J. Appl. Prob. **51,** 1081–1099 (2014)
Printed in England

DIRICHLET RANDOM WALKS

GÉRARD LETAC,* *Université Paul Sabatier*
MAURO PICCIONI,** *Sapienza Università di Roma*

Abstract

This paper provides tools for the study of the Dirichlet random walk in $\mathbb{R}^d$. We compute explicitly, for a number of cases, the distribution of the random variable W using a form of Stieltjes transform of W instead of the Laplace transform, replacing the Bessel functions with hypergeometric functions. This enables us to simplify some existing results, in particular, some of the proofs by Le Caër (2010), (2011). We extend our results to the study of the limits of the Dirichlet random walk when the number of added terms goes to ∞, interpreting the results in terms of an integral by a Dirichlet process. We introduce the ideas of Dirichlet semigroups and Dirichlet infinite divisibility and characterize these infinite divisible distributions in the sense of Dirichlet when they are concentrated on the unit sphere of $\mathbb{R}^d$.

Keywords: Dirichlet process; Stieltjes transform; random flight; distributions in a sphere; hyperuniformity; infinite divisibility in the sense of Dirichlet

2010 Mathematics Subject Classification: Primary 60D99; 60F99

1. Introduction

In this paper we study the distribution of

$$W_{n,d} = X_1\Theta_1 + \cdots + X_n\Theta_n,$$

where $X = (X_1, \ldots, X_n) \sim \mathcal{D}(q_1, \ldots, q_n)$ is Dirichlet distributed and $\Theta_1, \ldots \Theta_n$ are independent, identically distributed (i.i.d.), uniformly distributed on the unit sphere S_{d-1} of the Euclidean space $\mathbb{R}^d$ and independent of X (the scalar product on $\mathbb{R}^d$ is $\langle x, y\rangle = x_1y_1 + \cdots + x_dy_d$). In other terms, we select independently and uniformly n random points on S_{d-1} and we take their barycenter according to the random weights $(X_1, \ldots, X_n)$.

We term (*improperly*) the random variable $W_{n,d}$ a Dirichlet random walk because if $Y_j \sim \gamma_{q_j}$, $j = 1, 2, \ldots,$ are independent and if $S_n = Y_1 + \cdots + Y_n$ then the sequence

$$\left(\sum_{j=1}^{n} Y_j\Theta_j\right)_{n\geq 1}$$

is an inhomogeneous random walk on E_d and $W_{n,d} \sim (1/S_n)\sum_{j=1}^{n} Y_j\Theta_j$ according to the well-known properties of gamma and Dirichlet distributions. The random variable $W_{n,d}$ leads to a simpler analysis than the study of $\sum_{j=1}^{n}\Theta_j$, initiated by Lord Kelvin, and the subject of a recent in-depth paper by Borwein *et al.* (2012), where $d = 2$ and $n = 3, 4, 5$.

Received 22 October 2013; revision received 22 January 2014.
* Postal address: Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France. Email address: letac@math.univ-toulouse.fr
** Postal address: Dipartimento di Matematica, Sapienza Università di Roma, 00185 Rome, Italia.

If U is any random variable concentrated in the closed unit sphere, B_d of E_d, our technique for characterizing the distribution of U is to consider the function defined on the interior $\mathring{B}_d$ of B_d by

$$y \mapsto T_p(U)(y) = \mathbb{E}\left(\frac{1}{(1 + \langle y, U\rangle)^p}\right),$$

where p is a fixed positive number. The function $T_p(U)$ can be considered as a kind of Stieltjes transform of the distribution of U. It is straightforward to prove that knowledge of the function $T_p(U)$ gives knowledge of the moments of the real variable $\langle y, U\rangle$. Since U is bounded, this identifies the distribution of U. However, recovering the distribution of U from $T_p(U)$ is not always easy; in particular, from $T_p(U)$ and $T_q(V)$ and if $p \neq q$, it may be unclear whether $U \sim V$ or not.

In all the cases considered here, the distribution of U is invariant by rotation and $y \mapsto T_p(U)(y)$ depends only on $\|y\|$. We borrow from Le Caër (2010) the following terminology. We say that U is hyperspherically uniform (*hyperuniform* for short) of type $k > d$ if U is the law of the projection of a random variable Θ uniform on the unit sphere S_{k-1} of $\mathbb{R}^k$ onto a subspace of dimension d. The term 'hyperuniform' is justified by the fact that when $k = d+2$ the induced distribution is uniform on the unit sphere B_d of $\mathbb{R}^d$ (for $k = 3$ and $d = 1$ this result is due to Archimedes).

Le Caër's paper (2010) is motivated by problems in metallurgy. It studies the cases where the Dirichlet random walk $W_{n,d}$ is hyperuniform of type k for some $k > d$. His main result is the description of the quadruplets (q, d, n, k) such that $W_{n,d}$ is hyperuniform of type k when $X \sim \mathcal{D}(q, q, \dots, q)$ (see Theorems 6 and 9 below). We refer to Le Caër's paper for the original proofs, for the bibliography concerning this problem, and for some interesting simulations. In Section 3 we provide new proofs of Le Caër's statements. Section 4 is devoted to explicit expressions of the distribution of $W_{n,d}$ when $X \sim \mathcal{D}(q, \dots, q)$, where Theorem 4 deals with the case where $q = d$ and gives the explicit form of the density of $\|W_{n,d}\|^2$ as a mixing of beta distributions. In Theorem 5 we deal with $q = 1$ and even d, with results close to those of Kolesnik (2009). In Section 5 we rule out the easy case where $n = 2$. In Section 6 we consider the limits of $W_{n,d}$ with $(X_1, \dots, X_n) \sim \mathcal{D}(Q/n, \dots, Q/n)$, when n goes to ∞ with Q and d fixed, and interprets the results in terms of an integral by a Dirichlet process. The examples in Section 6 lead to the idea of infinite divisibility in the sense of Dirichlet, introduced in Section 7.

2. Dirichlet random walks

Theorem 1. *Let* $(q_1, \dots, q_n)$ *be positive numbers and let* $(X_1, \dots, X_n) \sim \mathcal{D}(q_1, \dots, q_n)$. *We write* $Q = q_1 + \dots + q_n$. *Let* $\mathring{B}_d$ *be the interior of the unit sphere of* $\mathbb{R}^d$, S_{d-1} *its unit sphere and let* $\Theta_1, \dots \Theta_n$ *be i.i.d., each of them being uniformly distributed on* S_{d-1}. *We consider*

$$W_{n,d} = X_1\Theta_1 + \dots + X_n\Theta_n.$$

Then, for $y \in \mathring{B}_d$, *we have*

$$\mathbb{E}\left(\frac{1}{(1 + \langle y, W_{n,d}\rangle)^Q}\right) = \prod_{j=1}^{n} {}_2F_1\left(\frac{q_j}{2}, \frac{q_j+1}{2}; \frac{d}{2}; \|y\|^2\right). \quad (1)$$

In order to give a proof we need a lemma. In what follows $(a)_k$ is the familiar Pochhammer symbol $(a)_0 = 1$ and $(a)_k = a(a+1)\dots(a+k-1) = \Gamma(a+k)/\Gamma(a)$ for $k > 0$.

Lemma 1. *Let Θ be uniform on S_{d-1} and $y \in \mathring{B}_d$. For $p > 0$ and k, a positive integer, we have*

$$\mathbb{E}(|\langle y, \Theta\rangle|^p) = \|y\|^p \frac{\Gamma((p+1)/2)\Gamma(d/2)}{\Gamma(1/2)\Gamma((p+d)/2)}, \qquad \mathbb{E}(\langle y, \Theta\rangle^{2k}) = \|y\|^{2k} \left(\frac{1}{2}\right)_k \Big/ \left(\frac{d}{2}\right)_k, \tag{2}$$

$$\mathbb{E}\left(\frac{1}{(1+\langle y, \Theta\rangle)^p}\right) = {}_2F_1\left(\frac{p}{2}, \frac{p+1}{2}; \frac{d}{2}; \|y\|^2\right). \tag{3}$$

