

# Around the root of random multidimensional quadtrees

Gilbert Labelle, Louise Laforest, Xavier Provençal

► **To cite this version:**

Gilbert Labelle, Louise Laforest, Xavier Provençal. Around the root of random multidimensional quadtrees. Chassaing, Philippe and others. Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities, 2006, Nancy, France. Discrete Mathematics and Theoretical Computer Science, DMTCS Proceedings vol. AG, Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities, pp.335-344, 2006, DMTCS Proceedings. <hal-01184686>

**HAL Id: hal-01184686**

**<https://hal.inria.fr/hal-01184686>**

Submitted on 17 Aug 2015

**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.

# Around the root of random multidimensional quadrees

Gilbert Labelle<sup>1,†</sup>, Louise Laforest<sup>1</sup> and Xavier Provençal<sup>1</sup>

<sup>1</sup> *LaCIM, Université du Québec à Montréal,  
Montréal (Québec) Canada H3C 3P8  
{labelle.gilbert, laforest.louise}@uqam.ca, provenca@math.uqam.ca*

We analyse the distribution of the root pattern of randomly grown multidimensional point quadrees. In particular, exact, recursive and asymptotic formulas are given for the expected arity of the root.

**Keywords:** point quadrees, asymptotics, random structures.

## 1 Introduction

A random quadtree is a data structure for multidimensional data access (5). More precisely, in dimension  $d$ , a sequence  $\mathcal{P} = (P_1, P_2, \dots, P_n)$  of points  $P_i$  in the unit cube  $[0, 1]^d \subseteq \mathbb{R}^d$  is represented by a  $d$ -dimensional quadtree  $\mathcal{T}$  constructed recursively by the following rules :

- If  $n = 0$ , ( $\mathcal{P} = \emptyset$ ), then  $\mathcal{T}$  is empty.
- If  $n > 0$ , then the first point  $P_1$  is made the root of the tree. The  $2^d$  root subtrees are then constructed recursively from the  $2^d$  sublists of points

$$\mathcal{P}_{00\dots 0}, \mathcal{P}_{00\dots 1}, \dots, \mathcal{P}_{\varepsilon_1\varepsilon_2\dots\varepsilon_d}, \dots, \mathcal{P}_{11\dots 1}, \quad \varepsilon = \varepsilon_1\varepsilon_2\dots\varepsilon_d \in \{0, 1\}^d, \quad (1.1)$$

defined by restricting  $\mathcal{P} \setminus P_1$  to the  $2^d$  hyperoctants indexed by the binary words  $\varepsilon = \varepsilon_1\varepsilon_2\dots\varepsilon_d$  of length  $d$  that are determined by the root node  $P_1$ .

The hyperoctant indexed by the word  $\varepsilon = \varepsilon_1\varepsilon_2\dots\varepsilon_d$  is, by convention, the one which contains the vertex  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d)$  of the unit cube  $[0, 1]^d$ . Each subtree is numbered by the decimal equivalent of the corresponding hyperoctant's index. For each node, the subtrees are drawn, from left to right, in ascending order. In the case of usual quadrees ( $d = 2$ ) the quadrants are indexed as in Figure 1(a). A sequence of random points and associated quadtree are given in Figure 1(b) and (c).

We define the *root pattern* of a  $d$ -dimensional quadtree  $\mathcal{T}$  by the set  $R$  of binary words  $\varepsilon_1\varepsilon_2\dots\varepsilon_d$  for which  $\mathcal{P}_{\varepsilon_1\varepsilon_2\dots\varepsilon_d}$  is not empty. Obviously, the cardinality of  $R$  is the arity of the root of  $\mathcal{T}$ . From example illustrated in Figure 1(b) and (c), one easily sees that the root pattern is

$$\{00, 01, 10\}. \quad (1.2)$$

We are interested here in the analysis of the root pattern of randomly grown quadrees built from  $n$  uniformly distributed independent random points of the unit hypercube  $[0, 1]^d$  in  $d$ -dimensional space.

