics as $D^2 (\mathcal{P}_N; \mathbf{A}) = \frac{1}{N^2} \cdot \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n (\xi_i \cdot \xi_j)$ (Pycke's statistic is the only one computed removing the elements with $i = j$ in the double sum over i and j), and we have reported the values of $N \cdot D^2 (\mathcal{P}_N; \mathbf{A})$ in Table 1, along with the other values of interest.[4] The values of the cumulative distribution function have been computed through X_ν and $X_\nu^\star$, where $\nu = 1000$. The seven statistics yield similar results in the sense that all of them fail to reject the null hypothesis of uniformity, even if one of them (Giné's statistic) comes very near to reject it at 90%. The respective error bounds are also displayed in parentheses. The difference in the bounds for the centered and the uncentered cases is eloquent. It should also be remarked that the bounds for the Cui-Freeden statistic are coherent with the rates of decrease derived in Remark 2 (iii).

---

[4]In these computations, we do not consider $s-$energies since they cause some peculiar problems that will be dealt with in a separate paper.

TABLE 1. Computation of the test statistics on the data described in the text.

Name	A_n^2	$\sum_{n=1}^{\infty}\frac{2n+1}{4\pi A_n^2}\cdot P_n(t)$	Asymptotic distribution
Rayleigh	$\begin{cases}\frac{3}{4\pi} & n=1 \\ +\infty & n\neq 1\end{cases}$	t	$\chi_n^2(3)/3$
Bingham	$\begin{cases}\frac{5}{4\pi} & n=2 \\ +\infty & n\neq 2\end{cases}$	$-\frac{1}{2}\cdot(1-3\cdot t^2)$	$\chi_n^2(5)/5$
Beran/Ajne	$\begin{cases}+\infty & n\text{ even} \\ \left(n\cdot\frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)}\right)^2 & n\text{ odd}\end{cases}$	$\frac{1}{4}-(2\pi)^{-1}\cos^{-1}t$	$\sum_{n=1}^{\infty}\frac{\left[B\left(\frac{2n-1}{2},\frac{3}{2}\right)\right]^2\cdot\chi_n^2(4n-1)}{16\pi\left[\Gamma\left(\frac{3}{2}\right)\right]^2}$
Giné	$\begin{cases}+\infty & n\text{ odd} \\ \frac{n-1}{n+2}\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}\right)^2 & n\text{ even}\end{cases}$	$\frac{1}{2}-2\pi^{-1}\sin\cos^{-1}t$	$\sum_{n=1}^{\infty}\frac{(2n-1)\left[B\left(\frac{2n-1}{2},\frac{3}{2}\right)\right]^2\chi_n^2(4n+1)}{8\pi(n+1)\left[\Gamma\left(\frac{3}{2}\right)\right]^2}$
Cui-Freeden	$(2n+1)n(n+1)$	$\frac{1}{4\pi}\left[1-2\ln\left(1+\sqrt{\frac{1-t}{2}}\right)\right]$	$\sum_{n=1}^{\infty}\frac{\chi_n^2(2n+1)}{4\pi(2n+1)n(n+1)}$
Freeden-Schreiner	$n^2(n+1)^2$	$\frac{1}{4\pi}\cdot\left\{1-\frac{\pi^2}{6}+\mathcal{L}_2\left(\frac{1+t}{2}\right)\right\}$	$\sum_{n=1}^{\infty}\frac{\chi_n^2(2n+1)}{4\pi n^2(n+1)^2}$
Pycke	$n(n+1)$	$-\frac{1}{4\pi}\cdot\ln\frac{e}{2}(1-t)$	$\sum_{n=1}^{\infty}\frac{\chi_n^2(2n+1)-(2n+1)}{4\pi n(n+1)}$