Proof of Lemma 1. Consider the Gaussian random variable $Z = (Z_1, \dots, Z_d)$ valued in $\mathbb{R}^d$ with distribution $N(0, I_d)$. Then $Z/\|Z\|$ is uniformly distributed on S_{d-1} and $\langle y, \Theta\rangle \sim \|y\|Z_1/\|Z\|$. Furthermore, $\langle y, \Theta\rangle$ is a symmetric random variable and the knowledge of its distribution is given by the knowledge of the distribution of its square. However,

$$\begin{aligned}\frac{1}{\|y\|^2}\langle y, \Theta\rangle^2 &\sim \frac{Z_1^2}{Z_1^2 + \cdots + Z_d^2} \\ &\sim \beta\left(\frac{1}{2}, \frac{d-1}{2}\right) \\ &= \frac{1}{B(1/2, (d-1)/2)} z^{1/2-1}(1-z)^{(d-1)/2-1}\mathbf{1}_{(0,1)}(z)\,\mathrm{d}z\end{aligned}$$

since the Z_j^2 are independent chi square random variables (RV) with one degree of freedom. Equation (2) is now clear. Now we prove (3) by using the second part of (2):

$$\begin{aligned}\mathbb{E}\left(\frac{1}{(1+\langle y, \Theta\rangle)^p}\right) &= \sum_{i=0}^{\infty} \frac{(p)_i}{i!}(-\|y\|)^i \mathbb{E}\left(\left(\frac{1}{\|y\|}\langle y, \Theta\rangle\right)^i\right) \\ &= \sum_{j=0}^{\infty} \frac{(p)_{2j}}{(2j)!}\|y\|^{2j}\frac{(1/2)_j}{(d/2)_j} \\ &= \sum_{j=0}^{\infty} \frac{(p/2)_j((p+1)/2)_j}{(d/2)_j\, j!}\|y\|^{2j} \\ &= {}_2F_1\left(\frac{p}{2}, \frac{p+1}{2}; \frac{d}{2}; \|y\|^2\right).\end{aligned}$$

Proof of Theorem 1. We apply the following principle (see Chamayou and Letac (1994)). If $X = (X_1, \dots, X_n) \sim \mathcal{D}(q_1, \dots, q_n)$ with $Q = q_1 + \cdots + q_n$, and if $f = (f_1, \dots, f_n)$ is such that $f_i > 0$ for all i, then

$$\mathbb{E}\left(\frac{1}{\langle f, X\rangle^Q}\right) = \frac{1}{\prod_{i=1}^n f_i^{q_i}}. \tag{4}$$

Condition by $(\Theta_1, \dots, \Theta_n)$ and apply (4) to $f_i = 1 + \langle y, \Theta_i\rangle$, we obtain, from (3) and, (4)

$$\begin{aligned}\mathbb{E}\left(\frac{1}{(1+\langle y, W\rangle)^Q}\right) &= \mathbb{E}\left(\mathbb{E}\left(\frac{1}{(1+\langle y, W\rangle)^Q} \,\Big|\, (\Theta_1, \dots, \Theta_n)\right)\right) \\ &= \mathbb{E}\left(\frac{1}{\prod_{i=1}^n (1+\langle y, \Theta_i\rangle)^{q_i}}\right) \\ &= \prod_{j=1}^n {}_2F_1\left(\frac{q_j}{2}, \frac{q_j+1}{2}; \frac{d}{2}; \|y\|^2\right).\end{aligned}$$

In what follows we will need the following simple identities relating hypergeometric functions with the function defined for $|z| < 1$ by

$$G(z) = \frac{2}{z}(1 - \sqrt{1-z}) = \frac{2}{1+\sqrt{1-z}}. \tag{5}$$

Lemma 2. *We have, for* $|z| < 1$,

$$ {}_2F_1\left(\frac{c}{2}, \frac{c+1}{2}; c+1; z\right) = G(z)^c \quad \text{if } c > 0, \tag{6}$$

$$ {}_2F_1\left(\frac{c}{2}, \frac{c+1}{2}; c; z\right) = \frac{1}{\sqrt{1-z}} G(z)^{c-1} \quad \text{if } c > 1, \tag{7}$$

$$ \sum_{n=1}^{\infty} \frac{(1/2)_n}{n!} \frac{z^n}{2n} = \log G(z). \tag{8}$$

Proof. Equation (6) is a consequence of ${}_2F_1(c/2, (c+1)/2; c+1; 4u-4u^2) = {}_1F_0(c; u)$, which is a particular case of the classical and nontrivial identity

$$ {}_2F_1\big(2a, 2b; a+b+\tfrac{1}{2}; u\big) = {}_2F_1\big(a, b; a+b+\tfrac{1}{2}; 4u-4u^2\big) \quad \text{true for } 0 \le u \le \tfrac{1}{2}.$$

Equation (7) is a consequence of the Euler identity

$$ {}_2F_1(p, q; r; z) = (1-z)^{r-p-q} {}_2F_1(r-p, r-q; r; z)$$

applied to $p = c/2$, $q = (c+1)/2$, and $r = c$, then using (6). If f is the left-hand side of (8) it is easily seen that

$$ f'(z) = \frac{1}{2z}\left(\frac{1}{\sqrt{1-z}} - 1\right), \qquad f(z) = \int_0^z f'(t)\,\mathrm{d}t = \int_{\sqrt{1-z}}^1 \frac{\mathrm{d}s}{1+s} = \log G(z)$$

by the change of variable $t = 1 - s^2$.

The next proposition shows that under certain circumstances the computation of the moments of $\|W_{n,d}\|^2$ is easy with knowledge of $T_a(W_{n,d})(y)$.

Proposition 1. *Let Θ be uniform on the unit sphere S_{d-1}. Assume that Θ is independent of the random variable $R \in [0, 1]$, and denote $W = R\Theta$. Let $a > 0$ and suppose that $T_a(W)$ has the form of the left-hand side of (9)–(11) for some $b > 0$. Then the even moments of R have the form of the right-hand side. We have*

$$ T_a(W)(y) = \frac{1}{(1-\|y\|^2)^b} \implies \mathbb{E}(R^{2k}) = \frac{(b)_k (d/2)_k}{(a/2)_k((a+1)/2)_k}, \tag{9}$$

$$ T_a(W)(y) = G(\|y\|^2)^b \implies \mathbb{E}(R^{2k}) = \frac{(b)_{2k}(d/2)_k}{(a)_{2k}(b+1)_k}, \tag{10}$$

$$ T_a(W)(y) = \frac{1}{\sqrt{1-\|y\|^2}} G(\|y\|^2)^{b-1} \implies \mathbb{E}(R^{2k}) = \frac{(b)_{2k}(d/2)_k}{(a)_{2k}(b)_k}. \tag{11}$$

Proof. We first expand $\mathbb{E}((1+\langle y, W\rangle)^{-a})$ in powers of $\|y\|^2$. Using the fact that the odd moments of $\langle y, W\rangle$ are 0 and using (2), we obtain

$$\begin{aligned}\mathbb{E}\bigg(\frac{1}{(1+\langle y, W\rangle)^a}\bigg) &= \sum_{k=0}^{\infty}\frac{(a)_{2k}}{(2k)!}\mathbb{E}(\langle y, W\rangle^{2k}) \\ &= \sum_{k=0}^{\infty}\frac{(a)_{2k}}{(2k)!}\mathbb{E}(\langle y, \Theta\rangle^{2k})\mathbb{E}(R^{2k}) \\ &= \sum_{k=0}^{\infty}\frac{(a)_{2k}}{(2k)!}\frac{(1/2)_k}{(d/2)_k}\mathbb{E}(R^{2k})\|y\|^{2k}. \end{aligned} \tag{12}$$

Observe also that

$$\bigg(\frac{1}{2}\bigg)_k = \frac{(2k)!}{2^{2k}k!}, \qquad \frac{(b/2)_k((b+1)/2)_k}{(1/2)_k}\frac{(2k)!}{k!} = (b)_{2k}.$$

Using the fact that the power series of $G(z)^b$ and $G(z)^{b-1}/\sqrt{1-z}$ are known by Lemma 2, we obtain the results of Proposition 1.

We now apply the equations from Proposition 1 with R^2 beta distributed. Actually

$$R^2 \sim \beta(p, q) \Longleftrightarrow \mathbb{E}(R^{2k}) = \frac{(p)_k}{(p+q)_k}. \tag{13}$$

Proposition 2. *The following hold.*

(i) *If $W_{n,d}$ is a Dirichlet random walk governed by $\mathcal{D}(d-1, \dots, d-1)$ and if $R_{n,d} = \|W_{n,d}\|$ then*

$$R_{n,d}^2 \sim \beta\bigg(\frac{d}{2}, \frac{(n-1)(d-1)}{2}\bigg).$$

The same law is obtained if one of the Dirichlet parameters is set to d.

(ii) *If $W_{n,d}$ is a Dirichlet random walk governed by $\mathcal{D}(d/2-1, \dots, d/2-1)$ and if $R_{n,d} = \|W_{n,d}\|$ then*

$$R_{n,d}^2 \sim \beta\bigg(\frac{d}{2}, (n-1)\bigg(\frac{d}{2}-1\bigg)\bigg).$$

The same law is obtained if one of the Dirichlet parameters is set to $d/2$.

Proof. The first statement follows from (1) and (9) with $a = 2b = n(d-1)$ and $a-1 = 2(b-1) = n(d-1)$. The second statement follows from (10) with $a = b = n(d/2-1)$ and from (11) with $a = b = n(d/2-1)+1$.

