



HAL
open science

–statistics based on the Green’s function of the Laplacian on the circle and the sphere

J.-R. Pycke

► **To cite this version:**

J.-R. Pycke. –statistics based on the Green’s function of the Laplacian on the circle and the sphere. Statistics and Probability Letters, 2009, 77 (9), pp.863. 10.1016/j.spl.2006.11.009 . hal-00547982

HAL Id: hal-00547982

<https://hal.science/hal-00547982>

Submitted on 18 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

U -statistics based on the Green's function of the Laplacian on the circle and the sphere

J.-R. Pycke

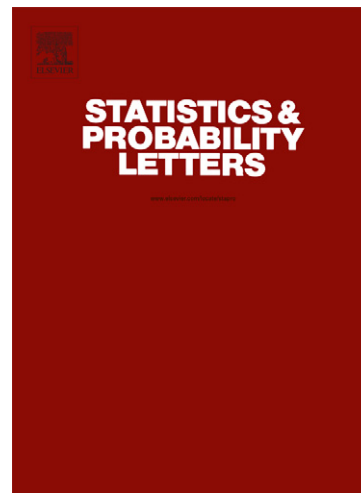
PII: S0167-7152(06)00344-0
DOI: doi:10.1016/j.spl.2006.11.009
Reference: STAPRO 4535

To appear in: *Statistics & Probability Letters*

Received date: 23 November 2004
Revised date: 16 October 2006
Accepted date: 26 November 2006

Cite this article as: J.-R. Pycke, U -statistics based on the Green's function of the Laplacian on the circle and the sphere, *Statistics & Probability Letters*, doi:10.1016/j.spl.2006.11.009

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

U –statistics based on the Green’s function of the Laplacian on the circle and the sphere

J.-R. Pycke*

October 16, 2006

*Université d’Évry, département de Mathématiques, Boulevard F. Mitterrand, 91 025 Évry cedex, France ¹

Abstract

We show that the Watson and Cramér-von Mises statistics are related to Green’s function of the Laplacian on a circle. A generalization leads to a new U-statistic whose kernel is the Green function of the Laplacian on the sphere.

Keywords and phrases: directional statistics, tests of uniformity, U- and V-statistics, Cramér-von Mises statistic, Anderson-Darling statistic, Watson statistic.

1 Introduction

In the field of directional statistics the problem of testing uniformity remains widely open, especially when the sample space is the unit sphere. The present paper deals with the case where the sample space is the unit circle or the unit sphere. For general surveys about tests for uniformity on the circle and the sphere the reader is referred to Fisher (1993) p.64–71 and Mardia and Jupp (2000), §6.3 and §10.4.1. Many tests for uniformity on the circle and the sphere fit into the general framework established by Giné (1975), whose most important features are expounded in Mardia and Jupp (2000), §10.8. In the present paper we propose a new approach based upon the following remarks. The problem of testing uniformity on the unit circle is closely related to that of testing uniformity on $[0, 1]$. A classical way to test the hypothesis

H_0 : a random sample x_1, \dots, x_n (with order statistics $x_{(1)} < \dots < x_{(n)}$) has been drawn from a population uniformly distributed on $[0, 1]$,

¹E-mail address : jrpycke@univ-evry.fr

is based on the rejection of H_0 for large values taken by one of the celebrated statistics

$$(1.1) \quad W_n^2 = \sum_{i=1}^n \left(x_{(i)} - \frac{i - \frac{1}{2}}{n} \right)^2 + \frac{1}{12n}, \quad (\text{Cramér-von Mises})$$

$$(1.2) \quad U_n^2 = W_n^2 - n(\bar{x} - \frac{1}{2})^2, \quad (\text{Watson})$$

$$(1.3) \quad A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) \{ \log x_{(i)} + \log[1 - x_{(n-i+1)}] \}, \quad (\text{Anderson-Darling})$$

where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ (see Durbin (1973), formulas (4.17) – (4.18) p. 27 and (5.4.2) p. 36). Elementary computations enable to express these Cramér-von Mises type statistics in the alternative form of the von Mises functionals (or V-statistics)

$$(1.4) \quad U_n^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{(|x_i - x_j| - 1/2)^2}{2} - \frac{1}{24} \right\} \quad \text{with the kernel}$$

$$(1.5) \quad K_U(x_1, x_2) := \frac{(|x_1 - x_2| - 1/2)^2}{2} - \frac{1}{24} \\ = \sum_{k=1}^{\infty} \frac{2 \sin(2k\pi x_1) \sin(2k\pi x_2) + 2 \cos(2k\pi x_1) \cos(2k\pi x_2)}{4k^2\pi^2};$$

$$(1.6) \quad W_n^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{x_i^2 - x_i + x_j^2 - x_j - |x_i - x_j|}{2} + \frac{1}{3} \right\} \quad \text{with}$$

$$(1.7) \quad K_W(x_1, x_2) := \frac{x_1^2 - x_1 + x_2^2 - x_2 - |x_1 - x_2|}{2} + \frac{1}{3} = \sum_{k=1}^{\infty} \frac{2 \cos(k\pi x_1) \cos(k\pi x_2)}{k^2\pi^2};$$

$$(1.8) \quad A_n^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{ -\log[\max(x_i, x_j) - x_i x_j] - 1 \}.$$