Let  $Q_n[R]$  be the probability that such a randomly grown quadtree  $\mathcal{T}$  has a given root pattern  $R$ . Using inclusion-exclusion principle and  $d$ -dimensional integrals we have established in (10) that

$$Q_n[R] = \sum_{S \subseteq R} (-1)^{|R|-|S|} J_n[S] \quad (1.3)$$

where  $J_n[S]$  is the probability that the root pattern is a subset of  $S$  which is given by

$$J_n[S] = \int_0^1 \cdots \int_0^1 (f_S(t_1, t_2, \dots, t_d))^{n-1} dt_1 dt_2 \dots dt_d \quad (1.4)$$

<sup>†</sup>Research supported by the Natural Sciences and Engineering Research Council of Canada.

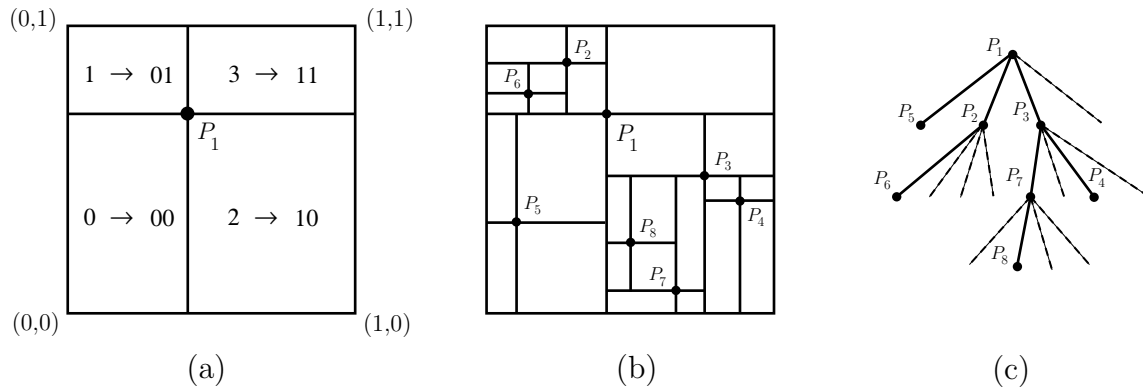


Fig. 1: (a) Quadrants (b) Sequence of random points (c) Associated quadtree.

where

$$f_S(t_1, t_2, \dots, t_d) = \sum_{\varepsilon \in S} t_1^{<\varepsilon_1>} t_2^{<\varepsilon_2>} \dots t_d^{<\varepsilon_d>} \tag{1.5}$$

is a polynomial linear in each variable  $t_1, t_2, \dots, t_d$  for which the notation  $t_i^{<\varepsilon_i>}$  means

$$t_i^{<\varepsilon_i>} = \begin{cases} t_i & \text{if } \varepsilon_i = 0, \\ (1 - t_i) & \text{if } \varepsilon_i = 1. \end{cases} \tag{1.6}$$

In fact,  $f_S(t_1, t_2, \dots, t_d)$  is the probability that an independent uniformly distributed random point  $(t_1, t_2, \dots, t_d) \in [0, 1]^d$  belongs to one of the hyperoctants encoded by the binary words of  $S$ .

In the next section, we give an algorithm for the computation of  $J_n[S]$  as finite combinatorial sums with inputs  $S$  and  $d$ . In fact, using the Maple package *gfun* (4), it can be seen that for dimensions 2 and 3 the  $J_n[S]$  are polynomially recursive sequences.

We present in Section 3 compact integral formulas for  $J_n[S]$  which can be asymptotically evaluated using Laplace transforms and the classical Watson’s Lemma (3). In Section 4, explicit, recursive and asymptotic expressions are given for the expected arity of the root of randomly grown multidimensional quadtrees following a suggestion by Flajolet et al. in (7).

Our results are related to the study of maxima in hypercubes (see (1), (2)). See also (8), (12), (13) and (15) for other kinds of random trees.