Note that when d, m, and n are positive integers such that

$$(n-1)(d-1) = (m-1)(d-2) = t$$

we obtain, from the previous result, four different representations as Dirichlet random walks for the same law $\beta(d/2, t/2)$ for the squared radius. For example, with $d = 3$, $n = 2$, $m = 3$ the same law $\beta(\frac{3}{2}, 1)$ for the squared radius is obtained from $X\Theta_1 + (1-X)\Theta_2$, $X'\Theta_1 + (1-X')\Theta_2$, $Y_1\Theta_1 + Y_2\Theta_2 + Y_3\Theta_3$, and $Y_1'\Theta_1 + Y_2'\Theta_2 + Y_3'\Theta_3$, where $X \sim \beta(2, 2)$, $Y \sim \beta(2, 3)$, $X' \sim \mathcal{D}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, and $Y' \sim \mathcal{D}(\frac{3}{2}, \frac{1}{2}, \frac{1}{2})$. This example was already noted in Le Caër (2010).

3. Hyperuniformity

It is easy to describe hyperuniform random vectors by introducing the standard Gaussian random variable

$$Z = (Z_1, \dots, Z_k) \sim N(0, I_k).$$

Then $U = Z/\|Z\|$ is uniformly distributed on the unit sphere of $\mathbb{R}^k$ and $(Z_1/\|Z\|, \dots, Z_d/\|Z\|)$ is hyperuniform on $\mathbb{R}^d$ of type k. We have

$$\|U\|^2 = \frac{Z_1^2 + \dots + Z_d^2}{Z_1^2 + \dots + Z_k^2} \sim \beta_{d/2,(k-d)/2}$$

by standard beta gamma algebra. Note that, for $k - d = 2$, the distribution $\beta_{d/2,1}$ is the measure induced on the squared radius by the Lebesgue measure in B_d, in agreement with Archimedes' theorem. Thus, we can rephrase and extend the definition in the introduction by allowing $k > d$ to be an arbitrary real number. There is no particular problem in allowing $k > d$ to be a real number.

Definition. Let U be a random variable valued in the unit sphere of $\mathbb{R}^d$ such that its distribution is invariant by rotation (namely $a(U) \sim U$ for all $a \in \mathbb{O}(\mathbb{R}^d)$). We say that U is hyperuniform of type $k > d$ if $\|U\|^2 \sim \beta(d/2, (k-d)/2)$.

With Proposition 2 we immediately have the first main result of Le Caër (2010).

Theorem 2. *Let $X = (X_1, \dots, X_n) \sim \mathcal{D}(q + m, q, \dots, q)$ with $m = 0$ or 1. Let $(\Theta_j)_{j=1}^n$ be i.i.d. and uniform on the unit sphere of $\mathbb{R}^d$. Then $W_{n,d} = X_1\Theta_1 + \dots + X_n\Theta_n$ is hyperuniform of type $k > d$ in the unit sphere of $\mathbb{R}^d$ irrespectively of $m = 0, 1$, if*

(i) *$d \geq 2$, $n \geq 2$, $q = d - 1$, and $k = n(d-1) + 1$;*

(ii) *$d \geq 3$, $n \geq 2$, $q = d/2 - 1$, and $k = n(d-2) + 2$.*

Proof. This is immediately obtained from Proposition 2.

Proposition 3. *If U in the unit sphere of dimension d is invariant by rotation, then U is hyperuniform of type $k > d$ if and only if, for any $p > 0$, we have*

$$T_p(U)(y) = {}_2F_1\left(\frac{p}{2}, \frac{(p+1)}{2}; \frac{k}{2}; \|y\|^2\right).$$

Proof. From Lemma 1 and (13),

$$\begin{aligned}\mathbb{E}\left(\frac{1}{(1 + \langle U, y\rangle)^p}\right) &= \mathbb{E}\left(\frac{1}{(1 + \langle \Theta, \|U\| y\rangle)^p}\right) \\ &= \mathbb{E}\left({}_2F_1\left(\frac{p}{2}, \frac{(p+1)}{2}; \frac{d}{2}; \|U\|^2\|y\|^2\right)\right) \\ &= {}_2F_1\left(\frac{p}{2}, \frac{(p+1)}{2}; \frac{k}{2}; \|y\|^2\right).\end{aligned}$$

Corollary 1. *Let $X = (X_1, \dots, X_n) \sim D(q_1, \dots, q_n)$ and $Q = q_1 + \dots + q_n$. Let $\Theta_1, \dots, \Theta_n$ be i.i.d., and uniform on the unit sphere of $\mathbb{R}^d$ and independent of X. Then $W = X_1\Theta_1 + \dots +$*

$X_n\Theta_n$ is hyperuniform of type $k > d$ if and only if, for $|z| < 1$,

$$\prod_{i=1}^{n} {}_2F_1\left(\frac{q_i}{2}, \frac{(q_i+1)}{2}; \frac{d}{2}; z\right) = {}_2F_1\left(\frac{Q}{2}, \frac{(Q+1)}{2}; \frac{k}{2}; z\right).$$

Proof. This is an immediate consequence of Theorem 1 and Proposition 3, with z replacing the term $\|y\|^2$.

The second main result of Le Caër (2010) is a partial converse of Theorem 2 and its proof and includes an unexplained miracle.

Theorem 3. *Let $X = (X_1, \ldots, X_n) \sim \mathcal{D}(q, \ldots, q)$. Let $(\Theta_j)_{j=1}^n$ be i.i.d. and uniform on the unit sphere of $\mathbb{R}^d$. Then $W_{n,d} = X_1\Theta_1 + \cdots + X_n\Theta_n$ is hyperuniform of type $k > d$ in the unit sphere of E only if*

(i) *either $d \ge 2$, $q = d-1$, $n \ge 2$, and $k = n(d-1)+1$; or*

(ii) *$d \ge 3$, $q = d/2 - 1$, $n \ge 2$, and $k = n(d-2)+2$.*

Proof. By Proposition 3 and Corollary 1 the random variable $W_{n,d}$ is hyperuniform of order k if and only if the following holds:

$$ {}_2F_1\left(\frac{q}{2}, \frac{q+1}{2}; \frac{d}{2}; z\right)^n = {}_2F_1\left(\frac{nq}{2}, \frac{nq+1}{2}; \frac{k}{2}; z\right).$$

With the notation

$$a_i = \frac{(q/2)_i((q+1)/2)_i}{i!\,(d/2)_i} = \frac{(q)_{2i}}{2^{2i}i!\,(d/2)_i}, \qquad b_i = \frac{(nq/2)_i((nq+1)/2)_i}{i!\,(k/2)_i} = \frac{(nq)_{2i}}{2^{2i}i!\,(k/2)_i},$$

for $i = 1, 2, \ldots,$ we have

$$(1 + a_1z + a_2z^2)^n \equiv 1 + b_1z + b_2z^2 \mod z^3.$$

Expanding the left-hand side term through the multinomial formula, we obtain two equations:

$$b_1 = na_1, \tag{14}$$

$$b_2 = na_2 + \frac{n(n-1)}{2}a_1^2. \tag{15}$$

Equality (14) implies $k(q+1) = d(nq+1)$. Equality (15), after simplification, becomes

$$\frac{(nq+1)(nq+2)(nq+3)}{k(k+2)} = \frac{(q+1)(q+2)(q+3)}{d(d+2)} + (n-1)\frac{q(q+1)^2}{d^2}.$$

Replacing k by $d(nq+1)/(q+1)$ in this equation and simplifying again leads to

$$\frac{(q+1)(nq+2)(nq+3)}{d(nq+1)+2q+2} = \frac{(q+2)(q+3)}{d+2} + (n-1)\frac{q(q+1)}{d}.$$

Next, set $x = d/(q+1)$, obtaining (for fixed q) an equation in x:

$$\frac{(nq+2)(nq+3)}{x(nq+1)+1} = \frac{(q+2)(q+3)}{x(q+1)+1} + (n-1)\frac{q}{x}. \tag{16}$$

Now we attend to a little *miracle* of the theory. Equation (16) is a second order equation in x, whose solutions $x = 1$ and $x = 2$ depend neither on n nor q. From this we have that (16) has only $q = d-1$ and $q = d/2 - 1$ as solutions, as desired.

Remark. From Theorems 2 and 3 the only Dirichlet walks $W_{n,d}$ governed by $(X_1, \dots, X_n) \sim \mathcal{D}(q, \dots, q)$ are uniform (that they are hyperuniform with type $k = d + 2$) if and only if

(i) $d = 2,\ n = 3,\ q = 1;$

(ii) $d = 3,\ n = 2,\ q = 2;$

(iii) $d = 3,\ n = 3,\ q = \frac{1}{2};$

(iv) $d = 4,\ n = 2,\ q = 1.$

4. Density of the Dirichlet random walk in the case where $\mathcal{D}(q, \dots, q)$

Let us study the density of the squared radius of the Dirichlet random walk $W_{n,d}$ governed by $\mathcal{D}(q, \dots, q)$, with $n \geq 2$. The cases in which $q = d - 1$ and $q = d/2 - 1$ have already been dealt with in Theorem 2. First, we rule out the simple cases where $d = 1$ and q is arbitrary (Proposition 4). We present the main results of this section for the case where $q = d$ (Theorem 4), and the case where $q = 1$ and d is even (Theorem 5). The case where $q = d$ is the subject of Le Caër (2011). The case where $q = d = 2$ is considered by Beghin and Orsingher (2010). The case where $(q, d) = (1, 6)$ is the subject of Kolesnik (2009).