Name	$N\cdot D^2(\mathcal{P}_N;\mathbf{A})$	p−value UNC	p−value CEN
Rayleigh	0.7539653	0.5198599 (—)	0.5198599 (—)
Bingham	1.325083	0.2500217 (—)	0.2500217 (—)
Beran/Ajne	0.2226953	0.4652019 (0.002812024)	0.4656521 (8.269991e-11)
Giné	0.7638484	0.1023675 (0.008622067)	0.1027497 (7.939330e-12)
Cui-Freeden	0.0881051	0.3134682 (0.006405284)	0.3142182 (2.822021e-11)
Freeden-Schreiner	0.0744169	0.4369813 (4.392691e-06)	0.436982 (1.234307e-16)
Pycke	0.1531135	— (—)	0.09439549 (6.35351e-06)

7. Proofs

In the following $\omega\left(\cdot\right)$ will denote the uniform measure on the sphere Ω and $\omega^{\star}\left(\cdot\right) := \omega\left(\cdot\right)/\omega\left(\Omega\right)$ the uniform probability measure corresponding to $\omega\left(\cdot\right)$.

Proof of Theorem 4.1. We show that $D^2\left(\mathcal{P}_N; \mathbf{A}\right)$ is a *degenerate V−statistic* with respect to $\omega\left(\cdot\right)$ (see [40, Chapter 5]). Indeed, from the definition of the generalized discrepancy, it can be seen that

$$N \cdot D^2\left(\mathcal{P}_N; \mathbf{A}\right) = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} h\left(\xi_i, \xi_j\right)$$

where $h\left(\xi_i, \xi_j\right) = \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(\xi_i \cdot \xi_j\right)$ is called the *kernel*. Since $h\left(\xi_i, \xi_i\right) = \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2}$, we can write:

$$N \cdot D^2\left(\mathcal{P}_N; \mathbf{A}\right) = \frac{2}{N} \cdot \sum_{i=1}^{N} \sum_{j=1}^{i-1} h\left(\xi_i, \xi_j\right) + \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2}.$$

The first term is a $U-statistic$ (see [40, Chapter 5]). From the addition theorem of spherical harmonics, we have

$$
\begin{aligned}
h\left(\xi_i, \xi_j\right) &= \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(\xi_i \cdot \xi_j\right) \\
&= \sum_{n=1}^{\infty} \frac{1}{A_n^2} \cdot \sum_{k=-n}^{n} Y_{n,k}\left(\xi_i\right) \cdot Y_{n,k}\left(\xi_j\right),
\end{aligned}
$$

so that it is simple to see that $\int_{\Omega} h\left(\eta, \xi\right) \omega^{\star}\left(\mathrm{d}\eta\right) = \int_{\Omega} \int_{\Omega} h\left(\eta, \xi\right) \omega^{\star}\left(\mathrm{d}\eta\right) \omega^{\star}\left(\mathrm{d}\xi\right) = 0$. In this case the kernel is said to be *degenerate*. A very important role in this situation is played by the eigenvalues of the integral operator

$$\mathcal{A}m\left(\xi\right) = \int_{\Omega} h\left(\eta, \xi\right) m\left(\eta\right) \omega^{\star}\left(\mathrm{d}\eta\right).$$

We can identify the eigenvalues of $\mathcal{A}$ with the sequence $\{(\frac{1}{4\pi A_n^2})^{\otimes(2n+1)}; n \in \mathbb{N}\}$ where $x^{\otimes i}$ means that x is repeated i times. The presence of the term $(4\pi)^{-1}$ is due to the fact that the eigenvalues of $\mathcal{A}$ are defined with respect to the uniform measure on the sphere. The reason is very simple to see if $\mathbf{A}$ is invertible. Indeed, from the fact that the pseudodifferential operator $\mathbf{A}$ is invariant (see equation on p. 597 in [10]) and invertible, we get $\mathbf{A}^{-2}Y_{n,j} = A_n^{-2} \cdot Y_{n,j}$, $n = 0, 1, \ldots, j = -n, \ldots, n$. However, the eigenvalues of $\mathcal{A}$ are defined with respect to the probability measure $\omega^{\star}\left(\cdot\right)$, so that

$$\mathcal{A}Y_{n,j} = \frac{1}{\omega\left(\Omega\right)} \cdot \mathbf{A}^{-2}Y_{n,j} = \frac{1}{4\pi A_n^2} \cdot Y_{n,j}.$$