For basic definitions and results about U - and V - statistics, see Koroljuk and Borovskich (1994), Chapter 1. The explicit *Kac-Siebert* or *Karhunen-Loève expansions* (1.5) and (1.7) of the Watson and Cramér-von Mises kernels (K_U and K_W respectively) are given in Durbin (1973), formula (5.6.7) p. 38 and Dym and McKean (1972) p. 60. The definition and usefulness of such an expansion of the kernel, in the study of the corresponding Cramér-von Mises, U - or V - statistic is well-known. Some basic facts will be shortly recalled in Section 3. For details, the reader is referred to Shorack and Wellner (1986), Chapter 5 (for Cramér-von Mises type statistics) and Koroljuk and Borovskich (1994), § 4.3. (for U - and V - statistics). Unfortunately, such explicit expansions can rarely be derived. However, recent advances (see Deheuvels and Martynov (2003), Henze and Nikitin (2002), Lachal (2001) and Pycke (2003)) show a renewal of the interest in this field, in which a central problem is to find a general method for deriving explicit expansions having a statistical interest. The aim of this paper is to show how the three classical examples mentioned above can be generalized. It leads to a new family of statistics since to our knowledge, the asymptotic distribution given by (4.39) in Proposition 4.2 is not that of an already known test for uniformity on the sphere.

To start with, relations (3.14) of Proposition 3.2 and (3.22) of Proposition 3.3 show that K_U and K_W are simply related to the zero-mean Green function of the Laplacian on the circle, say G_1 . In Section 2, we recall the definition of the Green's function of the Laplacian on a compact manifold M , and give in Proposition 2.1 a Karhunen-Loève expansion of this function, in the particular case where $\dim M \leq 3$. In Propositions 3.2 and 3.3, we show that the general way to cover the circle by an interval leads to new V-statistics whose kernel, defined by (3.15) and (3.24), are as well simply related to G_1 . In Section 4 we generalize these ideas to the sphere. This interpretation enables us to introduce a new U-statistic, defined in (4.39), arising as a generalization of the Watson statistic in order to test uniformity on the sphere. Its kernel is the centered Green function of the Laplacian on the sphere, for which we obtain different explicit orthogonal decompositions in Proposition 4.1. The asymptotic distribution of a subset of principal components of the new statistic (in the line of Durbin and Knott (1972)) is given in Propositions 4.3 and 4.4. In particular, Proposition 4.3 shows that the Anderson-Darling statistic can be viewed as the radial principal component, with respect to a pole, of our new statistic. Generalizations of these ideas to a wider class of compact manifolds will be studied in a forthcoming paper.

Before stating our results, we recall some basic facts about Green's function and Karhunen-Loève expansions.

2 Green's function and Karhunen-Loève expansions

For details concerning the Green's function of the Laplacian, the reader is referred to Aubin (1982), Chapter 4, particularly § 2.3 p. 108.

Assume M is a compact and connected, C^∞ , Riemannian manifold without boundary having volume V . The density of the Riemannian measure on M (see Chavel (1984), Chapter 1, § 2) is denoted by dQ . We let $L^2(M)$ be the space of real measurable functions f on M for which $\int_M f^2(Q)dQ < \infty$, equipped with the usual inner product

$$(f|g) = \int_M f(Q)g(Q)dQ, \quad f, g \in L^2(M).$$

For any C^k , $k \geq 2$, function f , $\Delta f := \operatorname{div}(\operatorname{grad} f)$ denotes the Laplacian of f . A *Green's function of the Laplacian* is a function $G(P, Q)$ defined on $M \times M$ satisfying, in the distributional sense,

$$-\Delta_Q G(P, Q) = \delta_P(Q) - 1/V$$

where δ_P denotes the Dirac distribution at P (see formula (14) p.108 in Aubin (1982), keeping in mind that the author has an opposite sign convention for Δ). From Aubin (1982), Theorem 4.13 p.108, we know the following properties. On the one hand there always exists a Green function of the Laplacian on M . On the other hand there is only one Green's function of the Laplacian, say G_M , satisfying moreover

$$(2.9) \quad \int_M G_M(P, Q)dQ = 0, \quad \forall P \in M.$$

This function will be referred to in the sequel as the *zero-mean Green's function of the Laplacian*. Consider the integral equation

$$(2.10) \quad \lambda \int_M K(P, Q)\phi(Q)dQ = \phi(P)$$

where the symmetric, measurable kernel K satisfies

$$\int_{M \times M} K^2(P, Q)dPdQ < \infty.$$

According to the general theory of such integral equations, (2.10) has a denumerable set of eigenvalues λ_k , $k = 1, 2, \dots$, which are matched by a denumerable set of eigenfunctions ϕ_k , $k = 1, 2, \dots$, forming an orthonormal set of functions in $L^2(M)$. The Fourier series of the kernel $K(P, Q)$ regarded as a function of Q , with respect to the orthonormal system $\{\phi_k(Q)\}$ is

$$(2.11) \quad K(P, Q) = \sum_{k=1}^{\infty} \frac{\phi_k(P)\phi_k(Q)}{\lambda_k}, \quad \text{with} \quad \int_M \phi_k(Q)\phi_\ell(Q)dQ = \begin{cases} 1 & \text{if } k = \ell, \\ 0 & \text{if } k \neq \ell. \end{cases}$$

This series converge in the mean to K . If the integral operator is positive, the eigenvalues can be rearranged to satisfy $0 < \lambda_1 \leq \lambda_2 \leq \dots$ and we refer to (2.11) as to the Karhunen-Loève (or K-L) expansion of K .