The structure of our recursive trees is given by the underlying independant uniform random points added successively to the tree (6), (9), (11) and (14). And instead of studying degree of vertices, we study what we call the root pattern of the tree which leads us some results about the arity (or number of children) of the root.

## 2 Combinatorial sums for root pattern probabilities

In this section, we first develop an explicit formula and an algorithm to compute the  $J_n[S]$  probabilities.

**Theorem 2.1** Probabilities  $J_n[S]$  can be expressed explicitly with multiple sums :

$$J_n[S] = \frac{1}{n(n!)^{d-1}} \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{\prod_{i=1}^d \left( (\sum_{\varepsilon_i=0} k_\varepsilon)! (\sum_{\varepsilon_i=1} k_\varepsilon)! \right)}{\prod_{\varepsilon \in S} k_\varepsilon!}. \tag{2.1}$$

**Proof.** Using multinomial formula, we get, considering the expression for  $f_S(t_1, t_2, \dots, t_d)$  :

$$J_n[S] = \int_0^1 \int_0^1 \dots \int_0^1 \left( f_S(t_1, t_2, \dots, t_d) \right)^{n-1} dt_1 dt_2 \dots dt_d \tag{2.2}$$

$$= \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{(n-1)!}{\prod_{\varepsilon \in S} k_\varepsilon!} \int_0^1 \int_0^1 \dots \int_0^1 \prod_{\varepsilon \in S} \left( t^{(\varepsilon)} \right)^{k_\varepsilon} dt_1 dt_2 \dots dt_d \tag{2.3}$$

where

$$\prod_{\varepsilon \in S} \left( t^{(\varepsilon)} \right)^{k_\varepsilon} = \prod_{\varepsilon \in S} \left( t_1^{(\varepsilon_1)} t_2^{(\varepsilon_2)} \dots t_d^{(\varepsilon_d)} \right)^{k_\varepsilon} \tag{2.4}$$

$$= \prod_{\varepsilon \in S} \left( t^{(\varepsilon_1)} \right)^{k_\varepsilon} \prod_{\varepsilon \in S} \left( t^{(\varepsilon_2)} \right)^{k_\varepsilon} \dots \prod_{\varepsilon \in S} \left( t^{(\varepsilon_d)} \right)^{k_\varepsilon}. \tag{2.5}$$

Each of these products can be written in the following form

$$\prod_{\varepsilon \in S} \left( t_i^{(\varepsilon_i)} \right)^{k_{\varepsilon_i}} = t_i^{\sum_{\varepsilon_i=0} k_{\varepsilon_i}} (1 - t_i)^{\sum_{\varepsilon_i=1} k_{\varepsilon_i}}. \tag{2.6}$$

We can thus rewrite the  $J_n[S]$  probabilities and evaluate them using Beta functions as follows

$$J_n[S] = \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{(n-1)!}{\prod_{\varepsilon \in S} k_\varepsilon!} \int_0^1 \dots \int_0^1 \prod_{i=1}^d \left( t_i^{\sum_{\varepsilon_i=0} k_\varepsilon} (1 - t_i)^{\sum_{\varepsilon_i=1} k_\varepsilon} \right) dt_1 \dots dt_d \tag{2.7}$$

$$= \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{(n-1)!}{\prod_{\varepsilon \in S} k_\varepsilon!} \prod_{i=1}^d \int_0^1 t_i^{\sum_{\varepsilon_i=0} k_\varepsilon} (1 - t_i)^{\sum_{\varepsilon_i=1} k_\varepsilon} dt_i \tag{2.8}$$

$$= \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{(n-1)!}{\prod_{\varepsilon \in S} k_\varepsilon!} \prod_{i=1}^d \frac{(\sum_{\varepsilon_i=0} k_\varepsilon)! (\sum_{\varepsilon_i=1} k_\varepsilon)!}{n!} \tag{2.9}$$

