Available online at www.sciencedirect.com

SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 341 (2005) 515–518

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistique

Des statistiques liées à la fonction de Green du laplacien

Jean-Renaud Pycke

LSTA, université Paris 6, case courrier 158, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 21 janvier 2005 ; accepté le 14 mars 2005

Présenté par Paul Deheuvels

Résumé

Dans cette Note, nous montrons qu'un test d'uniformité pour un échantillon de points d'un espace 2-point homogène peut être basé sur la fonction de Green du laplacien. Nous retrouvons les célèbres statistiques de Watson, Cramér–von Mises et Anderson–Darling comme des cas particuliers de cette classe de statistiques. ***Pour citer cet article : J.-R. Pycke, C. R. Acad. Sci. Paris, Ser. I 341 (2005).***

Abstract

Statistics related to Green's function of the Laplacian. In this Note, we show that an invariant test of uniformity for a sample from a compact 2-point homogeneous space can be based on the Green function of the Laplacian. The three celebrated Watson, Cramér–von-Mises and Anderson–Darling statistics are shown to be particular cases of this family of statistics. ***To cite this article: J.-R. Pycke, C. R. Acad. Sci. Paris, Ser. I 341 (2005).***

1. Introduction

Soit M un espace 2-point homogène connexe, compact, muni de sa mesure canonique de densité dP en $P \in M$, induisant une densité également notée d sur toute sous variété (voir [3], Chapitre II). Si l'on souhaite tester l'hypothèse

H_0 : l'échantillon $P_1, \ldots, P_n$ provient d'une population uniformément distribuée sur M,

il est naturel de chercher à construire une statistique qui soit, tout comme la distribution testée, invariante par le groupe des isométries de M (voir par exemple [7]). Dans cette Note, nous montrons que des U- ou V-statistiques de ce type peuvent être obtenues à partir de la fonction de Green du laplacien sur M. Nos principaux résultats sont basés sur le développement orthogonal du noyau de ces statistiques, qui implique les propriétés asymptotiques bien connues exposées dans [10], §4.3. Le développement bilinéaire donné dans le Théorème 2.2 généralise celui connu pour la statistique de Watson (cas $d = 1$). Sa composante radiale permet de construire un test introduit au Théorème 2.3 qui peut être interprété comme la composante principale radiale d'un test d'uniformité, généralisant les statistiques de Cramér–von Mises (cas $d = 1$) et d'Anderson–Darling ($d = 2$), dont nous donnons ainsi une interprétation géomé-

Adresse e-mail : pycke@ccr.jussieu.fr (J.-R. Pycke).

1631-073X/$ – see front matter © 2005 Académie des sciences. Publié par Elsevier SAS. Tous droits réservés.
doi:10.1016/j.crma.2005.03.017

Tableau 1
Notation

Notation	Espace	$\dim_{\mathbb{R}} M$
$M(d,d-1,\kappa)=\mathcal{S}^d(R)$	sphère euclidienne dans $\mathbb{R}^{d+1}$, de rayon $R=(2\sqrt{\kappa})^{-1}$	$d\in\mathbb{N}^*$
$M(d,0,\kappa)=P_\kappa^d(\mathbb{R})$	projectif réel de courbure constante κ	$d\in\mathbb{N}^*$
$M(d,1,\kappa)=P_\kappa^d(\mathbb{C})$	projectif complexe de courbure sectionnelle holomorphe 4κ	$d=4,6,8,\ldots$
$M(d,3,\kappa)=P_\kappa^d(\mathbb{H})$	projectif sur les quaternions de courbure sectionnelle maximale 4κ	$d=8,12,16,\ldots$
$M(16,7,\kappa)=P_\kappa^{16}(\mathbb{C}ay)$	plan projectif sur les octaves de Cayley de courbure sectionnelle maximale 4κ	$d=16$

trique, respectivement sur le cercle et sur la sphère. Les cas $d=2$ et 3, qui donnent de nouvelles statistiques, sont étudiés dans les exemples 2 et 3. En particulier, on y définit par (8) une statistique très simple pour un test d'uniformité sur la sphère S^2, qui généralise celle que Watson avait introduite pour le cercle S^1. Elle est basée sur la moyenne géométrique des espacements entre les points de l'échantillon.