For the case where $d = 1$, for arbitrary $q > 0$.

For this case we can deal with a slightly more general case. A Dirichlet random walk occurs when $p = \frac{1}{2}$ in the following proposition.

Proposition 4. *Let $X = (X_1, \dots, X_n) \sim \mathcal{D}(q, \dots, q)$ (n times) and let $\epsilon_1, \dots, \epsilon_n$ be i.i.d. random variables independent of X with distribution $p\delta_1 + (1 - p)\delta_{-1}$. Then the law of $Y = X_1\epsilon_1 + \dots + X_n\epsilon_n$ is*

$$(1 - p)^n \delta_{-1}(\mathrm{d}y) + p^n \delta_1(\mathrm{d}y)$$
$$+ \Gamma(nq)2^{-nq-1} \sum_{k=1}^{n-1} \binom{n}{k} p^k (1 - p)^{n-k} \frac{(1 + y)^{kq-1}(1 - y)^{(n-k)q-1}}{\Gamma(kq)\Gamma((n - k)q)} 1_{(-1,1)}(y)\,\mathrm{d}y.$$

Proof. This is easy by conditioning on $K = \epsilon_1 + \dots + \epsilon_n$.

For the case where $q = d > 1$.

With $R = R_{n,d} = \|W_{n,d}\|$, Theorem 1 and (9) show that, for $q = d$,

$$\mathbb{E}(R^{2k}) = \frac{((nd + n)/2)_k}{((nd + 1)/2)_k} \frac{(d/2)_k}{(nd/2)_k}.$$

In other terms, if $Z \sim R^2$ and X are independent then $X \sim \beta((nd + 1)/2, (n - 1)/2)$ implies

$$Y = XZ \sim \beta\left(\frac{d}{2}, \frac{(n - 1)d}{2}\right). \tag{17}$$

This is a consequence of (13). However, finding the distribution of Z from the multiplicative convolution equation (17) is not easy. Here is the result.

Theorem 4. *Suppose that $d > 1$. Consider the Dirichlet random walk $W_{n,d}$ governed by $\mathcal{D}(d, \dots, d)$, denote $R_{n,d} = \|W_{n,d}\|$ and let f be the density of $V = R^2_{n,d}$. Then the Mellin transform M of f is given by*

$$M(s) = \int_0^1 f(v)v^s\,\mathrm{d}v = C \frac{\Gamma((nd + n)/2 + s)\Gamma(d/2 + s)}{\Gamma((nd + 1)/2 + s)\Gamma(nd/2 + s)}, \tag{18}$$

where C is determined by $M(0) = 1$. Furthermore,

(i) *if $n = 2N+1$ is odd then M is a rational function of the form*

$$M(s) = C\frac{(Nd + (d+1)/2 + s)_N}{(d/2+s)_{Nd}} = \sum_{k=0}^{Nd-1} \frac{A_k}{d/2+k+s}$$

and $f(v) = \sum_{k=0}^{Nd-1} A_k v^{d/2+k-1}$, where, for $k = 0, \dots, Nd-1$, we have

$$A_k = (-1)^k C \frac{(Nd + 1/2 - k)_N}{k!\,(Nd-k-1)!};$$

(ii) *if $n = 2N$ is even then*

$$M(s) = C\frac{\Gamma(d/2+s)}{\Gamma(Nd+1/2+s)}(s+Nd)_N$$

and f is a mixture of the beta densities

$$f(v)\,\mathrm{d}v = \sum_{k=0}^{N} r_k \beta\left(\frac{d}{2}+k, Nd - \frac{d-1}{2} - k\right)(\mathrm{d}v),$$

with the weights $r_0, \dots, r_N$, that are positive, satisfy $\sum_{k=0}^{N} r_k = 1$, and have the explicit value

$$r_k = \frac{1}{(Nd)_N}\left(\frac{d}{2}\right)_k \left(d\left(N - \frac{1}{2}\right)\right)_{N-k} \binom{N}{k}, \qquad k = 0, \dots, N. \tag{19}$$

Proof. With the notation X, Y, Z used in (17) we can claim that $\mathbb{E}(X^s)\mathbb{E}(Z^s) = \mathbb{E}(Y^s)$ for $s > 0$, which implies (18). When $n = 2N+1$ is odd the A_k are computed by partial fraction expansion and $1/(d/2+k+s)$ is the Mellin transform of the function $v^{d/2+k-1}\mathbf{1}_{(0,1)}(v)$.

When $n = 2N$ the situation is more complicated. Since $((x)_k/k!)_{k\geq 0}$ is a basis for real polynomials, every polynomial $P(x)$ of degree N can be written as

$$P(x) = \sum_{k=0}^{N} p_k \frac{(x)_k}{k!}$$

for some coefficients $p_0, \dots, p_N$. This implies that we have $P(-j) = \sum_{k=0}^{j}(-1)^k p_k \binom{j}{k}$ that we invert into $p_k = \sum_{j=0}^{k}(-1)^j P(-j)\binom{k}{j}$. We now apply these remarks to the polynomial $P(x) = (x+A)_N/(Nd)_N$, with $A = d(N - \frac{1}{2})$. Taking $x = s + d/2$, we obtain $P(s+d/2) = (s+Nd)_N/(Nd)_N$, thus,

$$\frac{(s+Nd)_N}{(Nd)_N} = \sum_{k=0}^{N} r_k \frac{(s+d/2)_k}{(d/2)_k}. \tag{20}$$

We proceed to an explicit calculation of the r_ks as follows:

$$r_k = \frac{1}{(Nd)_N}\left(\frac{d}{2}\right)_k \sum_{j=0}^{k} \frac{(-1)^j}{j!}\frac{1}{(k-j)!}(A-j)_N = \frac{1}{(Nd)_N}\left(\frac{d}{2}\right)_k (A)_{N-k} a_k,$$

where

$$a_k = \sum_{j=0}^{k} (-1)^j \frac{(A-j)_j}{j!} \frac{(A+N-k)_{k-j}}{(k-j)!}.$$

For computing the a_ks we use their generating function and the change of index $k' = k - j$:

$$\begin{aligned}\sum_{k=0}^{\infty} a_k z^k &= \sum_{k=0}^{\infty}\sum_{j=0}^{k}(-1)^j \frac{(A-j)_j}{j!} z^j \frac{(A+N-k)_{k-j}}{(k-j)!} z^{k-j} \\ &= \sum_{j=0}^{\infty}(-1)^j \frac{(A-j)_j}{j!} z^j \sum_{k'=0}^{\infty} \frac{(A+N-j-k')_{k'}}{k'!} z^{k'} \\ &= \sum_{j=0}^{\infty}(-1)^j \frac{(A-j)_j}{j!} z^j (1+z)^{A+N-j} \\ &= (1+z)^{A+N-1} \sum_{j=0}^{\infty} \frac{(A-j)_j}{j!} \left(\frac{-z}{1+z}\right)^j \\ &= (1+z)^{A+N-1}\left(1 - \frac{z}{1+z}\right)^{A-1} \\ &= (1+z)^N,\end{aligned}$$

which proves $a_k = \binom{N}{k}$ and (19). Finally, setting $s = 0$ in (20) we have $\sum_{k=0}^{N} r_k = 1$. To conclude the proof, we now have, from (18) and (20),

$$M(s) = C \sum_{k=0}^{N} r_k \frac{\Gamma(d/2+s)}{\Gamma(Nd+1/2+s)} \frac{(s+d/2)_k}{(d/2)_k},$$

where $C = (\Gamma(Nd+1/2))/(\Gamma(d/2))$ and we observe that

$$C \frac{\Gamma(d/2+s)}{\Gamma(Nd+1/2+s)} \frac{(s+d/2)_k}{(d/2)_k} = \frac{\Gamma(Nd+1/2)}{\Gamma(d/2+k)} \frac{\Gamma(d/2+k+s)}{\Gamma(Nd+1/2+s)}$$

is the Mellin transform of $\beta(d/2+k, Nd-(d-1)/2-k)$.

For the case where d is even and $q = 1$.

The next theorem (Theorem 5) is inspired by Kolesnik (2009) who computes the distribution of W for $d = 6$ when W is a Dirichlet random walk governed by $\mathcal{D}(1, 1, \ldots, 1)$ but when n is random and Poisson distributed. More specifically, let $Y_0, \ldots, Y_n, \ldots$ be i.i.d. RV such that $\Pr(Y_0 > y) = \mathrm{e}^{-cy}$, let $N(t)$ be a Poisson process such that $\mathbb{E}(N(t)) = \lambda t$ and let $\Theta_0, \ldots, \Theta_n, \ldots$ be i.i.d. and uniform on the unit sphere of $\mathbb{R}^d$. Denote

$$X(t) = \sum_{i=0}^{N(t)} Y_i \Theta_i, \qquad S(t) = \sum_{i=0}^{N(t)} Y_i, \qquad W(t) = \frac{X(t)}{S(t)}.$$

Kolesnik computes the distribution of $X(t)$ for $d = 6$. Here we are rather interested in the distribution of $W(t)|N(t) = n-1$, which is independent of $S(t)|N(t) = n-1$.