In all other cases (that is when $A_n = +\infty$ for at least one value of $n \in \mathbb{N}$), the appearance of the $(4\pi)^{-1}$ term is linked to the fact that the orthonormality of the eigenfunctions of the operator $\mathcal{A}$ are defined with respect to the probability measure $\omega^{\star}\left(\cdot\right)$, while the spherical harmonics are orthonormal with respect to the measure $\omega\left(\cdot\right)$. In the following, the eigenvalues of $\mathcal{A}$, counted with their multiplicities and arranged in order of decreasing absolute value, will be denoted as $\{\lambda_n; n \in \mathbb{N}\}$.

(i) We show that, if the points are uniformly distributed, then $D\left(\mathcal{P}_N; \mathbf{A}\right)$ converges to 0. When $\mathbb{E}\left|h\left(\eta, \xi\right)\right| < \infty$, the strong law of large numbers holds for

$D\left(\mathcal{P}_N; \mathbf{A}\right)$ (see [40, Theorem A, p. 190]) and $D^2\left(\mathcal{P}_N; \mathbf{A}\right) \xrightarrow{\text{as}} 0$. But $\mathbb{E}\left|h\left(\eta, \xi\right)\right| = \mathbb{E}\left|\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(\eta \cdot \xi\right)\right|$ and this is guaranteed by the condition $\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2} < \infty$.

(ii) For degenerate $U-$statistics, the asymptotic distribution is given in [40, Theorem on p. 194]. Therefore, we get

$$N \cdot D^2\left(\mathcal{P}_N; \mathbf{A}\right) \xrightarrow{\mathcal{D}} \sum_{n=1}^{\infty} \frac{1}{4\pi A_n^2} \cdot \sum_{k=-n}^{n} \left[\chi_{n,k}^2\left(1\right) - 1\right] + \sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2}$$

where $\{\chi_{n,k}^2\left(1\right); n = 0, 1, \ldots; k = -n, \ldots, n\}$ is a sequence of independent chi-squared random variables with one degree of freedom. Using the classical relation $\sum_{k=-n}^{n} \chi_{n,k}^2\left(1\right) = \chi_n^2\left(2n+1\right)$ the result follows.

(iii) As concerns the uniform bound, from [3, Theorem 1.1] we have

$$\left|\mathbb{P}\left\{N \cdot D^2\left(\mathcal{P}_N; \mathbf{A}\right) \leq y\right\} - \mathbb{P}\left\{\sum_{n=1}^{\infty} \frac{\chi_n^2\left(2n+1\right)}{4\pi A_n^2} \leq y\right\}\right|$$

$$\leq \frac{\exp\left\{\frac{c\sqrt{\gamma_2}}{|\lambda_{13}|}\right\}}{N} \cdot \left(\frac{\gamma_3}{\gamma_2^{\frac{3}{2}}} + \frac{\gamma_{2,2}}{\gamma_2^2}\right) \leq \frac{\exp\left\{\frac{c\sqrt{\gamma_2}}{|\lambda_{13}|}\right\}}{N} \cdot \left(\frac{\gamma_4^{\frac{3}{4}}}{\gamma_2^{\frac{3}{2}}} + \frac{\gamma_{2,2}}{\gamma_2^2}\right)$$

where λ_{13} is the 13th element of the sequence of eigenvalues of $\mathcal{A}$ (counted with their multiplicities and arranged in order of decreasing absolute value), and

(7.1) $$\gamma_s = \mathbb{E}\left|h\left(\xi_i, \xi_j\right)\right|^s, \qquad \gamma_{s,r} = \mathbb{E}\left\{\mathbb{E}\left[\left|h\left(\xi_i, \xi_j\right)\right|^s \mid \xi_i\right]\right\}^r.$$

We get

$$\gamma_2 = \mathbb{E}\left|h\left(\xi_i, \xi_j\right)\right|^2 \asymp \sum_{n=1}^{\infty} \frac{2n+1}{A_n^4}$$

and

$$\gamma_4 = \mathbb{E}\left|h\left(\xi_i, \xi_j\right)\right|^4 \lesssim \left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^4}\right) \cdot \left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^2}\right)^2.$$