Proposition 2.1. *The eigenvalue problem*

$$\Delta\varphi = -\lambda\varphi$$

has a complete orthonormal system of C^∞ -eigenfunctions $\varphi_0, \varphi_1, \varphi_2, \dots$ in $L^2(M)$ and corresponding eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \uparrow \infty$. Furthermore, if $\dim M \leq 3$, one has in $L^2(M \times M)$ the bilinear expansion of the zero-mean Green function

$$(2.12) \quad G_M(P_1, P_2) = \sum_{k=1}^{\infty} \frac{\varphi_k(P_1)\varphi_k(P_2)}{\lambda_k}.$$

Proof. The first assertion is a classical result from Riemannian geometry, see e.g. Chavel (1984), Theorem 1 p. 8. It is clear that $\phi_0(Q) = 1/\sqrt{V}$; consequently the orthogonality relations imply that for each $k \geq 1$, $\int_M \phi_k = \sqrt{V} \int_M \phi_k \phi_0 = 0$. Hence from formula (15) p. 108 in Aubin (1982), we infer

$$\lambda_k \int_M G_M(P, Q)\phi_k(Q)dQ = \phi_k(P), \quad (k \geq 1).$$

In view of $\lim_{k \rightarrow \infty} \lambda_k^{-1} = 0$, we know that $\phi \mapsto \int_M G_M(\cdot, Q)\phi(Q)dQ$ is a compact operator on $L^2(M)$ (see e.g. Proposition 1.3.10 in Zhu (1990)). Furthermore we know from Weyl's asymptotic formula (see Chavel (1984), formula (50) p. 9), that

$$\lambda_k \sim \text{constant} \times k^{2/\dim M}, \quad \text{hence} \quad \sum_{k=1}^{\infty} \lambda_k^{-2} < \infty \iff \dim M \in \{1, 2, 3\}.$$

This means that in our case, $\phi \mapsto \int_M G_M(P, Q)\phi(Q)dQ$ is a Hilbert-Schmidt operator (see Zhu (1990), §1.4). It also implies that if $\dim M \leq 3$, the series on the right-hand side of (2.12) converges in $L^2(M \times M)$ toward a function, say $L(P, Q)$. This function is the kernel of the integral operator $\phi \mapsto \int_M L(P, Q)\phi(Q)dQ$ on $L^2(M)$. In order to prove that it coincides with $G_M(P, Q)$, it is sufficient to check it on their action on φ_k , $k \geq 0$, since the latter span $L^2(M)$. For $k = 0$, the result is a consequence of (2.9). And for each $k \in \mathbb{N}^*$, it is readily checked that

$$\int_M L(P, Q)\varphi_k(Q)dQ = \varphi_k(P)/\lambda_k,$$

which completes the proof. \square

3 The Watson, Cramér-von Mises statistics and the circle

U_n^2 was introduced by Watson (1961) for use with observations P_1, \dots, P_n recorded on the circumference of the circle. For the circle of radius R

$$S^1(R) := \{Re^{i\theta} : \theta \in \mathbb{R}\},$$

endowed with the measure $dP = Rd\theta$, a point $P \in S^1(\mathbb{R})$ corresponding to the argument θ will be denoted by $P(\theta)$ and we write $\theta = \arg P$. The principal value of the argument of P , denoted by $\text{Arg } P$, is the argument satisfying $-\pi < \theta \leq \pi$. The north pole corresponds to $\theta = 0$, the south pole to $\theta = \pi$. We denote by $d(P_1, P_2)$ the distance on the circle between the two points $P_1(\theta_1)$ and $P_2(\theta_2)$. The Laplacian of $P \mapsto f(\theta)$ where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a 2π -periodic function is given by

$$\Delta f = \frac{1}{R^2} \frac{d^2 f}{d\theta^2}.$$

Proposition 3.1. *The zero-mean Green's function of the Laplacian on $S^1(R)$ is*

$$G_1(P_1, P_2) := \frac{[\pi R - d(P_1, P_2)]^2}{4\pi R} - \frac{\pi R}{12}.$$

Moreover, one has the pointwise converging bilinear expansion

$$(3.13) \quad G_1(P_1, P_2) = \sum_{k=1}^{\infty} \frac{s_k(P_1)s_k(P_2) + c_k(P_1)c_k(P_2)}{(k^2/R^2)} = \sum_{k=1}^{\infty} \frac{c_k[d(P_1, P_2)/R]}{(k^2/R^2)}$$

with $s_k(P) = (\pi R)^{-1/2} \sin(k\theta)$ and $c_k(\theta) = (\pi R)^{-1/2} \cos(k\theta)$.