Rappelons tout d'abord (voir [5] p. 78, [8] exemples 1 à 4 p. 340–341, [9] Chapitre I, [4] p. 202) qu'un tel espace, est caractérisé par un triplet (d,q,κ) tel que $d\in\mathbb{N}^*$, $(q,\kappa)\in\{0,1,3,7,d-1\}\times(0,\infty)$ où d est la dimension de M sur $\mathbb{R}$ et κ une courbure. En notant $M(d,q,\kappa)$ l'espace correspondant, les cas possibles sont donnés dans le Tableau 1. Nous omettrons les indices (n,d,κ) lorsqu'ils sont déterminés sans ambiguïté. Le diamètre δ_M de $M(d,q,\kappa)$, la surface $A(r)$ d'une boule de centre quelconque et de rayon $r\in[0,\delta_M]$ sont donnés par

$$\delta_M:=\max_{P,Q\in M(d,q,\kappa)} d(P,Q)=\frac{\pi}{2\sqrt{\kappa}},\quad \text{et}\quad A(r)=A[d,q,\kappa](r)=\frac{2\pi^{d/2}}{\Gamma(d/2)}S_\kappa^{d-1}(r)C_\kappa^q(r),\quad (0\leqslant r\leqslant\delta_M)$$

avec $S_\kappa(r):=(1/\sqrt{\kappa})\sin(\sqrt{\kappa}r)$, $C_\kappa(r):=\cos(\sqrt{\kappa}r)$ et nous noterons $T_\kappa=S_\kappa/C_\kappa$. Le volume de cette boule est donc

$$V(r)=V[d,q,\kappa](r)=\int_0^r A[d,q,\kappa](\rho)\,\mathrm{d}\rho=\left(\frac{\pi}{\kappa}\right)^{d/2}\frac{\Gamma((q+1)/2)}{\Gamma((q+1+d)/2)}I\left[\frac{d}{2},\frac{q+1}{2};\sin^2(\sqrt{\kappa}r)\right]$$

où I est la fonction Bêta incomplète définie par $I[a,b;x]:=\Gamma(a+b)/[\Gamma(a)\Gamma(b)]\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$ pour $a,b>0$ et $x\in[0,1]$. Fixons un pôle $N\in M$, et soit $P\mapsto r:=d(N,P)$ la fonction distance associée à ce pôle. Soit Δ_M le laplacien sur M, avec la convention de signe telle que toutes ses valeurs propres soient négatives (donc l'opposé de l'opérateur introduit dans [3], Chapitre II, § G.III p. 126). Les valeurs propres de $(-\Delta_M)$ sont les réels $\lambda_k=\lambda_k[d,q,\kappa]=2\kappa k(2k+d+q-1)$, $(k\in\mathbb{N})$. Munissons $L^2(M)$ de la norme induite par le produit scalaire $(f|g):=\int_M f(P)g(P)\,\mathrm{d}P$, $f,g\in L^2(M)$. Pour tout $k\in\mathbb{N}$, l' espace propre E_k associé à λ_k est de dimension finie et contient une unique fonction C^∞ ne s'annulant pas en N, de norme 1 et qui soit sphérique zonale relativement à N, c'est-à-dire ne dépende que de $r=d(P,N)$ (voir [7]). Notées $P\mapsto f_k^0(r)$, ces fonctions sont liées aux polynômes de Jacobi $P_k^{(\alpha,\beta)}(x):=\frac{(-1)^k}{k!2^k}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{\mathrm{d}^k}{\mathrm{d}x^k}[(1-x)^{k+\alpha}(1+x)^{k+\beta}]$, $(\alpha,\beta>-1,\ k\in\mathbb{N})$ par la relation, pour $k\in\mathbb{N}$,

$$\begin{aligned} f_k^0(r) &= \left(\frac{\kappa}{\pi}\right)^{d/4}\widetilde{P}_k^{(\frac{d-2}{2},\frac{q-1}{2})}[C_\kappa(2r)]\\ &:=\left(\frac{\kappa}{\pi}\right)^{d/4}\left\{\frac{k!(\frac{d+q-1}{2}+2k)\Gamma(\frac{d}{2})\Gamma(\frac{d+q-1}{2}+k)}{\Gamma(\frac{d}{2}+k)\Gamma(\frac{q+1}{2}+k)}\right\}^{1/2}P_k^{(\frac{d-2}{2},\frac{q-1}{2})}\big[C_\kappa(2r)\big]\end{aligned}$$

(voir [1] p. 65). Pour $k\geqslant 1$, l'orthogonal E_k^0 de $f_k^0(r)$ dans E_k est constitué de fonctions dont la moyenne est nulle sur toute sphère de centre N. Une base orthonormale de E_k^0 formée de fonctions C^∞ sera notée $\{f_k^i\colon 1\leqslant i\leqslant \dim E_k-1\}$. Il existe une unique fonction $G_M(P,Q)$, dite *fonction de Green centrée du laplacien*, solution de $(-\Delta_Q)G_M=\delta_P-1/V_M$, avec $\int_M G_M(P,Q)\,\mathrm{d}Q=0$ où δ_P est la distribution de Dirac en P et $V_M:=V(\delta)$ le volume total de M (voir [2], Chapitre 4, § 2.3 p. 108).