As concerns $\gamma_{2,2}$, we use the relation $\sum_{j=-n}^{n} \left|Y_{n,j}\left(\xi\right)\right|^2 = \frac{2n+1}{4\pi}$ to get

$$\gamma_{2,2} = \mathbb{E}\left\{\mathbb{E}\left[\left|h\left(\xi_i, \xi_j\right)\right|^2 \mid \xi_i\right]\right\}^2$$

$$= \mathbb{E}\left\{\sum_{n=1}^{\infty} \frac{1}{A_n^4} \cdot \sum_{k=-n}^{n} \left|Y_{n,k}\left(\xi_i\right)\right|^2\right\}^2$$

$$\asymp \left(\sum_{n=1}^{\infty} \frac{2n+1}{A_n^4}\right)^2. \qquad \qquad \square$$

Proof of Theorem 5.1. We start with the uncentered case. Consider the following bound obtained through the inversion formula

$$
\begin{aligned}
\left|\mathbb{P}\left\{X \leq x\right\} - \mathbb{P}\left\{X_\nu \leq x\right\}\right| &\leq \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1}{|t|}\cdot\left|\phi_X(t)-\phi_{X_\nu}(t)\right|\mathrm{d}t \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1}{|t|}\cdot\left|\phi_{X_\nu}(t)\right|\cdot\left|\frac{\phi_X(t)}{\phi_{X_\nu}(t)}-1\right|\mathrm{d}t \\
&= \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1}{|t|}\cdot\left|\phi_{X_\nu}(t)\right|\cdot\left|\mathbb{E}e^{it(X-X_\nu)}-1\right|\mathrm{d}t \\
&\leq \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{1}{|t|}\cdot\left|\phi_{X_\nu}(t)\right|\cdot\mathbb{E}\left|e^{it(X-X_\nu)}-1\right|\mathrm{d}t \\
&\leq \frac{\sqrt{\mathbb{E}\left(X-X_\nu\right)^2}}{\pi}\cdot\int_{-\infty}^{+\infty}\left|\phi_{X_\nu}(t)\right|\mathrm{d}t
\end{aligned}
$$

where the third step comes from the fact that X_ν and $X - X_\nu$ are independent and the fifth step from the inequalities $\left|e^{ix}-1\right|\leq|x|$ and $\mathbb{E}\left|X-X_\nu\right|\leq\sqrt{\mathbb{E}\left(X-X_\nu\right)^2}$ (see [42, Lemma 4.2, Chapter 13, p. 352], and [40, p. 197] for a similar majorization). Then we have