Proof. It is clear that $\{s_k, c_k : k \geq 1\}$ is a complete orthonormal set of zero-mean eigenfunctions of the Laplacian. The first equality in (3.13) then follows from the equality (2.12) in Proposition 2.1. The second equality is a consequence of the identity $\cos(\pm|\theta_1 - \theta_2|) = \cos \theta_1 \cos \theta_2 \pm \sin \theta_1 \sin \theta_2$, in view of the equality $d(P_1, P_2)/R = \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|)$. \square

A way to wrap a distribution on $[0, 1]$ around the circle is to set

$$\arg P := 2q\pi x + c, \quad x \in [0, 1],$$

with $c \in \mathbb{R}$ and $q \in \mathbb{N}^*$. The next Proposition shows that for $q = 1$, this wrapping transforms Watson's kernel, defined on $[0, 1] \times [0, 1]$, into the Green's function of the circle, and generalizes Watson's kernel for other values of q .

Proposition 3.2. *Assume $x_1, x_2 \in [0, 1]$ and $c \in \mathbb{R}$. If $\arg P_i := 2\pi x_i + c$, $i = 1, 2$, then*

$$(3.14) \quad K_U(x_1, x_2) = G_1(P_1, P_2)/(2\pi R).$$

More generally, if $\arg P_i := 2q\pi x_i + c$, $i = 1, 2$, with $q \in \mathbb{N}^*$, then one has

$$(3.15) \quad \frac{1}{2\pi R} G_1 [P_1(2\pi q x_1 + c), P_2(2\pi q x_2 + c)] \\ = \frac{(|qx_1 - [qx_1] - qx_2 + [qx_2]| - \frac{1}{2})^2}{2} - \frac{1}{24} =: K_q(x_1, x_2)$$

and the pointwise converging bilinear expansion

$$(3.16) \quad K_q(x_1, x_2) = 2 \sum_{k \in \mathbb{N}^*} \frac{\cos(2kq\pi x_1) \cos(2kq\pi x_2) + \sin(2kq\pi x_1) \sin(2kq\pi x_2)}{4k^2\pi^2}$$

Proof. For $x \in \mathbb{R}$, we set $\{x\} = x - [x]$. Obviously, $\{x\} = x$ for $x \in [0, 1)$. For $q \in \mathbb{N}^*$ and $\arg P_i := 2q\pi x_i + c$, $i = 1, 2$, one has

$$(3.17) \quad d(P_1, P_2) = 2\pi R \min(|\{qx_1\} - \{qx_2\}|, 1 - |\{qx_1\} - \{qx_2\}|).$$

Thus we have $d(P_1, P_2) - \pi R = \pm 2\pi R(|\{qx_1\} - \{qx_2\}| - 1/2)$, which implies (3.14) (for $q = 1$) and the first equality in (3.15). The bilinear expansion (3.16) follows readily from (3.13), (3.17) and the identity

$$\cos[2k\pi \min(|\{qx_1\} - \{qx_2\}|, 1 - |\{qx_1\} - \{qx_2\}|)] = \cos(2kq\pi[x_1 - x_2]).$$

Since K_q is continuous on $[0, 1] \times [0, 1]$, the pointwise convergence is a consequence of Mercer's theorem. \square

Corollary 3.1. *For each $q \in \mathbb{N}^*$, the V -statistic*

$$(3.18) \quad V_{q,n}^2 := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K_q(x_i, x_j)$$

satisfies, under H_0 ,

$$(3.19) \quad \lim_{n \rightarrow \infty} P(nV_{q,n}^2 > x) = \sum_{j \in \mathbb{N}^*} (-1)^{j-1} 2e^{-2j^2\pi^2 x}.$$

Proof. The eigenvalues appearing in the Karhunen-Loève expansion (3.16) are independent of q . Consequently (see Koroljuk and Borovskich (1994), formula (4.3.18) p.142), for each q , the asymptotic distribution of $nV_{q,n}^2$ is that of $nV_{1,n}^2$ which is the Watson statistic. The asymptotic distribution of the latter is given in Watson (1961), formula (22). \square

We now show that the Cramér-von Mises statistic is obtained as the radial part, with respect to the north pole, of a test of uniformity on the circle. Let $s : P \mapsto s \cdot P$ with $\text{Arg}(s \cdot P) = -\text{Arg } P$ denote the reflection in the axis passing through the poles. Clearly,

$$(3.20) \quad c_k(s \cdot P) = c_k(P), \quad \text{whereas } s_k(s \cdot P) = -s_k(P).$$