Before stating Theorem 5, we need a presentation of the hypergeometric function ${}_2F_1(\frac{1}{2}, 1; D; z)$ when $D \geq 2$ is a positive integer. The main point of Proposition 5 is that this function is actually a polynomial $\sum_{k=1}^{D-1} B_k G(z)^k$ with respect to the function G defined by (5).

Proposition 5. *Let D be a positive integer ≥ 2. Then, for $0 < u < \frac{1}{2}$,*

$$ {}_2F_1\left(\frac{1}{2}, 1; D; 4u - 4u^2\right) = \frac{(D-1)!B_D(u)}{4^{D-1}(1-u)^{D-1}} = \sum_{k=1}^{D-1} \frac{B_k}{(1-u)^k}, \tag{21} $$

where $B_D(u)$ is a polynomial of degree $D-2$ defined as follows: if $A_D(u) = u^{D-1}B_D(u)$ then $A'_D(u) = 4(1-2u)A_{D-1}(u)$, $A_D(0) = 0$, and $A_1(u) = 1/(1-2u)$. In particular, $B_1 = 1$ for $D = 2$, $B_1 = \frac{4}{3}$ and $B_2 = -\frac{1}{3}$ for $D = 3$, and $B_1 = \frac{8}{5}$, $B_2 = -\frac{7}{10}$ and $B_3 = \frac{1}{10}$ for $D = 4$.

Proof. For $|z| < 1$, we have

$$ {}_2F_1\left(\frac{1}{2}, 1; D; z\right) = (D-1)!\, \frac{1}{z^{D-1}} H_D(z), $$

where

$$ H_D(z) = \sum_{n=0}^{\infty} \frac{(1/2)_n}{(n+D-1)!}\, z^{n+D-1}. $$

Therefore, $H_D(0) = H'_D(0) = \cdots = H_D^{(D-2)}(0) = 0$ and, for $0 \leq i < D$, we have

$$ \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^i H_D(z) = H_{D-i}(z), \qquad \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^{D-1} H_D(z) = H_1(z) = \frac{1}{\sqrt{1-z}}. $$

We now show by induction on $D > 1$ that $u \mapsto A_D(u) = H_D(4u - 4u^2)$ is a polynomial of degree $2D-3$ and valency $D-1$. Note that $A_1(u) = H_1(4u-4u^2) = 1/(1-2u)$. Since, for $D \geq 2$, $A'_D(u) = 4(1-2u)A_{D-1}(u)$, and since $A_D(0) = 0$, then $A_2(u) = 4u$ and the property is true for $D = 2$.

Suppose that the property is true for some $2 \leq D$. Then $A'_{D+1}(u) = 4(1-2u)A_D(u)$ and the induction hypothesis shows that A_{D+1} is a polynomial of degree $2D-1$. Introduce the polynomial B_D such that $A_D(u) = u^{D-1}B_D(u)$. It exists from the induction hypothesis. Since $A_{D+1}(0) = 0$, we have $A_{D+1}(u) = 4\int_0^u (1-2v)v^{D-1}B_D(v)\,\mathrm{d}v$. Therefore, the valency of A_{D+1} is D and the induction hypothesis is extended. The remainder is straightforward.

Theorem 5. *Let $W_{n,2D}$ be a Dirichlet random walk in the unit sphere of $\mathbb{R}^{2D}$ governed by $\mathcal{D}(1, 1, \ldots, 1)$ (n times), where D is a positive integer ≥ 2. Define the sequence $\{p_i, i = n, n+1, \ldots, n(D-1)\}$ by*

$$ \left(\sum_{k=1}^{D-1} B_k z^k\right)^n = \sum_{i=n}^{n(D-1)} p_i z^i, $$

where $B_1, \ldots, B_{D-1}$ are the numbers defined by (21). Then

$$ T_n(W_{n,d})(y) = \sum_{i=n}^{n(D-1)} p_i\, {}_2F_1\left(\frac{i}{2}, \frac{i+1}{2}; i+1; \|y\|^2\right). $$

In particular, if $R_{n,d} = \|W_{n,d}\|$ we have the moments of $R^2_{n,d}$,

$$ \mathbb{E}(R^{2k}_{n,d}) = \frac{(D)_k}{(n)_{2k}} \sum_{i=n}^{n(D-1)} p_i \frac{(i)_{2k}}{(i+1)_k} \tag{22} $$

as well as the Mellin transform of $R^2_{n,d}$,

$$\mathbb{E}(R^{2s}_{n,d}) = \frac{(n-1)!}{(D-1)!}\sum_{i=n}^{n(D-1)} i p_i \frac{(s)_D}{(s)_{i+1}}\frac{(2s)_i}{(2s)_n}. \tag{23}$$

Proof. From Theorem 1 we have

$$\mathbb{E}\bigg(\frac{1}{(1+\langle y, W_{n,d}\rangle)^n}\bigg) = \big({}_2F_1\big(\tfrac{1}{2}, 1; D; \|y\|^2\big)\big)^n.$$

From Proposition 5 we have

$$\big({}_2F_1\big(\tfrac{1}{2}, 1; D; 4u-4u^2\big)\big)^n = \sum_{i=n}^{n(D-1)} p_i \frac{1}{(1-u)^i}.$$

With the notation (5) for G and using (6), we obtain

$$\begin{aligned}\big({}_2F_1\big(\tfrac{1}{2}, 1; D; z\big)\big)^n &= \sum_{i=n}^{n(D-1)} p_i G(z)^i \\ &= \sum_{i=n}^{n(D-1)} p_i\, {}_2F_1\bigg(\frac{i}{2}, \frac{i+1}{2}; i+1; z\bigg) \\ &= \sum_{k=0}^{\infty}\frac{z^k}{k!}\bigg(\sum_{i=n}^{n(D-1)} p_i \frac{(i/2)_k((i+1)/2)_k}{(i+1)_k}\bigg) \\ &= \sum_{k=0}^{\infty}\frac{z^k}{k!}\bigg(\sum_{i=n}^{n(D-1)} p_i \frac{(i)_{2k}}{2^{2k}(i+1)_k}\bigg).\end{aligned}$$

We write $W_{n,d} = R_{n,d}\Theta$ and apply (12) to (22). For proving (23) we rewrite (22) as follows:

$$\begin{aligned}\mathbb{E}(R^{2k}_{n,d}) &= \sum_{i=n}^{n(D-1)} p_i \frac{\Gamma(D+k)}{\Gamma(D)\Gamma(k)}\frac{\Gamma(i+1)\Gamma(k)}{\Gamma(i+1+k)}\frac{\Gamma(i+2k)}{\Gamma(i)\Gamma(2k)}\frac{\Gamma(n)\Gamma(2k)}{\Gamma(n+2k)} \\ &= \frac{(n-1)!}{(D-1)!}\sum_{i=n}^{n(D-1)} i p_i \frac{(k)_D}{(k)_{i+1}}\frac{(2k)_i}{(2k)_n}.\end{aligned}$$

Therefore, (23) is correct when s is a nonnegative integer. Now observe that the right-hand side $H(s)$ of (23) is a rational fraction. More specifically, it is a linear combination of the rational fractions $((s)_D/(s)_{i+1})((2s)_i/(2s)_n)$ and the difference between the degree of the denominator and that of the numerator is $n+1-D$ and does not depend on i. If $n+1 > D$ the rational fraction H is a linear combination of partial fractions of the form $1/(s+j) = \int_0^1 v^{s+j-1}\,\mathrm{d}v$, and H is the Mellin transform of a polynomial P_H restricted to $(0, 1)$. Since $\mathbb{E}(R^{2k}_{n,d}) = \int_0^1 v^k P_H(v)\,\mathrm{d}v$, this implies that

$$R^2_{n,d} \sim P_H(v)\mathbf{1}_{(0,1)}(v)\,\mathrm{d}v$$

and, therefore, (23) holds for all $s \geq 0$. If $n+1 \leq D$ the rational fraction H is the sum of a polynomial Q of degree $\leq D-n-1$ and of a linear combination of partial fractions of the form $1/(s+j) = \int_0^1 v^{s+j-1}\,\mathrm{d}v$.

We claim that Q is the zero polynomial. If $\deg Q > 0$ then $k \mapsto \mathbb{E}(R^{2k})$ is unbounded, which is impossible since $0 \le R_{n,d} \le 1$. If Q is a nonzero constant c, this implies that $\lim_{k\to\infty} \mathbb{E}(R_{n,d}^{2k}) = c$ which means that $\Pr(R_{n,d}^2 = 1) = c$. This fact is impossible since $R_{n,d} = \|W_{n,d}\| = \|X_1\Theta_1 + \cdots + X_n\Theta_n\| = 1$ implies that all Θ_j are equal, which has zero probability if $d > 1$. Therefore, $Q = 0$ and we conclude that (23) holds as in the case where $n + 1 > D$.