$$
\begin{aligned}
\phi_{X_\nu}(t) &= \prod_{n=1}^{\nu}\left(1-\frac{2it}{4\pi A_n^2}\right)^{-\frac{2n+1}{2}} = \exp\left\{-\sum_{n=1}^{\nu}\frac{2n+1}{2}\ln\left(1-\frac{it}{2\pi A_n^2}\right)\right\} \\
&= \exp\left\{-\sum_{n=1}^{\nu}\frac{2n+1}{4}\ln\left(1+\frac{t^2}{4\pi^2 A_n^4}\right)\right. \\
&\qquad\qquad \left. -i\sum_{n=1}^{\nu}\frac{2n+1}{2}\arctan\left(-\frac{t}{2\pi A_n^2}\right)\right\} \\
&= \exp\left\{-\sum_{n=1}^{\nu}\frac{2n+1}{4}\ln\left(1+\frac{t^2}{4\pi^2 A_n^4}\right)\right\} \\
&\qquad \cdot\left\{\cos\left[\sum_{n=1}^{\nu}\frac{2n+1}{2}\arctan\left(-\frac{t}{2\pi A_n^2}\right)\right]\right. \\
&\qquad\qquad \left. -i\cdot\sin\left[\sum_{n=1}^{\nu}\frac{2n+1}{2}\arctan\left(-\frac{t}{2\pi A_n^2}\right)\right]\right\} \\
\left|\phi_{X_\nu}(t)\right| &= \exp\left\{-\sum_{n=1}^{\nu}\frac{2n+1}{4}\ln\left(1+\frac{t^2}{4\pi^2 A_n^4}\right)\right\} = \prod_{n=1}^{\nu}\left(1+\frac{t^2}{4\pi^2 A_n^4}\right)^{-\frac{2n+1}{4}} \\
&= \frac{1}{\prod_{n=1}^{\nu}\left(1+\frac{t^2}{4\pi^2 A_n^4}\right)^{\frac{2n+1}{4}}} \leq \frac{1}{\left(1+t^2\cdot\sum_{n=1}^{\nu}\frac{2n+1}{12\pi^2 A_n^4}\right)^{\frac{3}{4}}}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left(X - X_\nu\right)^2 &= \mathbb{E}\left(\sum_{n=\nu+1}^{\infty} \frac{\chi_n^2 \left(2n+1\right)}{4\pi A_n^2}\right)^2 \\
&= \mathbb{E}\left(\sum_{n=\nu+1}^{\infty} \frac{\chi_n^2 \left(2n+1\right)}{4\pi A_n^2} - \sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2 + \left(\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2 \\
&= \sum_{n=\nu+1}^{\infty} \frac{1}{16\pi^2 A_n^4} \cdot \mathbb{E}\left(\chi_n^2 \left(2n+1\right) - 2n - 1\right)^2 + \left(\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2 \\
&= \sum_{n=\nu+1}^{\infty} \frac{2\left(2n+1\right)}{16\pi^2 A_n^4} + \left(\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\left|\mathbb{P}\left\{X \leq x\right\} - \mathbb{P}\left\{X_\nu \leq x\right\}\right| \\
&\leq \frac{\sqrt{2\sum_{n=\nu+1}^{\infty} \frac{2n+1}{16\pi^2 A_n^4} + \left(\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2}}{\pi} \cdot \int_{-\infty}^{+\infty} \frac{1}{\left(1 + t^2 \cdot \frac{3}{4}\sum_{n=1}^{\nu} \frac{2n+1}{16\pi^2 A_n^4}\right)^{\frac{3}{4}}} \mathrm{d}t \\
&= \frac{1}{\pi}\sqrt{\frac{2\sum_{n=\nu+1}^{\infty} \frac{2n+1}{16\pi^2 A_n^4} + \left(\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2}{\frac{3}{4}\sum_{n=1}^{\nu} \frac{2n+1}{16\pi^2 A_n^4}}} \cdot \int_0^{+\infty} \frac{1}{\left(1+z\right)^{\frac{3}{4}} z^{\frac{1}{2}}} \mathrm{d}z \\
&= \sqrt{\frac{2\sum_{n=\nu+1}^{\infty} \frac{2n+1}{16\pi^2 A_n^4} + \left(\sum_{n=\nu+1}^{\infty} \frac{2n+1}{4\pi A_n^2}\right)^2}{\frac{3}{4}\sum_{n=1}^{\nu} \frac{2n+1}{16\pi^2 A_n^4}}} \cdot \frac{1}{\pi} \cdot B\left(\frac{1}{2}, \frac{1}{4}\right)
\end{aligned}
$$

where B is the beta function.

In the centered case, on the other hand, $\left|\phi_{X_\nu^\star}\left(t\right)\right|$ yields

$$
\begin{aligned}
\left|\phi_{X_\nu^\star}\left(t\right)\right| &= \left|\phi_{X_\nu}\left(t\right)\right| = \frac{1}{\prod_{n=1}^{\nu}\left(1 + \frac{t^2}{4\pi^2 A_n^4}\right)^{\frac{2n+1}{4}}} \\
&\leq \frac{1}{\left(1 + \frac{t^2}{4\pi^2 A_1^4}\right)^{\frac{3}{4}}\left(1 + t^2 \cdot \sum_{n=2}^{\nu} \frac{2n+1}{20\pi^2 A_n^4}\right)^{\frac{5}{4}}} \\
&\leq \frac{1}{\left(1 + t^2 \cdot \sum_{n=2}^{\nu} \frac{2n+1}{20\pi^2 A_n^4}\right)^{\frac{5}{4}}},
\end{aligned}
$$