Suppose we have a sample $P_1(\theta_1), \dots, P_n(\theta_n)$ on $S^1(R)$. A natural way to build a V -statistic depending only on $d(P_i, N) = |\theta_i|$, $i = 1, \dots, n$, is, in view of (3.20), to use the kernel

$$(3.21) \quad \frac{G(P_1, P_2) + G(P_1, s \cdot P_2)}{2} = \sum_{k=1}^{\infty} \frac{c_k(P_1)c_k(P_2)}{(k^2/R^2)},$$

whose bilinear expansion has been obtained on combining (3.13) and (3.20). We now show that the kernel obtained from G_1 on $S^1(R) \times S^1(R)$ by this averaging process corresponds to K_W on $[0, 1] \times [0, 1]$ after a suitable change of variables. For $x \in \mathbb{R}$ we use the notation

$$d_{2\mathbb{Z}}(x) := \min_{\ell \in \mathbb{Z}} |x - 2\ell|.$$

Proposition 3.3. *If $P_1, P_2 \in S^1(R)$, then*

$$(3.22) \quad \frac{G(P_1, P_2) + G(P_1, s \cdot P_2)}{2} = \frac{\pi R}{2} K_W\left(\frac{d(P_1, N)}{\pi R}; \frac{d(P_2, N)}{\pi R}\right).$$

Conversely, if $x_1, x_2 \in [0, 1]$ and $\arg P_i := \pi x_i$, ($i = 1, 2$), one has

$$(3.23) \quad K_W(x_1, x_2) = \frac{2}{\pi R} \cdot \frac{G_1(P_1, P_2) + G_1(P_1, s \cdot P_2)}{2}$$

More generally, assume $q \in \mathbb{N}^$, $x_1, x_2 \in (0, 1)$ and*

$$\arg P_i := q\pi x_i, \quad (i = 1, 2)$$

Then the kernel

$$(3.24) \quad L_q(x_1, x_2) := \frac{G_1(P_1, P_2) + G_1(P_1, s \cdot P_2)}{2} = \frac{\pi R}{2} K_W [d_{2\mathbb{Z}}(qx_1), d_{2\mathbb{Z}}(qx_2)]$$

has the bilinear expansion

$$(3.25) \quad L_q(x_1, x_2) = 2 \sum_{k \in \mathbb{N}^*} \frac{\cos(kq\pi x_1) \cos(kq\pi x_2)}{k^2 \pi^2}$$

Proof. The bilinear expansion (3.21), with $\arg P_i := q\pi x_i$, becomes

$$(3.26) \quad \frac{G(P_1, P_2) + G(P_1, s \cdot P_2)}{2} = \sum_{k=1}^{\infty} \frac{\cos[kq\pi x_1] \cos[kq\pi x_2]}{(k^2\pi/R)} \\ = \sum_{k=1}^{\infty} \frac{\cos[k\pi d_{2\mathbb{Z}}(qx_1)] \cos[k\pi d_{2\mathbb{Z}}(qx_2)]}{(k^2\pi/R)} = \frac{\pi R}{2} K_W[d_{2\mathbb{Z}}(qx_1), d_{2\mathbb{Z}}(qx_2)],$$

the last equality following from (1.7) and the fact that $d_{2\mathbb{Z}}(qx) \in [0, 1]$ for each $x \in \mathbb{R}$. Hence (3.24) and (3.25) are proved. By taking $q = 1$ we obtain (3.22) and (3.23) in view of $d(P_i, N)/(\pi R) = x_i$, ($i = 1, 2$). \square

4 Tests of uniformity on the sphere

Suppose now that wish to test the hypothesis

H_0 : a random sample P_1, \dots, P_n has been drawn from a population uniformly distributed on the unit sphere

$$S^2 := \{(x, y, z) \in \mathbb{R}^3, x^2 + y^2 + z^2 = 1\}.$$

On S^2 the generic point $Q(x, y, z)$ has spherical coordinates $(\theta, \phi) \in [0, \pi] \times [0, 2\pi]$ with

$$(4.27) \quad x = \sin \theta \sin \phi, \quad y = \sin \theta \cos \phi, \quad z = \cos \theta.$$

The Riemannian measure is

$$dQ = \sin \theta d\theta d\phi.$$

For the following basic facts, see e.g. Magnus *et al.* (1966), § 4.9. The Laplacian of $f : P \mapsto f(\theta, \phi)$ is given by

$$\Delta f = (\sin \theta)^{-1} \partial_\theta (\sin \theta \partial_\theta f) + (\sin \theta)^{-2} \partial_\phi^2 f.$$

For each $\ell \in \mathbb{N}$, $-\ell(\ell+1)$ is an eigenvalue of the Laplacian with multiplicity $2\ell+1$. The Legendre polynomials and associated functions of the first kind are defined by

$$P_\ell(x) = (2^\ell \ell!)^{-1} \frac{d^\ell}{dx^\ell} (x^2 - 1)^\ell, \quad (\ell \in \mathbb{N}) \text{ and}$$

$$P_\ell^m(x) = (-1)^m (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_\ell(x), \quad (1 \leq m \leq \ell, -1 \leq x \leq 1),$$