2. Résultats généraux

Théorème 2.1. *La fonction de Green centrée du Laplacien sur M est*

$$\begin{aligned}G_M(P_1,P_2)=g_M\big[d(P_1,P_2)\big]&:=\frac{1}{V_M}\int_{d(P_1,P_2)}^{\delta}\frac{V_M-V(r)}{A(r)}\,\mathrm{d}r-\frac{1}{V_M^2}\int_0^{\delta}\frac{V(r)[V_M-V(r)]}{A(r)}\,\mathrm{d}r\\&=\frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\left\{\int_{d(P_1,P_2)}^{\delta}\frac{I[\frac{q+1}{2},\frac{d}{2};\cos^2(\sqrt{\kappa}r)]}{S_\kappa^{d-1}(r)C_\kappa(r)}\,\mathrm{d}r-\int_0^{\delta}\frac{I[\frac{q+1}{2},\frac{d}{2};\cos^2(\sqrt{\kappa}r)]I[\frac{d}{2},\frac{q+1}{2};\sin^2(\sqrt{\kappa}r)]}{S_\kappa^{d-1}(r)C_\kappa(r)}\,\mathrm{d}r\right\}\\&=\sum_{k=1}^{\infty}\frac{f_k^0(r_1)f_k^0(r_2)}{\lambda_k}+\sum_{k=1}^{\infty}\frac{\sum_{\ell=1}^{(\dim E_k-1)}f_k^\ell(P_1)f_k^\ell(P_2)}{\lambda_k}.\end{aligned}\tag{1}$$

La série (1) *converge au sens des distributions quel que soit* $d=\dim M\in\mathbb{N}^*$*, dans* $L^2(M)$ *si* $\dim M\leqslant 3$*, ponctuellement si* $\dim M=1$*. Si* $d(N,P_i)=r_i$*,* $(i=1,2)$*, la valeur moyenne de* $P\mapsto G_M(P_1,P)$ *lorsque* P *parcourt sur la sphère* $S(r_2)$ *de centre* N*, contenant* P_2*, est donnée par*

$$G_M^{\mathcal{M}}(r_1,r_2):=\frac{1}{A(r_2)}\int_{S(r_2)}G_M(P_1,P)\,\mathrm{d}P=\sum_k\frac{f_k^0(r_1)f_k^0(r_2)}{\lambda_k}.\tag{2}$$

Notons $\chi^2(m)$ une v.a. de loi χ^2 à m degrés de liberté.

Théorème 2.2. *Si* $\dim M\in\{2,3\}$ (*resp.* $\dim M=1$)*, alors sous* H_0*, la* U*-statistique* (*resp. la* V*-statistique*)

$$U_M(P_1,\dots,P_n):=\frac{2}{(n-1)}\sum_{1\leqslant i<j\leqslant n}G_M(P_i,P_j),\ \left(\text{resp. } V_M(P_1,\dots,P_n):=\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n G_M(P_i,P_j)\right)\tag{3}$$

tend en loi vers la v.a. $\sum_{k=1}^\infty\frac{\chi_k^2(\dim E_k)-\dim E_k}{2\kappa k(2k+d+q-1)}$, (*resp.* $\sum_{k=1}^\infty\frac{\chi_k^2(\dim E_k)}{4\kappa k^2}$) *où les* χ_k^2*,* $k\geqslant 1$*, sont indépendantes.*

Le test suivant revient à se limiter à tester la répartition « radiale » de l'échantillon, puisque H_0' est clairement impliquée par H_0 si on a posé $r_i=d(N,P_i)$. Remarquons que les fonctions A, V, G_M et g_M sont définies pour des valeurs non entières de d et q, même si la variété correspondante $M(d,q,\kappa)$ n'existe pas. On pourrait tenir compte de cette extension dans le théorème suivant. En particulier l'identité (4) généraliserait (2) à des valeurs non entières.