and $\left|\mathbb{E}e^{it(X-X_\nu^\star)} - 1\right|$ becomes

$$
\begin{aligned}
\left|\mathbb{E}e^{it(X-X_\nu^\star)} - 1\right| &= \left|\mathbb{E}e^{it(X-X_\nu^\star)} - 1 - \mathbb{E}it\left(X - X_\nu^\star\right)\right| \\
&\leq \mathbb{E}\left|e^{it(X-X_\nu^\star)} - 1 - it\left(X - X_\nu^\star\right)\right| \\
&\leq \frac{t^2 \cdot \mathbb{E}\left|X - X_\nu^\star\right|^2}{2}
\end{aligned}
$$

(always through [42, Lemma 4.2, Chapter 13, p. 352]) where

$$\mathbb{E}\left(X - X_\nu^\star\right)^2 = \mathbb{V}\left(\sum_{n=\nu+1}^{\infty} \frac{\chi_n^2\left(2n+1\right)}{4\pi A_n^2}\right) = \sum_{n=\nu+1}^{\infty} \frac{2n+1}{8\pi^2 A_n^4}.$$

Therefore,

$$\left|\mathbb{P}\left\{X \le x\right\} - \mathbb{P}\left\{X_\nu^\star \le x\right\}\right| \le \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{|t|} \cdot \left|\phi_{X_\nu^\star}\left(t\right)\right| \cdot \mathbb{E}\left|e^{it(X-X_\nu^\star)} - 1\right| \mathrm{d}t$$

$$\le \frac{2}{\pi} \frac{\mathbb{E}\left|X - X_\nu^\star\right|^2}{2} \cdot \int_0^{+\infty} \frac{t}{\left(1 + t^2 \cdot \frac{4}{5} \sum_{n=2}^{\nu} \frac{2n+1}{16\pi^2 A_n^4}\right)^{\frac{5}{4}}} \mathrm{d}t$$

$$= \frac{\sum_{n=\nu+1}^{\infty} \frac{2n+1}{16\pi^2 A_n^4}}{\sum_{n=2}^{\nu} \frac{2n+1}{16\pi^2 A_n^4}} \cdot \frac{5}{4\pi} \cdot B\left(1, \frac{1}{4}\right).$$

The other relations are obvious from $A_n^{-1} \sim k \cdot n^{-1-\alpha}$, using the fact that

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \frac{2n+1}{4\pi A_n^2} \cdot P_n\left(\xi_i \cdot \xi_j\right)\right]^2 = \sum_{n=1}^{\infty} \frac{2n+1}{16\pi^2 A_n^4}. \qquad \square$$