$$= \frac{(n-1)!}{(n!)^d} \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{\prod_{i=1}^d \left( (\sum_{\varepsilon_i=0} k_\varepsilon)! (\sum_{\varepsilon_i=1} k_\varepsilon)! \right)}{\prod_{\varepsilon \in S} k_\varepsilon!} \tag{2.10}$$

$$= \frac{1}{n(n!)^{d-1}} \sum_{\sum_{\varepsilon \in S} k_\varepsilon = n-1} \frac{\prod_{i=1}^d \left( (\sum_{\varepsilon_i=0} k_\varepsilon)! (\sum_{\varepsilon_i=1} k_\varepsilon)! \right)}{\prod_{\varepsilon \in S} k_\varepsilon!}. \tag{2.11}$$

■

We are now able to describe our algorithm.

Algorithm - Formula for  $J_n[S]$  with binomial coefficients

Input dimension  $d$ , number of points  $n$ , set  $S$  of binary words

Output formula to compute  $J_n[S]$

Step 1 : Build a table with, on the first line,  $i$  followed by each  $k_\varepsilon$  of  $S$ . Let the last two columns be  $\sum_{\varepsilon_i=0} k_\varepsilon$  and  $\sum_{\varepsilon_i=1} k_\varepsilon$ . There will be  $d$  others lines, each of them numbered from 1 to  $d$  (which constitutes the column below  $i$ ). Below the first  $k_\varepsilon$ , write down the sequence of 0 and 1 (including leading 0's if any) that forms the binary representation of  $\varepsilon$ . Do the same for each others  $k_\varepsilon$ .

Step 2 : Fill the last column in the following way : for each line, write down the binomial coefficient  $\binom{n-1}{\alpha}$  where  $\alpha = \sum_{\varepsilon_i=0} k_\varepsilon$  or  $\alpha = \sum_{\varepsilon_i=1} k_\varepsilon$ .

Step 3 : Form the following fractional expression : The numerator will be the multinomial coefficient  $\binom{n-1}{(k_\varepsilon)_{\varepsilon \in S}}$  where  $(k_\varepsilon)_{\varepsilon \in S}$  denotes the sequence of all  $k_\varepsilon$  where  $\varepsilon$  runs through  $S$  and the denominator will be the product of the binomial coefficients in the last column. To obtain  $J_n[S]$ , make the sum of each fractional expression obtained when  $\sum_{\varepsilon \in S} k_\varepsilon = n - 1$  and divide the result by  $n^d$ .

**Example.** Calculation of  $J_n[S]$  for  $d = 6$  and  $S = \{1, 7, 25\} = \{000001, 000111, 011001\}$ , we get

$i$	$k_1$	$k_7$	$k_{25}$	$\sum_{\varepsilon_i=0} k_\varepsilon$	$\sum_{\varepsilon_i=1} k_\varepsilon$	binomial coefficient
1	0	0	0	$k_1 + k_7 + k_{25}$	0	$\binom{n-1}{0}$
2	0	0	1	$k_1 + k_7$	$k_{25}$	$\binom{n-1}{k_{25}}$
3	0	0	1	$k_1 + k_7$	$k_{25}$	$\binom{n-1}{k_{25}}$
4	0	1	0	$k_1 + k_{25}$	$k_7$	$\binom{n-1}{k_7}$
5	0	1	0	$k_1 + k_{25}$	$k_7$	$\binom{n-1}{k_7}$
6	1	1	1	0	$k_1 + k_7 + k_{25}$	$\binom{n-1}{0}$

If we rename  $k_1$  by  $i$ ,  $k_7$  by  $j$  and  $k_{25}$  by  $k$  for more readability, we obtain

$$J_n[S] = \frac{1}{n^6} \sum_{i+j+k=n-1} \frac{\binom{n-1}{i,j,k}}{\binom{n-1}{k} \binom{n-1}{j}} = \frac{1}{n^6} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1-i} \frac{\binom{i+j}{i}}{\binom{n-1}{i+j} \binom{n-1}{j}}. \tag{2.12}$$

Note that for large  $S$  the computation of  $J_n[S]$  can be simplified using the following duality formula which involves  $J_k[\mathbb{C}S]$ , where  $\mathbb{C}S$  denotes the complement of the set  $S$ ,

$$J_n[S] = \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} J_k[\mathbb{C}S]. \tag{2.13}$$

This follows from (1.4) using the fact that  $f_S = 1 - f_{\mathbb{C}S}$ . Section 5 contains complete tables for  $J_n[S]$  in dimension  $d = 2$  and  $d = 3$  computed using our Algorithm and the above duality.