Example. *For the case where* $q = 1$, $d = 6$, *and* $n = 2$. We have seen, in Proposition 5 that since $D = d/2 = 3$, we have

$$ {}_2F_1\big(\tfrac{1}{2}, 1; 3; z\big) = \tfrac{4}{3}G(z) - \tfrac{1}{3}G^2(z) $$

which implies

$$ \big({}_2F_1\big(\tfrac{1}{2}, 1; 3; z\big)\big)^2 = \tfrac{16}{9}G^2(z) - \tfrac{8}{9}G^3(z) + \tfrac{1}{9}G^4(z), $$

that is, to say $p_2 = \frac{16}{9}$, $p_3 = -\frac{8}{9}$, $p_4 = \frac{1}{9}$. Careful calculations from (23) gives

$$ \mathbb{E}(R_{6,2}^{2s}) = \frac{8}{3+s} - \frac{20}{3}\frac{1}{4+s} = \int_0^1 v^s\left(8v^2 - \frac{20}{3}v^3\right) \mathrm{d}v. $$

In other terms $R_{6,2} \sim (8v^2 - \frac{20}{3}v^3)\mathbf{1}_{(0,1)}(v)\,\mathrm{d}v$. This is equivalent to Equation (13) of Kolesnik (2009). Applying the same method as above for $n = 3$ would provide Equation (15) of Kolesnik.

5. The case $\mathcal{D}(q_1, q_2)$

Since the Dirichlet distribution $\mathcal{D}(q_1, q_2)$ is nothing but the distribution of $(X, 1-X)$, where $X \sim \beta(q_1, q_2)$, it is almost trivial to directly study the Dirichlet random walks for $n = 2$, but it offers the opportunity to check general formulae in this particular case.

Proposition 6. *Let* $X \sim \beta(q_1, q_2)$, Θ_1, *and* Θ_2 *be three independent random variables such that the* Θ_i *are uniformly distributed on the unit sphere of* $\mathbb{R}^d$. *Let* $R_{2,d} = \|W_{2,d}\|$, *where* $W_{2,d}$ *is the Dirichlet random walk, such that*

$$ W_{2,d} = X\Theta_1 + (1-X)\Theta_2. \tag{24} $$

Then the Mellin transform of $H_{2,d} = 1 - R_{2,d}^2$ *is*

$$ \mathbb{E}(H_{n,d}^s) = C\frac{\Gamma(q_1+s)\Gamma(q_2+s)\Gamma((d-1)/2+s)}{\Gamma((q_1+q_2)/2+s)\Gamma((q_1+q_2+1)/2+s)\Gamma(d-1+s)}, $$

where C *is the normalizing constant such that the right-hand side is 1 when* $s = 0$.

Proof. Setting $Z = \frac{1}{2}(1 - \langle\Theta_2, \Theta_1\rangle)$, we have

$$ 1 - H_{2,d} = R_{n,d}^2 \sim \|X\Theta_1 + (1-X)\Theta_2\|^2 = 1 - 4X(1-X)Z. $$

Since X and Z are independent, we have $\mathbb{E}(H_{2,d}^s) = \mathbb{E}((4X(1-X))^s)\mathbb{E}(Z^s)$.

Since $X \sim \beta(q_1, q_2)$, then

$$ \mathbb{E}((4X(1-X))^s) = C_1\frac{\Gamma(q_1+s)\Gamma(q_2+s)}{\Gamma((q_1+q_2)/2+s)\Gamma((q_1+q_2+1)/2+s)} $$

with a suitable normalizing constant C_1. From Lemma 1, conditioning first on Θ_2, we know that $\langle\Theta_2, \Theta_1\rangle$ is symmetric and that $\langle u, \Theta_1\rangle^2 \sim \beta(\frac{1}{2}, (d-1)/2)$; this easily implies that

$Z \sim \beta((d-1)/2, (d-1)/2)$. From this the computation of $\mathbb{E}(Z^s)$ can be done and this leads to the proof of the proposition.

Example. If we apply the above proposition to $\mathcal{D}(q, q)$ and $\mathcal{D}(q, q+1)$, we obtain in both cases the same result:

$$\mathbb{E}(H_{2,d}^s) = \frac{\Gamma(d-1)\Gamma(q+1/2)\Gamma(q+s)\Gamma((d-1)/2+s)}{\Gamma(q+1/2+s)\Gamma(d-1+s)\Gamma(q)\Gamma((d-1)/2)}.$$

The above equality states that $H_{2,d} \sim XY$, where $X \sim \beta((d-1)/2, (d-1)/2)$ and $Y \sim \beta(q, \frac{1}{2})$ are independent. When $q = d-1$,

$$\mathbb{E}(H_{2,d}^s) = \frac{\Gamma(d-1/2)}{\Gamma((d-1)/2)} \frac{\Gamma((d-1)/2+s)}{\Gamma(d-1/2+s)}$$

and, therefore, $H_{2,d} \sim \beta((d-1)/2, d/2)$ and $R_{2,d}^2 \sim \beta(d/2, (d-1)/2)$, which is in agreement with Proposition 2.

Corollary 2. *If X is uniform on $(0, 1)$ the Mellin transform of $H_{2,d} = 1 - \|W_{2,d}\|^2$, as defined by (24), is as follows:*

(i) *if $d = 2$ then $\mathbb{E}(H_{2,d}^s) = 1/(1+2s)$, which implies $H_{2,d} \sim (1/2h^{1/2})\mathbf{1}_{(0,1)}(h)\,\mathrm{d}h$ and $R_{2,d}^2 \sim \beta(1, \frac{1}{2})$;*

(ii) *if $D \geq 2$ is an integer then*

$$\mathbb{E}(H_{2,2D}^s) = \frac{(2D-2)!}{(3/2)_{D-2}} \frac{(s+3/2)_{D-2}}{(s+1)_{2D-2}};$$

(iii) *if $D \geq 1$ is an integer then*

$$\mathbb{E}(H_{2,2D+1}^s) = \frac{\Gamma(3/2)(2D-1)!}{(D-1)!} \frac{\Gamma(s+1)}{\Gamma(s+3/2)} \frac{1}{(s+D)_D}.$$

Example. If $d = 6$, part (ii) shows that

$$\begin{aligned}\mathbb{E}(H_{2,6}^s) &= 16\frac{s+3/2}{(s+1)(s+2)(s+3)(s+4)} \\ &= \tfrac{20}{3}\frac{1}{s+4} - \frac{12}{s+3} + \frac{4}{s+2} + \tfrac{4}{3}\frac{1}{s+1} \\ &= \int_0^1 \left(\tfrac{20}{3}h^3 - 12h^2 + 4h + \tfrac{4}{3}\right)h^s\,\mathrm{d}h.\end{aligned}$$

Therefore the distribution of $R_{2,6}^2 = 1 - H_{2,6}$ is

$$\left(\tfrac{20}{3}(1-v)^3 - 12(1-v)^2 + 4(1-v) + \tfrac{4}{3}\right)\mathbf{1}_{(0,1)}(v)\,\mathrm{d}v = \left(8v^2 - \tfrac{20}{3}v^3\right)\mathbf{1}_{(0,1)}(v)\,\mathrm{d}v.$$

This result, of course, coincides with the result of the last example in Section 4.

6. Limits of Dirichlet walks

It is quite natural to ask what happens in Theorem 1 as $n \to \infty$ and $q_j = Q/n$ for $j = 1, \ldots, n$, for some positive constant $Q > 0$. The answer turns out to be rather surprising.

Theorem 6. *Let $q_j = Q/n$ for $j = 1, \dots, n$ in the definition of $W_{n,d}$. The following hold.*

(i) *If $n \to \infty$ then, for $y \in \mathring{B}_d$, we have $T_Q(W_{n,d})(y) \to \mathrm{e}^{L_d(\|y\|^2)}$, where*

$$L_d(z) = \frac{1}{2} \sum_{k=1}^{\infty} \frac{(1/2)_k}{(d/2)_k} \frac{z^k}{k}.$$