Théorème 2.3. *Pour tout* $d\in\mathbb{N}^*$*, on a*

$$\begin{aligned}G_M^{\mathcal{M}}(r_1,r_2)&=\sum_{k=1}^{\infty}\frac{\widetilde{P}_k^{(\frac{d-2}{2},\frac{d-2}{2})}[\cos(\sqrt{\kappa}r_1)]\widetilde{P}_k^{(\frac{d-2}{2},\frac{d-2}{2})}[\cos(\sqrt{\kappa}r_2)]}{4\kappa k(k+d-1)},\quad\left(0\leqslant r_1,r_2\leqslant\frac{\pi}{2\sqrt{\kappa}}\right)\\&=\begin{cases}g_M(\delta-r_1)+g_M(r_2)-g_M(\delta) & \text{si } r_1\leqslant r_2,\\ g_M(r_1)+g_M(\delta-r_2)-g_M(\delta) & \text{si } r_1\geqslant r_2.\end{cases}\\&=\frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\int_0^{\delta}\frac{I[\frac{d}{2},\frac{d}{2};\cos^2(\sqrt{\kappa}r)]\{1_{\{r\geqslant\max(\pi-r_1,\pi-r_2)\}}+1_{\{r\geqslant\max(r_1,r_2)\}}-I[\frac{d}{2},\frac{d}{2};\sin^2(\sqrt{\kappa}r)]\}}{S_\kappa^{d-1}(r)C_\kappa^{d-1}(r)}\,\mathrm{d}r.\end{aligned}\tag{4}$$

Sous l'hypothèse

H_0' : *L'échantillon* $r_1,\dots,r_n$ *provient d'une distribution de densité* $A[d,d-1,\kappa]$ *sur* $(0,\frac{\pi}{2\sqrt{\kappa}})$,

la V*-statistique*

$$V_M^{\mathcal{M}}(r_1,\dots,r_n):=\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n G_M^{\mathcal{M}}(r_i,r_j)=-ng_M(\delta)+\frac{1}{n}\sum_{i=1}^n(2i-1)\big[g_M(\delta-r_{(i)})+g_M(r_{(i)})\big]\tag{5}$$

tend en distribution vers la v.a. $\sum_{k=1}^\infty\frac{\xi_k^2}{4\kappa k(k+d-1)}$ *où* $(\xi_k)_{k\geqslant1}$ *est une suite de variables* $\mathcal{N}(0,1)$ *indépendantes.*

Exemple 1. Pour le cercle $S^1(R) = M(1,0,1/(4R^2))$ de rayon R, on obtient

$$G_{1,0,\frac{1}{4R^2}}(P,Q) = \frac{[\pi R - d(P,Q)]^2}{4\pi R} - \frac{\pi R}{12} \quad \text{et}$$

$$G^{\mathcal{M}}_{1,0,\frac{1}{4R^2}}(r_1,r_2) = \frac{r_1^2 + (\pi R - r_2)^2}{4\pi R} - \frac{\pi R}{12}, \quad (0 \leqslant r_1, r_2 \leqslant \pi R). \tag{6}$$

Si pour $i = 1,\dots,n$, l'argument de P_i est $2\pi x_i$ avec $x_i \in [0,1]$, notons $x_{(1)} < \cdots < x_{(n)}$ cet échantillon ordonné et $\bar{x} := n^{-1}\sum_{i=1}^n x_i$. On obtient alors

$$\begin{aligned} V_{1,0,\frac{1}{4R^2}}(P_1,\dots,P_n) &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left\{\frac{(\pi R - d(P_i,P_j))^2}{4\pi R} - \frac{\pi R}{12}\right\} = \frac{2\pi R}{n}\sum_{i=1}^n\sum_{j=1}^n\left\{\frac{(|x_i - x_j| - 1/2)^2}{2} - \frac{1}{24}\right\} \\ &= 2\pi R U_n^2 \end{aligned} \tag{7}$$

où $U_n^2 = \sum_{i=1}^n (x_{(i)} - \frac{i-1/2}{n})^2 + \frac{1}{12n} - n(\bar{x} - \frac{1}{2})^2$ est la statistique de Watson (voir [6], (4.17) p. 27). Si ensuite on pose $d(P_i,N) = \pi y_i$ (donc $y_i \in [0,1]$) en notant $y_{(1)} < \cdots < y_{(n)}$ cet échantillon ordonné, on obtient $\frac{2}{\pi R}V^{\mathcal{M}}_{1,0,\frac{1}{4R^2}}(P_1,\dots,P_n) = W_n^2(y_1,\dots,y_n) := \sum_{i=1}^n (y_{(i)} - \frac{i-1/2}{n})^2 + \frac{1}{12n}$ où W_n^2 n'est autre que la statistique de Cramér–von Mises (voir [6], (4.18) p. 27).