### 3 Compact integrals and asymptotic forms for root pattern probabilities

The multidimensional integral (given by equation (1.4)) for the probability  $J_n[S]$  can be rewritten as a simple definite integral (see (10) for more details)

$$J_n[S] = \int_0^1 (1-x)^{n-1} \omega(x) dx = \int_0^\infty e^{-nt} \varphi(t) dt, \tag{3.1}$$

where

$$\omega(x) = \frac{d\mu(M_S(x))}{dx}, \quad \varphi(t) = \omega(1 - e^{-t}), \tag{3.2}$$

$$M_S(x) = \{(t_1, \dots, t_d) \in [0, 1]^d \mid f_S(t_1, \dots, t_d) > 1 - x\} \subseteq [0, 1]^d. \tag{3.3}$$

In (3.2),  $\mu$  denotes the usual Lebesgue measure. This shows that  $J_n[S]$  can be expressed as a Laplace transform

$$(\mathbb{L}\varphi)(n) \tag{3.4}$$

and from which asymptotic expressions for  $J_n[S]$  can be deduced using classical Watson Lemma (3) :

**Theorem 3.1** Consider sequence  $(a_n)_{n \geq 0}$  defined by integrals of the form

$$a_n = \int_0^T e^{-nt} \varphi(t) dt, \quad 0 < T \leq \infty \tag{3.5}$$

where function  $\varphi(t)$  admits a convergent expansion

$$\varphi(t) = c_0 t^{\beta_0} + c_1 t^{\beta_1} + c_2 t^{\beta_2} + \dots, \quad -1 < \beta_0 < \beta_1 < \beta_2 < \dots. \tag{3.6}$$

valid for all positive  $t$  small enough. Suppose also that there exists an integer  $n_0 \geq 0$  such that

$$\int_0^T e^{-n_0 t} |\varphi(t)| dt < \infty. \tag{3.7}$$

Then we have the following asymptotic expansion

$$a_n \sim c_0 \frac{\Gamma(\beta_0 + 1)}{n^{\beta_0+1}} + c_1 \frac{\Gamma(\beta_1 + 1)}{n^{\beta_1+1}} + c_2 \frac{\Gamma(\beta_2 + 1)}{n^{\beta_2+1}} + \dots, \quad n \rightarrow \infty. \quad (3.8)$$

■

Taking the formal sum of the Laplace transform of individual terms in function  $\varphi(t)$ ,

$$L[c_k t^{\beta_k}](n) = c_k \int_0^\infty e^{-nt} t^{\beta_k} dt = c_k \frac{\Gamma(\beta_k + 1)}{n^{\beta_k+1}}, \quad (3.9)$$

one obtains the terms of asymptotic expansion for  $a_n$ . In particular, this asymptotic expansion is independent of the bound  $T$  and depends only on the behaviour of  $\varphi(t)$  for small values of  $t$ .

More generally, if the function  $\omega(x)$  in (3.1) contains logarithmic terms, for example

$$\omega(x) = \sum_{k \geq 0} a_k x^k + \sum_{k \geq 0} b_k x^k \ln x + \sum_{k \geq 0} c_k x^k (\ln x)^2 + \dots \quad (3.10)$$

then the corresponding function  $\varphi(t)$  will also be of the form

$$\varphi(t) = \sum_{k \geq 0} \alpha_k t^k + \sum_{k \geq 0} \beta_k t^k \ln t + \sum_{k \geq 0} \gamma_k t^k (\ln t)^2 + \dots \quad (3.11)$$