(Magnus *et al.* (1966) p.174 and p. 232). An orthonormal basis of the eigenspace corresponding to $-\ell(\ell+1)$ is given by the $2\ell+1$ functions $\{f_\ell^m : -\ell \leq m \leq \ell\}$ defined by

$$(4.28) \quad f_\ell^0(P) := \left(\frac{2\ell+1}{4\pi}\right)^{1/2} P_\ell(\cos \theta), \quad (\ell \geq 0),$$

$$(4.29) \quad f_\ell^m(P) := \left\{ \frac{(2k+1)(\ell-m)!}{2\pi(\ell+m)!} \right\}^{1/2} \cos(m\phi) P_\ell^m(\cos \theta), \quad (1 \leq m \leq \ell),$$

$$(4.30) \quad f_\ell^m(P) := \left\{ \frac{(2k+1)(\ell-|m|)!}{2\pi(\ell+|m|)!} \right\}^{1/2} \sin(|m|\phi) P_\ell^{|m|}(\cos \theta), \quad (-\ell \leq m \leq -1).$$

We set, for $0 < \theta_1 \leq \theta_2 < \pi$,

$$(4.31) \quad \Phi_0(P_1, P_2) := \frac{2 \log 2 - 1 - \log(1 + \cos \theta_1)(1 - \cos \theta_2)}{4\pi},$$

$$(4.32) \quad \Phi_m(P_1, P_2) := \frac{\sin(m\phi_1) \sin(m\phi_2)}{2\pi|m|} \tan^{|m|} \frac{\theta_1}{2} \cot^{|m|} \frac{\theta_2}{2}, \quad (m \in -\mathbb{N}^*)$$

$$(4.33) \quad \Phi_m(P_1, P_2) := \frac{\cos(m\phi_1) \cos(m\phi_2)}{2\pi m} \tan^m \frac{\theta_1}{2} \cot^m \frac{\theta_2}{2}, \quad (m \in \mathbb{N}^*).$$

In the following Proposition we compute the Green's function of the Laplacian, say G_2 , for which we give several orthogonal decomposition. We will use Lemma 4.1, whose proof is postponed to the end of this section. For any $P \in S^2$, \vec{P} will denote the unit vector \overrightarrow{OP} , where O is the center of S^2 .

Proposition 4.1. *The zero-mean Green's function of the Laplacian on S^2 is*

$$(4.34) \quad G_2(P_1, P_2) = \frac{\log 2 - 1 - \log(1 - \vec{P}_1 \cdot \vec{P}_2)}{4\pi}.$$

G_2 has in L^2 the bilinear expansions

$$(4.35) \quad \begin{aligned} G_2(P_1, P_2) &= \sum_{\ell=1}^{\infty} \frac{f_{\ell}^0(P_1) f_{\ell}^0(P_2)}{\ell(\ell+1)} + \sum_{m \in \mathbb{Z}^*} \sum_{\ell=|m|}^{\infty} \frac{f_{\ell}^m(P_1) f_{\ell}^m(P_2)}{\ell(\ell+1)} \\ &= \sum_{m \in \mathbb{Z}} \Phi_m(P_1, P_2) \end{aligned}$$

where the Φ_m have the pointwise converging bilinear expansions,

$$(4.36) \quad \Phi_0(P_1, P_2) = \sum_{\ell=1}^{\infty} \frac{f_{\ell}^0(P_1) f_{\ell}^0(P_2)}{\ell(\ell+1)},$$

$$(4.37) \quad \Phi_m(P_1, P_2) = \sum_{\ell=|m|}^{\infty} \frac{f_{\ell}^m(P_1) f_{\ell}^m(P_2)}{\ell(\ell+1)}, \quad (m \in \mathbb{Z}^*).$$

Moreover one has, for each $q \in \mathbb{N}^*$ and $0 < \theta_1 \leq \theta_2$,

$$(4.38) \quad \sum_{m \in q\mathbb{Z}^*} \Phi_m(P_1, P_2) = -\frac{1}{4\pi} \log \left\{ 1 - 2 \tan \frac{\theta_1}{2} \cot \frac{\theta_2}{2} \cos[q(\phi_1 - \phi_2)] + \tan^2 \frac{\theta_1}{2} \cot^2 \frac{\theta_2}{2} \right\}.$$

Proof. The first equality in (4.35) follows from Proposition 2.1 and the definition of the functions appearing in the series. For (4.36), see the last formula of § 5.4.4 in Magnus *et al.* (1966) p.239, with $x = \cos \theta_2$ and $y = \cos \theta_1$. The equality (4.37) is a consequence of Theorem 2.1 and relation (2.7) p. 390 in Pycke (2003), with $\mu = m$, $\min(s, t) = (1 - \cos \theta_1)/2$, $\max(s, t) = (1 - \cos \theta_2)/2$ and the change of index $j = k - m + 1$. (4.38). It implies in turn the second equality in (4.35). Relation (4.38) is a direct consequence of Lemma 4.1, used with $\phi = q\phi_1$, $\phi' = q\phi_2$. Finally (4.38), for $q = 1$, when combined with (4.44) and (4.31), leads to (4.34). \square

We now state some statistical applications of the orthogonal decomposition of G_2 obtained in the preceding Proposition. They are consequences of the general result about U -statistics stated for example in Koroljuk and Borovskich (1994), Theorem 4.3.1. Note that the degeneracy condition used in this theorem is satisfied in our case, since it corresponds to (2.9). Thus for each $\ell \in \mathbb{N}^*$, one has the convergence in distribution

$$T_\ell^m(n) := n^{-1/2} \sum_{i=1}^n f_\ell^m(\xi_i) \rightarrow \tau_\ell^m, \quad (-\ell \leq m \leq \ell) \text{ as } n \rightarrow \infty,$$

where the (τ_ℓ^m) are independent standard Gaussian variables. Hence we can state the following result.