References

1. I.B. Alberink and V. Bentkus, *Berry-Esseen bounds for von Mises and U-statistics*, Liet. Mat. Rink. **41** (2001), no. 1, 1–20. MR1849804 (2002f:60036)
2. B. Bajnok, S.B. Damelin, J. Li, and G.L. Mullen, *A constructive finite field method for scattering points on the surface of d-dimensional spheres*, Computing **68** (2002), no. 2, 97–109. MR1901138 (2003h:11086)
3. V. Bentkus and F. Götze, *Optimal bounds in non-Gaussian limit theorems for U-statistics*, Ann. Probab. **27** (1999), no. 1, 454–521. MR1681161 (2000f:62032)
4. R.J. Beran, *Testing for uniformity on a compact homogeneous space*, J. Appl. Probability **5** (1968), 177–195. MR0228098 (37:3682)
5. J.S. Brauchart, *Points on an unit sphere in R^{d+1}, Riesz energy, discrepancy and numerical integration*, Ph.D. thesis, Graz University of Technology (TU-Graz), 2005.
6. M.J. Buckley and G.K. Eagleson, *An approximation to the distribution of quadratic forms in normal random variables*, Aust. J. Stat. **30A** (1988), no. Spec. Issue, 150–159.
7. C. Choirat and R. Seri, *The asymptotic distribution of quadratic discrepancies*, Monte Carlo and quasi-Monte Carlo methods 2004 (Berlin), Springer, 2006, pp. 61–76. MR2208702 (2006k:65006)
8. ———, *Statistical properties of generalized discrepancies*, Math. Comp. **77** (2008), no. 261, 421–446 (electronic). MR2353960 (2009c:62028)
9. ———, *Numerical properties of generalized discrepancies on spheres of arbitrary dimension*, J. Complexity **29** (2013), no. 2, 216–235. MR3018140
10. J. Cui and W. Freeden, *Equidistribution on the sphere*, SIAM J. Sci. Comput. **18** (1997), no. 2, 595–609. MR1433797 (97k:65024)
11. S.B. Damelin, *A walk through energy, discrepancy, numerical integration and group invariant measures on measurable subsets of Euclidean space*, Numer. Algorithms **48** (2008), no. 1-3, 213–235. MR2413284 (2009d:41002)
12. S.B. Damelin and P.J. Grabner, *Energy functionals, numerical integration and asymptotic equidistribution on the sphere*, J. Complexity **19** (2003), no. 3, 231–246. MR1984111 (2004e:41032)
13. S.B. Damelin, F.J. Hickernell, D.L. Ragozin, and X. Zeng, *On energy, discrepancy and group invariant measures on measurable subsets of Euclidean space*, J. Fourier Anal. Appl. **16** (2010), no. 6, 813–839. MR2737760 (2011k:49099)
14. R.B. Davies, *Numerical inversion of a characteristic function*, Biometrika **60** (1973), 415–417. MR0321152 (47:9685)

15. _____, *Statistical algorithms: Algorithm AS 155: The distribution of a linear combination of χ^2 random variables*, Applied Statistics **29** (1980), no. 3, 323–333.

16. M. Engliš and J. Peetre, *Green's functions for powers of the invariant Laplacian*, Canad. J. Math. **50** (1998), no. 1, 40–73. MR1618718 (99c:58146)

17. K.-T. Fang and Y. Wang, *Number-theoretic methods in statistics*, Monographs on Statistics and Applied Probability, vol. 51, Chapman & Hall, London, 1994. MR1284470 (95g:65189)

18. N.I. Fisher, T. Lewis, and B.J.J. Embleton, *Statistical analysis of spherical data*, Cambridge University Press, Cambridge, 1993. MR1247695 (94h:62001)

19. J. Fliege and U. Maier, *The distribution of points on the sphere and corresponding cubature formulae*, IMA J. Numer. Anal. **19** (1999), no. 2, 317–334. MR1686006

20. W. Freeden and M. Schreiner, *Spherical functions of mathematical geosciences. A scalar, vectorial, and tensorial setup*, Advances in Geophysical and Environmental Mechanics and Mathematics. Berlin: Springer. xv, 602 p., 2009.

21. E. Giné, *Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms*, Ann. Statist. **3** (1975), no. 6, 1243–1266. MR0388663 (52:9499)

22. I.S. Gradshteyn and I.M. Ryzhik, *Table of integrals, series, and products*, fifth ed., Academic Press Inc., San Diego, CA, 1996, CD-ROM version 1.0 for PC, MAC, and UNIX computers. MR1398882 (97c:00014)

23. G.G. Gregory, *Large sample theory for U-statistics and tests of fit*, Ann. Statist. **5** (1977), no. 1, 110–123. MR0433669 (55:6642)

24. J. Gustavsson, *Some sums of Legendre and Jacobi polynomials*, Math. Bohem. **126** (2001), no. 1, 141–149. MR1826477 (2002g:33011)

25. F.J. Hickernell, *A generalized discrepancy and quadrature error bound*, Math. Comp. **67** (1998), no. 221, 299–322. MR1433265 (98c:65032)

26. R.J. Hinde, *Constructing atom-molecule potential surfaces from ab initio data: a method combining quadrature and interpolation*, Comput. Phys. Commun. **130** (2000), no. 1, 1–11.