with suitable constants  $\alpha_k, \beta_k, \gamma_k \dots$ . The asymptotic expansion of  $J_n[S]$  can be computed by taking term by term Laplace transforms as above using series expansion of  $\varphi(t)$  in (3.1). For example, in dimension 3 for  $S = \{1, 2, 3, 4, 5, 6, 7\}$  we get

$$\mu(M_{\{1,2,3,4,5,6,7\}}(x)) = x - x \ln(x) + \frac{1}{2} x \ln(x)^2, \quad (3.12)$$

$$\omega(x) = \frac{1}{2} \ln(x)^2, \quad (3.13)$$

and the corresponding asymptotic expansion is

$$J_n[\{1, 2, 3, 4, 5, 6, 7\}] \sim \frac{\pi^2 + 6\gamma^2 + 12\gamma \ln(n) + 6(\ln(n))^2}{12n} + \frac{-1 + \gamma + \ln(n)}{2n^2} + \frac{9 - 2\gamma - 2\ln(n)}{24n^3} - \frac{1}{8n^4} + \frac{5 + 12\gamma + 12\ln(n)}{1440n^5} + \frac{1}{48n^6} + \dots, \quad (3.14)$$

where  $\gamma$  denotes the Euler's constant. Complete tables of the asymptotic expansions of  $J_n[S]$  for dimensions 2 and 3 will appear in an expanded form of the paper.

## 4 Distribution of the arity of the root and its mean value

We have already established in a previous paper the following lemma (10).

**Lemma 4.1** *The probability  $p_{n,k,d}$  that the root of a random quadtree of dimension  $d$  of  $n$  points has exactly  $k$  children is given by*

$$p_{n,k,d} = \sum_{\nu=0}^k (-1)^{k-\nu} \binom{2^d - \nu}{k - \nu} \sum_{|S|=\nu} J_n[S], \quad 0 \leq k \leq 2^d. \quad (4.1)$$

**Theorem 4.1** *For fixed  $n$  and  $d$ , the generating series (polynomial in fact) of the arity of the root of a random  $d$ -dimensional quadtree of  $n$  points is given by*

$$f_{n,d}(x) = \sum_{k=0}^{2^d} p_{n,k,d} x^k = \sum_{\nu=0}^{2^d} \sigma_{n,\nu,d} x^\nu (1-x)^{2^d-\nu} \quad (4.2)$$

where

$$\sigma_{n,\nu,d} = \sum_{|S|=\nu} J_n[S]. \quad (4.3)$$

**Proof.** Use Lemma 4.1. ■

The following corollaries are closely related to similar results concerning maxima in a random set of samples from  $[0, 1]^d$  (see, for example, (1) and (2)).

**Corollary 4.1** *The expected arity  $E_{n,d}$  of the root of a random quadtree of dimension  $d$  of  $n$  points are given by*

$$E_{n,d} = 2^d (1 - J_n[\{1, 2, \dots, 2^d - 1\}]) \tag{4.4}$$

$$= 2^d \left( \binom{n-1}{1} \frac{1}{2^d} - \binom{n-1}{2} \frac{1}{3^d} + \dots + (-1)^k \binom{n-1}{k} \frac{1}{k^d} + (-1)^n \frac{1}{n^d} \right) \tag{4.5}$$

**Proof.** For (4.4) use the fact that

$$E_{n,d} = \sum_{k=0}^{2^d} k p_{n,k,d} = f'_{n,d}(1) \tag{4.6}$$

and compute the derivative using the right hand side of (4.2). Only the terms corresponding to  $\nu = 2^d$  and  $\nu = 2^d - 1$  will survive. Using the fact that

$$J_n[\{1, 2, \dots, 2^d - 1\}] = \int_0^1 \int_0^1 \dots \int_0^1 (1 - t_1 t_2 \dots t_d)^{n-1} dt_1 dt_2 \dots dt_d, \tag{4.7}$$