Proposition 4.2. *Under the null hypothesis of uniformity*

$$(4.39) \quad U_{G_2}(n)(P_1, \dots, P_n) := \frac{2}{(n-1)} \sum_{1 \leq i < j \leq n} G_2(P_i, P_j) \rightarrow \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} \frac{(\tau_\ell^m)^2 - 1}{\ell(\ell+1)} \text{ as } n \rightarrow \infty.$$

Each $T_\ell^m(n)$ is a principal component of $U_{G_2}(n)$. The following result permits to compute the distribution of sums of squares of these principal components regrouped in subsets whose choice follows naturally from the orthogonal decomposition (4.35). Consider the V -statistics

$$(4.40) \quad V_0^2(n)(P_1, \dots, P_n) := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Phi_0(P_i, P_j),$$

$$(4.41) \quad V_m^2(n)(P_1, \dots, P_n) := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [\Phi_m(P_i, P_j) + \Phi_{-m}(P_i, P_j)], \quad (m \in \mathbb{N}^*)$$

Since for any $\ell \geq 1$, $f_\ell^0(P)$ depends only on the distance (or latitude) of P to the north pole, $nV_0^2(n)$ can be viewed as the radial part (with respect to the north pole) of a test of uniformity on the sphere. The next proposition shows that more precisely, it is the Anderson-Darling statistic computed from the distances to the pole, hence the latitudes of the points.

Proposition 4.3. *If P_1, \dots, P_n are points from the sphere, one has*

$$nV_0^2(n)(P_1, \dots, P_n) = A_n^2 \left(\frac{1 - \cos \theta_1}{2}, \dots, \frac{1 - \cos \theta_n}{2} \right).$$

Proof. This result follows readily from (1.8), (4.28), (4.31) and (4.36). □

Next Proposition deals with the non null values of m .

Proposition 4.4. *Assume $m \in \mathbb{N}^*$. One has*

$$\lim_{n \rightarrow \infty} nV_m^2(n) = \sum_{\ell \geq |m|} \frac{(\tau_\ell^m)^2 + (\tau_\ell^{-m})^2}{\ell(\ell+1)} =: A_m^2.$$

The random variable A_m^2 has a characteristic function given by

$$(4.42) \quad \mathbf{E} \exp\{itA_m^2\} = \prod_{\ell=m}^{\infty} \left(1 - \frac{2it}{\ell(\ell+1)}\right)^{-1}, \quad (t \in \mathbb{R}).$$

Hence

$$(4.43) \quad \lim_{n \rightarrow \infty} P(nV_m^2(n) \geq x) = 1_{\{x>0\}} \sum_{\ell=m}^{\infty} (-1)^{\ell-m} (2\ell+1) a_{m,\ell} e^{-\ell(\ell+1)x/2}.$$

with

$$\begin{cases} a_{1,\ell} = 1, \\ a_{m,\ell} = \prod_{k=1}^{m-1} \left\{ \frac{\ell(\ell+1)}{k(k+1)} - 1 \right\}, \quad m = 2, 3, \dots \end{cases}$$

Proof. For each $t \in \mathbb{C}$, (see Durbin (1973), relation (4.4.7) p.32)

$$\phi_m(t) := \prod_{\ell=m}^{\infty} \left(1 - \frac{2t}{\ell(\ell+1)}\right)^{-1} = \frac{-2\pi t \prod_{k=1}^{m-1} \left(1 - \frac{2t}{k(k+1)}\right)}{\cos(\pi\sqrt{(1+8t)/2})} = \sum_{\ell=m}^{\infty} \frac{c_\ell}{1 - (2t)/[\ell(\ell+1)]}$$

with

$$c_\ell = \lim_{t \rightarrow \ell(\ell+1)/2} \left\{ 1 - \frac{2t}{\ell(\ell+1)} \right\} \phi_m(t) = (-1)^{\ell-m} (2\ell+1) \prod_{k=1}^{m-1} \left(\frac{\ell(\ell+1)}{k(k+1)} - 1 \right).$$

We obtained the last equality by setting $2t = \ell(\ell+1) + \varepsilon$, so that, as $t \rightarrow \ell(\ell+1)/2$,

$$\cos(\pi\sqrt{(1+8t)/2}) \sim (-1)^{\ell+1} \sin(\pi\varepsilon/(2\ell+1)), \quad 1 - \frac{2t}{\ell(\ell+1)} = -\frac{\varepsilon}{\ell(\ell+1)}.$$