Exemple 2. $M(2,1,\frac{1}{4})$ est la sphère unité de $\mathbb{R}^3$. Les calculs donnent $G_{2,1,\frac{1}{4}}(P_1,P_2) = \frac{\log 2 - 1 - \log[1-\cos d(P_1,P_2)]}{4\pi}$, $G^{\mathcal{M}}_{2,1,\frac{1}{4}}(r_1,r_2) = -\frac{1}{4\pi}\log\frac{e(1+\cos r_1)(1-\cos r_2)}{4}$. En remarquant que $(2[1-\cos d(P,Q)])^{1/2} =: \delta(P,Q)$ est la distance entre P et Q dans $\mathbb{R}^3$, on obtient

$$U_{2,1,\frac{1}{4}}(P_1,\dots,P_n) = -(n/(4\pi))\log\bigl(e/4 \cdot g_n^2\bigr) \tag{8}$$

où $g_n := \prod_{1\leqslant i<j\leqslant n}\delta(P_i,P_j)^{\frac{2}{n(n-1)}}$ est la moyenne géométrique des distances dans $\mathbb{R}^3$ entre les points distincts de l'échantillon. Elle tend en loi vers la variable aléatoire $\sum_{k=1}^\infty \frac{\chi_k^2(2k+1)-(2k+1)}{k(k+1)}$ où les variables $\{\chi_k^2(2k+1)\colon k \geqslant 1\}$ sont indépendantes. En posant $x_i = (1-\cos r_i)/2$, $(1 \leqslant i \leqslant n)$, on obtient

$$4\pi V^{\mathcal{M}}_{2,1,\frac{1}{4}}(r_1,\dots,r_n) = -n - \frac{1}{n}\sum_{i=1}^n (2i-1)\log\bigl[x_{(i)}(1-x_{(n-i+1)})\bigr] = A_n^2(x_1,\dots,x_n) \tag{9}$$

où A_n^2 est la statistique d'Anderson–Darling (voir [6], (5.4.2) p. 36).

Remerciements

L'auteur remercie Paul Deheuvels, dont la suggestion d'établir un lien entre les développements de Karhunen–Loève explicites et la représentation de certains groupes de Lie a été à l'origine de ces recherches.

Références

[1] R. Askey, Orthogonal Polynomials and Special Functions, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1975.
[2] T. Aubin, Nonlinear Analysis on Manifolds, Monge–Ampère Equations, Grundlehren Math. Wiss. (Fundamental Principles of Mathematical Sciences), vol. 252, Springer-Verlag, New York, 1982.
[3] M. Berger, P. Gauduchon, E. Mazet, Le Spectre d'une Variété Riemannienne, Lecture Notes in Math., vol. 194, Springer-Verlag, 1971.
[4] A.L. Besse, Manifolds all of Whose Geodesics Are Closed, Ergeb. Math. Grenzgeb. (Results in Mathematics and Related Areas), vol. 93, Springer-Verlag, Berlin–New York, 1978 (ix+262 p.).
[5] I. Chavel, Riemannian Symmetric Spaces of Rank One, Lecture Notes in Pure and Appl. Math., vol. 5, Marcel Dekker, New York, 1972 (vii+81 p.).
[6] J. Durbin, Distribution Theory for Tests Based on the Sample Distribution Function, CBMS-NSF Regional Conf. Ser. in Appl. Math., vol. 9, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1973.
[7] M.E. Giné, Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms, Ann. Statist. 3 (6) (1975) 1243–1266.
[8] A. Gray, The volume of a small geodesic ball of a Riemannian manifold, Michigan Math. J. 20 (1973) 329–344.
[9] S. Helgason, Groups and Geometric Analysis. Integral Geometry, Invariant Differential Operators, and Spherical Functions, Pure Appl. Math., vol. 113, Academic Press, Orlando, FL, 1984 (xix+654 p.).
[10] V.S. Koroljuk, Yu.V. Borovskich, Theory of U-Statistics, Math. Appl., vol. 273, Kluwer Academic Publishers Group, Dordrecht, 1994 (x+552 p.).