(4.5) follows immediately by expansion of the integrand. ■

**Corollary 4.2** *The probabilities  $J_n[\{1, 2, \dots, 2^d - 1\}]$  can be written in the form*

$$J_n[\{1, 2, \dots, 2^d - 1\}] = \int_0^1 (1-x)^{n-1} \frac{(-\ln x)^{d-1}}{(d-1)!} dx. \tag{4.8}$$

Moreover,

$$\sum_{d \geq 1} J_n[\{1, 2, \dots, 2^d - 1\}] z^{d-1} = \frac{\frac{1}{n}}{\left(1 - \frac{z}{1}\right) \left(1 - \frac{z}{2}\right) \dots \left(1 - \frac{z}{n}\right)}. \tag{4.9}$$

**Sketch of the proof.** The first equation, (4.8) follows from (4.7) using appropriate change of variables. The second one (4.9) is obtained this way

$$\begin{aligned} \sum_{d \geq 1} J_n[\{1, 2, \dots, 2^d - 1\}] z^{d-1} &= \sum_{d \geq 1} \int_0^1 (1-x)^{n-1} \frac{(-\ln x)^{d-1}}{(d-1)!} z^{d-1} dx, \\ &= \int_0^1 (1-x)^{n-1} x^{-z} dx, \\ &= B(n, 1-z), \\ &= \frac{\Gamma(n)\Gamma(1-z)}{\Gamma(n+1-z)}, \\ &= \frac{\frac{1}{n}}{\left(1 - \frac{z}{1}\right) \left(1 - \frac{z}{2}\right) \dots \left(1 - \frac{z}{n}\right)}. \end{aligned}$$

■

**Corollary 4.3** *The expected value  $E_{n,d}$  can be expressed as the following finite sum involving generalized harmonic numbers.*

$$E_{n,d} = 2^d \left( 1 - \frac{1}{n} \sum_{\lambda \vdash d-1} \frac{H_\lambda(n)}{z_\lambda} \right), \tag{4.10}$$

where,  $\lambda = (\lambda_1, \lambda_2, \dots) \vdash d-1$  denotes a partition of  $d-1$  and then, let  $k_i$  be the number of parts of size  $i$  in  $\lambda$ ,

$$\begin{aligned} z_\lambda &= 1^{k_1} k_1! 2^{k_2} k_2! 3^{k_3} k_3! \dots \\ H_\lambda(n) &= H_{\lambda_1}(n) H_{\lambda_2}(n) H_{\lambda_3}(n) \dots = \prod_{i \geq 1} (H_i(n))^{k_i}, \end{aligned} \tag{4.11}$$

and, for  $k \geq 1$ , the generalized harmonic number  $H_k(n)$  is given by the sum,

$$H_k(n) = \frac{1}{1^k} + \frac{1}{2^k} + \frac{1}{3^k} + \dots + \frac{1}{n^k}. \tag{4.12}$$

**Proof.** From (4.9) we have

$$\begin{aligned} \sum_{d \geq 1} J_{n,d} z^{d-1} &= \frac{1}{n} \exp \left( -\ln \left( 1 - \frac{z}{1} \right) - \ln \left( 1 - \frac{z}{2} \right) - \dots - \ln \left( 1 - \frac{z}{n} \right) \right) \\ &= \frac{1}{n} \exp \left( \sum_{i=1}^n \left( \frac{z}{i} + \frac{1}{2} \left( \frac{z}{i} \right)^2 + \frac{1}{3} \left( \frac{z}{i} \right)^3 + \dots \right) \right) \\ &= \frac{1}{n} \exp \left( H_1(n)z + \frac{1}{2}H_2(n)z^2 + \frac{1}{3}H_3(n)z^3 + \dots \right) \\ &= \frac{1}{n} \sum_{d \geq 1} \left( \sum_{\lambda \vdash d-1} \frac{H_\lambda(n)}{z_\lambda} \right) z^{d-1}. \end{aligned} \tag{4.13}$$