Thus the density function of A_m^2 is

$$f(x) = \sum_{\ell=m}^{\infty} (-1)^{\ell-m} c_\ell \frac{\ell(\ell+1)}{2} \exp\left\{-\frac{\ell(\ell+1)t}{2}\right\}$$

from which (4.43) follows. □

Lemma 4.1. *If $0 \leq \theta \leq \theta' \leq \pi$, and $\phi, \phi' \in [0, 2\pi]$, then*

$$(4.44) \quad -\log\left(1 - 2 \tan \frac{\theta}{2} \cot \frac{\theta'}{2} \cos(\phi - \phi') + \tan^2 \frac{\theta}{2} \cot^2 \frac{\theta'}{2}\right) \\ = \sum_{m=1}^{\infty} \frac{2 \cos\{m(\phi - \phi')\}}{m} \tan^m \frac{\theta}{2} \cot^m \frac{\theta'}{2}$$

Furthermore

$$(4.45) \quad \log(1 - \vec{P}_1 \cdot \vec{P}_2) = \log([1 - \cos(\phi - \phi') \sin \theta \sin \theta' - \cos \theta \cos \theta']) \\ = \log\left(1 - 2 \tan \frac{\theta}{2} \cot \frac{\theta'}{2} \cos(\phi - \phi') + \tan^2 \frac{\theta}{2} \cot^2 \frac{\theta'}{2}\right) \\ + \log(1 + \cos \theta) + \log(1 - \cos \theta') - \log 2.$$

Proof. If $\theta < \theta'$, then $z := \tan \frac{\theta}{2} \cot \frac{\theta'}{2}$ satisfies $|z| < 1$. From the theory of orthogonal polynomials (see Szegő (1967), Chapter 4, formula 4.7.25), we know that

$$-\log(1 - 2z \cos(\phi - \phi') + z^2) = 2 \sum_{m=1}^{\infty} T_m [\cos(\phi - \phi')] z^m / m$$

where T_m is the Chebyshev polynomial, defined for $m \in \mathbb{N}$ by $T_m(\cos \cdot) = \cos(m \cdot)$. This leads to (4.44). In the case where $\theta = \theta'$, we have $\phi - \phi' \neq 0$ or 2π . The equality (4.44) then reduces to the classical Fourier series expansion (see Tolstov (1976), Chapter 5, section 11 p.147)

$$\sum_{m=1}^{\infty} \frac{2 \cos\{m(\phi - \phi')\}}{m} = -2 \log \left(2 \sin \frac{\phi - \phi'}{2} \right) = -\log [2 - 2 \cos(\phi - \phi')].$$

The first equality in (4.45) follows from the definition (4.27) of spherical coordinates. The second equality is a direct consequence of the identities

$$(1 + \cos \theta)(1 - \cos \theta') \left[1 + \tan^2 \frac{\theta}{2} \cot^2 \frac{\theta'}{2} \right] = 2 - 2 \cos \theta \cos \theta'$$

and

$$(1 + \cos \theta)(1 - \cos \theta') \cdot 2 \tan \frac{\theta}{2} \cot \frac{\theta'}{2} \cos(\phi - \phi') = 2 \sin \theta \sin \theta' \cos(\phi - \phi').$$

□

References

- Aubin, T. (1982). *Nonlinear analysis on manifolds. Monge-Ampère equations*, volume 252 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York.
- Chavel, I. (1984). *Eigenvalues in Riemannian geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL.
- Deheuvels, P. and Martynov, G. (2003). Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. In *High dimensional probability, III (Sandjberg, 2002)*, volume 55 of *Progr. Probab.*, pages 57–93. Birkhäuser, Basel.
- Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function*. Society for Industrial and Applied Mathematics, Philadelphia, Pa. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 9.
- Durbin, J. and Knott, M. (1972). Components of Cramér-von Mises statistics. I. *J. Roy. Statist. Soc. Ser. B*, **34**, 290–307.
- Dym, H. and McKean, H. P. (1972). *Fourier series and integrals*. Academic Press, New York. Probability and Mathematical Statistics, No. 14.

- Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.
- Giné, M. E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.*, **3**(6), 1243–1266.
- Henze, N. and Nikitin, Y. Y. (2002). Watson-type goodness-of-fit tests based on the integrated empirical process. *Math. Methods Statist.*, **11**(2), 183–202.
- Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U -statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.
- Lachal, A. (2001). Study of some new integrated statistics: computation of Bahadur efficiency, relation with non-standard boundary value problems. *Math. Methods Statist.*, **10**(1), 73–104.
- Magnus, W., Oberhettinger, F., and Soni, R. P. (1966). *Formulas and theorems for the special functions of mathematical physics*. Third enlarged edition. Die Grundlehren der mathematischen Wissenschaften, Band 52. Springer-Verlag New York, Inc., New York.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester. Revised reprint of *Statistics of directional data* by Mardia.
- Pycke, J.-R. (2003). Multivariate extensions of the Anderson-Darling process. *Statist. Probab. Lett.*, **63**(4), 387–399.
- Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.
- Szegő, G. (1967). *Orthogonal polynomials*. American Mathematical Society, Providence, R.I., third edition.
- Tolstov, G. P. (1976). *Fourier series*. Dover Publications Inc., New York.
- Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, **48**, 109–114.
- Zhu, K. H. (1990). *Operator theory in function spaces*, volume 139 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York.