(ii) *For $y \in \mathring{B}_d$ and Θ uniform on the sphere S_{d-1}, denote by U_1 the first component of Θ. Then*

$$L_d(\|y\|^2) = -\mathbb{E}\{\log(1 + \|y\|U_1)\}.$$

(iii) *We have, that $W_{n,d}$ tends weakly to $W^Q_{\infty,d}$, which is characterized by*

$$T_Q(W^Q_{\infty,d})(y) = \exp\{-Q\mathbb{E}\{\log(1 + \|y\|U_1)\}.$$

(iv) *Let $\{\Theta_i,\ i = 1, 2, \dots\}$ be a sequence of i.i.d. RV's uniform on the unit sphere S_{d-1} and let $\{\Pi_i,\ i = 1, 2, \dots\}$ be obtained from an independent i.i.d. sequence $\{Y_i, i = 1, 2, \dots\}$ of $\beta(1, Q)$ random variables, through $\Pi_1 = Y_1$ and $\Pi_n = Y_n \prod_{i=1}^{n-1}(1 - Y_i)$. Then*

$$W^Q_{\infty,d} \sim \sum_{i=1}^{\infty} \Pi_i \Theta_i.$$

Proof. Statement (i) results from Theorem 1 and the development of

$${}_2F_1\left(\frac{Q}{2n}, \frac{Q}{2n} + \frac{1}{2}; \frac{d}{2}; z\right) = 1 + \frac{Q}{n} L_d(z) + o\left(\frac{1}{n}\right),$$

from which we obtain

$$\left({}_2F_1\left(\frac{Q}{2n}, \frac{Q}{2n} + \frac{1}{2}; \frac{d}{2}; \|y\|^2\right)\right)^n \to \mathrm{e}^{QL_d(\|y\|^2)}.$$

Moreover, recalling that $W_{n,d} = \sum_{i=1}^n X_i^{(n)} \Theta_i$, with $X^{(n)} = (X_1^{(n)}, \dots, X_n^{(n)}) \sim \mathcal{D}(Q/n, \dots, Q/n)$, we have, for $y \in \mathring{B}_d$,

$$T_Q(W_{n,d})(y) = \mathbb{E}\left(\frac{1}{(\sum_{i=1}^n X_i^{(n)}(1 + \langle y, \Theta_i \rangle))^Q}\right),$$

which, by using (4), can be written as

$$\mathbb{E}\left(\frac{1}{\prod_{i=1}^n (1 + \langle y, \Theta_i \rangle)^{Q/n}}\right) = \left(\mathbb{E}\left(\frac{1}{(1 + \|y\|U_1)^{Q/n}}\right)\right)^n = (\mathbb{E}(\mathrm{e}^{X/n}))^n,$$

where $X = \log(1 + \|y\|U_1)^{-Q}$. For fixed $y \in \mathring{B}_d$, X is a bounded random variable. Thus, $\log(\mathbb{E}(\mathrm{e}^{X/n}))^n$ converges to $\mathbb{E}(X) = -Q\mathbb{E}(\log(1 + \|y\|U_1))$, which is Theorem 6(ii).

For Theorem 6(iii) observe that the sequence $\{W_{n,d}, n = 1, 2, \dots\}$ is clearly tight, since all these variables have support in S_{d-1}. So take any converging sequence and call $W^Q_{\infty,d}$ its limit. Since the function $w \mapsto f_y(w) = 1/(1 + \langle y, w \rangle)^Q$ defined in the unit sphere for $y \in \mathring{B}_d$ and $Q > 0$ is bounded and continuous, we have (along this subsequence)

$$\mathbb{E}\left(\frac{1}{(1 + \langle y, W_{n,d} \rangle)^Q}\right) \quad \text{converges to} \quad \mathbb{E}\left(\frac{1}{(1 + \langle y, W^Q_{\infty,d} \rangle)^Q}\right).$$

Therefore,

$$\mathbb{E}\Big(\frac{1}{(1+\langle y, W^Q_{\infty,d}\rangle)^Q}\Big) = \exp\{QL_d(\|y\|^2)\}$$

and since, as mentioned in the introduction, the knowledge of $T_Q(W^Q_{\infty,d})$ characterizes the law of $W^Q_{\infty,d}$ we have uniqueness of the law of $W^Q_{\infty,d}$ and the stated representation.

For Theorem 6(iv) we first observe that Theorem 6(ii) says that $W^Q_{\infty,d}$ is the mean vector of a Dirichlet random measure (sometimes called a Dirichlet process) with parameter measure equal to Q times the uniform distribution on S_{d-1}. To see this, apply the classical result about the Stieltjes transform of the mean of a Dirichlet random measure from Cifarelli and Regazzini (1979), (1990), conveniently stated in Theorem 2.1 of Lijoi and Prunster (2009). Even if this result is stated for a Dirichlet random measure concentrated on the real line, passing from $\mathbb{R}$ to $\mathbb{R}^d$ is standard: it is enough to apply the theorem to the real mean $\langle v, W^Q_{\infty,d}\rangle$ of the Dirichlet process generated by the measure $Q\alpha(\mathrm{d}u)$, where α is the distribution of U_1, $v \in S_{d-1}$ being arbitrary. Finally, recall from Sethuraman (1994) that $P = \sum_{i=1}^{\infty} \delta_{\Theta_i} \Pi_i$ is actually a Dirichlet random measure governed by Q times the uniform probability on S_{d-1}. Therefore $W^Q_{\infty,d} \sim \int_{S_{d-1}} \theta P(\mathrm{d}\theta)$ since both random variables have the same T_Q-transform given by Theorem 6(ii). Thus, the representation of $W^Q_{\infty,d}$ given in Theorem 6(iv) is obtained.

Theorem 6(iv) says that this limit of Dirichlet walks can be obtained as a walk with an infinite number of steps. This is a particular instance of a more general result appearing in Hjort and Ongaro (2005), (2006).

Finally, we investigate the random vectors in $\mathbb{R}^d$ whose Q-transform has the form $\exp\{QL_d(\|y\|^2)\}$. If $d = 1$, we obtain

$$\exp\{QL_1(\|y\|^2)\} = \frac{1}{(1-\|y\|^2)^{Q/2}}.$$

We apply Proposition 1 to this equation and it yields that the corresponding limiting distribution has the square radius $(R^Q_{\infty,1})^2$ distributed as $\beta(\frac{1}{2}, Q/2)$ and, therefore, the distribution of $W^Q_{\infty,1} = \pm R$ is

$$\frac{1}{B(Q/2, Q/2)}(1-w^2)^{Q/2-1}\mathbf{1}_{(-1,1)}(w)\,\mathrm{d}w.$$

This result is not easy to obtain as a limiting case of Proposition 4.

For $d = 2$, (8) shows that $\exp\{QL_2(\|y\|^2) = G(\|y\|^2)^Q\}$, where G is defined by (5). This time the distribution of $(R^Q_{\infty,2})^2$ is $\beta(\frac{1}{2}, Q)$.

What about $d \geq 3$? The function L_d, when d is odd, is explicitly computable. Here is the example for $d = 3$:

$$2L_3(z) = \sum_{k=1}^{\infty} \frac{z^k}{k(1+2k)} = \sum_{k=1}^{\infty} \frac{z^k}{k} - 2\sum_{k=1}^{\infty} \frac{z^k}{1+2k}.$$

Introducing

$$f(x) = \sum_{k=1}^{\infty} \frac{x^{2k+1}}{1+2k} = -x + \frac{1}{2}\log\frac{1+x}{1-x}, \qquad f'(x) = -1 + \frac{1}{2(x+1)} - \frac{1}{2(x-1)},$$

we obtain

$$L_3(\|y\|^2) = 1 + \tfrac{1}{2}\{(1-\|y\|)\log(1-\|y\|) - (1+\|y\|)\log(1+\|y\|)\}. \tag{25}$$

However, finding the distribution of $W^Q_{\infty,3}$ in the unit sphere of $\mathbb{R}^3$ such that

$$\mathbb{E}\bigg(\frac{1}{(1+\langle y, W^Q_{\infty,3}\rangle)^Q}\bigg) = \exp(QL_3(\|y\|^2)) \tag{26}$$

or, equivalently, finding the distribution of $\|W^Q_{\infty,3}\|^2$, seems to be a challenging open problem. Let us give some details about it. Consider the random vector Θ such that Θ is uniform on the unit sphere of $\mathbb{R}^3$ and denote by U its first coordinate. Then U is uniform on $(-1, 1)$ by the Archimedes theorem. Denoting $R = \|W^Q_{\infty,3}\|$ we claim that

$$\mathbb{E}\bigg(\frac{1}{(1+\langle y, W^Q_{\infty,3}\rangle)^Q}\bigg) = \mathbb{E}\bigg(\frac{1}{(1+\|y\|RU\rangle)^Q}\bigg) = \mathbb{E}(f_Q(R\|y\|)),$$

where $f_Q(t) = \frac{1}{2}\int_{-1}^{1}(1+ut)^{-Q}\,\mathrm{d}u$ is defined for $t \in (-1, 1)$ and is easy to compute. Replacing for simplicity $\|y\|$ by t in (25) and using (26), the problem of the explicit description of the distribution of $W^Q_{\infty,3}$ is now reduced to the following problem of harmonic analysis. For fixed $Q > 0$ find the unique distribution for $R \in [0, 1]$ such that, for all $t \in (0, 1)$,

$$\mathbb{E}(f_Q(Rt)) = \exp\bigg(Q\sum_{k=1}^{\infty}\frac{t^{2k}}{2k(2k+1)}\bigg) = (\exp(1-t)^{(1-t)/2t}(1+t)^{(-1-t)/2t})^Q.$$

For instance, for $Q = 1$, we want R such that

$$\mathbb{E}\bigg(\frac{1}{2tR}\log\frac{1+tR}{1-tR}\bigg) = \exp(1-t)^{(1-t)/2t}(1+t)^{(-1-t)/2t}.$$

7. Dirichlet infinite divisibility and Dirichlet semi groups

Limits of Dirichlet random walks or, in view of Theorem 6, means of Dirichlet random measures are examples of the following property of infinite divisibility.