27. J.P. Imhof, *Computing the distribution of quadratic forms in normal variables*, Biometrika **48** (1961), 419–426. MR0137199 (25:655)

28. W.B. Jordan, *A sum of Legendre polynomials: Problem 83-20 (in Solutions)*, SIAM Review **26** (1984), no. 4, 586–587.

29. A. Katanforoush and M. Shahshahani, *Distributing points on the sphere. I*, Experiment. Math. **12** (2003), no. 2, 199–209. MR2016706 (2004m:52056)

30. A.B.J. Kuijlaars and E.B. Saff, *Asymptotics for minimal discrete energy on the sphere*, Trans. Amer. Math. Soc. **350** (1998), no. 2, 523–538. MR1458327 (98e:11092)

31. H. Leeb, *Asymptotic properties of the spectral test, diaphony, and related quantities*, Math. Comp. **71** (2002), no. 237, 297–309 (electronic). MR1863001 (2002j:11086)

32. H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. MR1172997 (93h:65008)

33. M.J. Prentice, *On invariant tests of uniformity for directions and orientations*, Ann. Statist. **6** (1978), no. 1, 169–176. MR0458721 (56:16921)

34. _____, *Correction to: "On invariant tests of uniformity for directions and orientations" [Ann. Statist. 6 (1978), no. 1, 169–176;]*, Ann. Statist. **7** (1979), no. 4, 926. MR532258 (80k:62078)

35. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical recipes in C++*, Cambridge University Press, Cambridge, 2002. MR1880993 (2003a:65005)

36. J.-R. Pycke, *A decomposition for invariant tests of uniformity on the sphere*, Proc. Amer. Math. Soc. **135** (2007), no. 9, 2983–2993 (electronic). MR2317977 (2008i:62058)

37. _____, *U-statistics based on the Green's function of the Laplacian on the circle and the sphere*, Statist. Probab. Lett. **77** (2007), no. 9, 863–872. MR2363436 (2009b:62094)

38. S.O. Rice, *Distribution of quadratic forms in normal random variables—evaluation by numerical integration*, SIAM J. Sci. Statist. Comput. **1** (1980), no. 4, 438–448. MR610756 (82g:62037)

39. M. Schreiner, *On a new condition for strictly positive definite functions on spheres*, Proc. Amer. Math. Soc. **125** (1997), no. 2, 531–539. MR1353398 (97d:43005)

40. R.J. Serfling, *Approximation theorems of mathematical statistics*, John Wiley & Sons Inc., New York, 1980. MR595165 (82a:62003)

41. J. Sheil and I. O'Muircheartaigh, *Statistical algorithms: Algorithm AS 106: The distribution of non-negative quadratic forms in normal variables*, Applied Statistics **26** (1977), no. 1, 92–98.

42. G.R. Shorack, *Probability for statisticians*, Springer Texts in Statistics, Springer-Verlag, New York, 2000. MR1762415 (2001d:60002)

43. I.H. Sloan and R.S. Womersley, *Extremal systems of points and numerical integration on the sphere*, Adv. Comput. Math. **21** (2004), no. 1-2, 107–125. MR2065291 (2005b:65024)

44. A.W. van der Vaart, *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, Cambridge, 1998. MR1652247 (2000c:62003)

45. T.T. Wong, W.S. Luk, and P.A. Heng, *Sampling with Hammersley and Halton points*, Journal of Graphics Tools **2** (1997), no. 2, 9–24.

46. J.-T. Zhang, *Approximate and asymptotic distributions of chi-squared-type mixtures with applications*, J. Amer. Statist. Assoc. **100** (2005), no. 469, 273–285. MR2156837 (2006d:62050)

DEPARTMENT OF ECONOMICS, SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION, UNIVERSIDAD DE NAVARRA, EDIFICIO AMIGOS, 31080 PAMPLONA, SPAIN
   E-mail address: `cchoirat@unav.es`

DIPARTIMENTO DI ECONOMIA, UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA, VIA MONTE GENEROSO 71, 21100 VARESE, ITALY
   E-mail address: `raffaello.seri@uninsubria.it`