Definition. Let $W \sim \mu$ be a random variable on the unit sphere B_d of $\mathbb{R}^d$. We say that W or μ is *Dirichlet infinitely divisible* of type $Q > 0$ if, for all n, there exists a probability measure ν_n on the unit sphere such that the following occurs. If $Y = (Y_1, \dots, Y_n) \sim \mathcal{D}(Q/n, \dots, Q/n)$ and, independently, $W_1, \dots, W_n$ are i.i.d. with distribution ν_n, then

$$Y_1W_1 + \cdots + Y_nW_n \sim \mu.$$

Here is the following equivalence property.

Theorem 7. *The three following properties are equivalent:*

(i) *W is Dirichlet infinitely divisible of type Q;*

(ii) *for all n there exists a random variable $\tilde{W}^{(n)}$ on the unit sphere such that, for all $y \in \mathring{B}_d$,*

$$T_Q(W)(y) = [T_{Q/n}(\tilde{W}^{(n)})(y)]^n;$$

(iii) *there exists a random variable $\hat{W}$ on the unit sphere such that, for all $y \in \mathring{B}_d$, we have*

$$T_Q(W)(y) = \exp\{-Q\mathbb{E}(\log(1+\langle y, \hat{W}\rangle))\}.$$

Proof. For (i) to imply (ii) observe that, by assumption, $W \sim Y_1 W_1^{(n)} + \cdots + Y_n W_n^{(n)}$, where we have made explicit the dependence on n of the W_i. Thus,

$$\begin{aligned}
\mathbb{E}\bigg(\frac{1}{(1+\langle y, W\rangle)^Q}\bigg) &= \mathbb{E}\bigg(\frac{1}{(1+\langle y, Y_1 W_1^{(n)} + \cdots + Y_n W_n^{(n)}\rangle)^Q}\bigg) \\
&= \mathbb{E}\bigg[\mathbb{E}\bigg(\frac{1}{(1+\langle y, Y_1 W_1^{(n)} + \cdots + Y_n W_n^{(n)}\rangle)^Q} \,\bigg|\, W_1^{(n)}, \ldots, W_n^{(n)}\bigg)\bigg] \\
&= \mathbb{E}\bigg(\frac{1}{(1+\langle y, W_1^{(n)}\rangle)^{Q/n}} \cdots \frac{1}{(1+\langle y, W_n^{(n)}\rangle)^{Q/n}}\bigg) \\
&= \bigg[\mathbb{E}\bigg(\frac{1}{(1+\langle y, W_1^{(n)}\rangle)^{Q/n}}\bigg)\bigg]^n;
\end{aligned}$$

thus by taking $\tilde{W}^{(n)}$ equal in law to $W_1^{(n)}$ the representation is obtained.

Property (ii) implies property(i). This is obtained since $T_Q(W)$ characterizes the law of W.

Property (ii) implies property (iii). Again $\{\tilde{W}^{(n)}, n = 1, 2, \ldots\}$ is a tight sequence so we may assume that it has a limit $\hat{W}$. Arguing, as in the previous proof, we conclude that

$$\lim_{n\to\infty} [\mathbb{E}\bigg(\frac{1}{(1+\langle y, \tilde{W}^{(n)}\rangle)^{Q/n}}\bigg)]^n = \exp\{-Q\mathbb{E}(\log(1+\langle y, \hat{W}\rangle))\}.$$

Property (iii) implies property (ii). For any $q > 0$, the mean W^q of a Dirichlet random measure with parameter measure q times the law of $\hat{W}$ has the property

$$T_q(\hat{W}^q)(y) = \exp\{-q\mathbb{E}(\log(1+\langle y, \hat{W}\rangle))\}.$$

As a consequence Statement (ii) is obtained by taking $\tilde{W}^{(n)}$ equal in law to $W^{q/n}$.

Thus from the Dirichlet infinite divisibility property a stronger property follows, via the equivalent Statement (iii) of Theorem 7. Any Dirichlet infinite divisible random variable W^Q of type $Q > 0$ is indeed embedded in a Dirichlet semigroup of laws $\{\mu_q, q > 0\}$, defined according to the following definition, in the sense that $W^Q \sim \mu_Q$. Note that this semigroup is weakly continuous at 0 since μ_q converges to the distribution of $\hat{W}$ as $q \to 0$.

Definition. Let $Q \mapsto \mu_Q$ be a map from $(0, \infty)$ to the set of probability measures in the unit sphere B_d of $\mathbb{R}^d$. We say that this map is a *Dirichlet semigroup* if, for all n and all $q_1, \ldots, q_n > 0$, the following occurs. Taking $W_1, \ldots, W_n, Y$ independent such that $W_i \sim \mu_{q_i}$ and $Y = (Y_1, \ldots, Y_n) \sim \mathcal{D}(q_1, \ldots, q_n)$, then setting $Q = q_1 + \cdots + q_n$, we obtain

$$Y_1 W_1 + \cdots + Y_n W_n \sim \mu_Q.$$

For instance, when G is defined by (5),

$$T_Q(W_d^Q)(y) = G(\|y\|^2)^Q$$

describes $W_d^Q = R^Q\Theta$, with Θ uniform in S_{d-1} and $R^2 \sim \beta(d/2, Q+1-d/2)$, with $Q > d/2 - 1$. Thus, W_d^Q follows a hyperuniform law of type $2(Q+1)$. For $d = 2$, we have a Dirichlet semigroup and for $d = 1$ the restriction to $Q > 0$ is a Dirichlet semigroup as well. Similarly, the relation

$$T_Q(\bar{W}_d^Q)(y) = \frac{1}{(1-\|y\|)^{Q/2}}$$

describes $\bar{W}_d^Q = (\bar{R})^Q \Theta$, with Θ uniform distribution in S_{d-1} and

$$(\bar{R})^2 \sim \beta\left(\frac{d}{2}, \frac{Q+1-d}{2}\right)$$

, with $Q > d-1$. Thus, $\bar{W}_d^Q$ follows a hyperuniform law of type $Q+1$. Here we have a Dirichlet semigroup only for $d = 1$. However, by direct inspection of the parameters of these beta distributions it is seen that $\bar{W}_1^Q \sim W_1^{(Q-1)/2}$, and with a straightforward calculation we obtain $(1 + \bar{W}_1^Q)/2 \sim \beta(Q/2, Q/2)$. Thus, if such a family of symmetric beta distributions is reparametrized by $q = (Q-1)/2$, for $Q > 1$, it retains the Dirichlet semigroup property.

Note that, for any $d > 2$ in the first case and any $d > 1$ in the second, we still obtains a family of hyperuniform distributions with the Dirichlet semigroup property, but the index of the family runs on a parameter set of the form $(a, +\infty)$, with $a > 0$, which clearly remains an additive semigroup.

Acknowledgements

We thank A. D. Kolesnik, G. Le Caër, E. Orsingher, and M. Yor for interesting correspondence about the present paper. We are also grateful to the referee for his patient reading and his careful checking of our calculations.

References

Beghin, L. and Orsingher, E. (2010). Moving randomly amid scattered obstacles. *Stochastics* **82,** 201–229.

Borwein, J. M., Straub, A., Wan, J. and Zudilin, W. (2012). Densities of short uniform random walks. *Canad. J. Math.* **64,** 961–990.

Chamayou, J.-F. and Letac, G. (1994). A transient random walk on stochastic matrices with Dirichlet distributions. *Ann. Prob.* **22,** 424–430.

Cifarelli, D. M. and Regazzini, E. (1979). A general approach to Bayesian analysis of nonparametric problems. The associative mean values within the framework of the Dirichlet process. II. *Riv. Mat. Sci. Econom. Social.* **2,** 95–111 (in Italian).

Cifarelli, D. M. and Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.* **18,** 429–442. (Correction: **22** (1994), 1633–1634.)

Hjort, N. L. and Ongaro, A. (2005). Exact inference for random Dirichlet means. *Statist. Inference Stoch. Process.* **8,** 227–254.

Hjort, N. L. and Ongaro, A. (2006). On the distribution of random Dirichlet jumps. *Metron* **64,** 61–92.

Kolesnik, A. D. (2009). The explicit probability distribution of a six-dimensional random flight. *Theory Stoch. Process.* **15,** 33–39.

Le Caër, G. (2010). A Pearson random walk with steps of uniform orientation and Dirichlet distributed lengths. *J. Statist. Phys.* **140,** 728–751.

Le Caër, G. (2011). A new family of solvable Pearson–Dirichlet random walks. *J. Statist. Phys.* **144,** 23–45.

Lijoi, A. and Prünster, I. (2009). Distributional properties of means of random probability measures. *Statist. Surveys* **3,** 47–95.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4,** 639–650.