■

In particular, for  $d = 2$ ,

$$\begin{aligned} E_{n,2} &= 4 \left( 1 - \frac{1}{n} \left( \frac{H_{1,1}(n)}{z_{1,1}} + \frac{H_2(n)}{z_2} \right) \right) \\ &= 4 - \frac{4}{n} \left( \frac{\left( 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right)^2}{1!2!} + \frac{\left( 1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{n^2} \right)}{2!1!} \right) \\ &= 4 - \frac{2}{n} \left( H_1(n)^2 + H_2(n) \right). \end{aligned} \tag{4.14}$$

To get asymptotic expressions for the  $E_{n,d}$  we can also use Laplace's transform of (4.8) or the explicit expression from (4.10) using Euler-MacLaurin summation formula to get the asymptotic of  $H_k(n)$ . We obtain

$$\begin{aligned} E_{n,1} &= 2 - 2n^{-1} \\ E_{n,2} &\sim 4 - 4 \frac{\ln(n) + \gamma}{n} - 2n^{-2} + \dots \\ E_{n,3} &\sim 8 - 2/3 \frac{6(\ln(n))^2 + \pi^2 + 12\gamma \ln(n) + 6\gamma^2}{n} - 4 \frac{\ln(n) + \gamma - 1}{n^2} + \dots \\ E_{n,4} &\sim 16 - 4/3 \frac{2(\ln(n))^3 + 2\gamma^3 + 6\gamma^2 \ln(n) + \pi^2 \ln(n) + 6\gamma(\ln(n))^2 + \pi^2\gamma + 4\zeta(3)}{n} \\ &\quad - 2/3 \frac{6(\ln(n))^2 - 12\gamma + 6\gamma^2 + 12\gamma \ln(n) + \pi^2 - 12 \ln(n)}{n^2} + \dots \end{aligned} \tag{4.15}$$

Using *gfun* we compute the following recurrence for  $E_{n,d}$ ,  $d = 2, 3$  :

$$\begin{aligned} E_{1,2} &= 0, & E_{2,2} &= 1, \\ (n-2)n^2 E_{n,2} &= (n-1)(2n^2 - 4n + 1)E_{n-1,2} - (n-1)^2(n-2)E_{n-2,2} \end{aligned} \tag{4.16}$$

$$\begin{aligned} E_{1,3} &= 0, & E_{2,3} &= 1, & E_{3,3} &= \frac{46}{27}, \\ (n-2)n^3 E_{n,3} &= (n-3)(n-2)(n-1)^2 E_{n-3,3} - (n-1)(n-2)(3n^2 - 7n + 3)E_{n-2,3} \\ &\quad + (n-1)(3n^3 - 8n^2 + 5n - 1)E_{n-1,3} \end{aligned} \tag{4.17}$$

Longer tables will appear in an extended form of this paper including the probabilities  $Q_n[R]$  and an analysis of the variance of the expected arity.

We can partition the set of all sets  $S$  with a given cardinality and a given root pattern (up to rotations and mirror transformations) into orbits. Every set  $S$  in a fixed orbit gives the same value of  $J_n[S]$ . This leads us to use an orbit's representative  $S$  and multiply  $J_n[S]$  by the cardinality of this orbit.



## 5 Explicit tables for dimensions 2 and 3

When two sets of hyperoctants  $S$  and  $S'$  are equivalent under rotation and reflection (that is, are in the same orbit under the action of the hyperoctahedral group) we have  $J_n[S] = J_n[S']$ . Formula (1.3) for the root pattern probabilities can be simplified using sums over orbit's representative. If we collect, for a given dimension and a given cardinality, all sets  $S$  which are equivalent under rotations and reflections in a set called an orbit, we can simplify formula (4.1) for the root pattern probability using an orbit's representative multiplied by the cardinality of the orbit. This leads to the following tables.

### 5.1 Bidimensional case

Orbit's representative	Size	$J_n[S]$
{}	1	0
{0}	4	$\frac{1}{n^2}$
{0, 1}	4	$\frac{1}{n}$
{0, 3}	2	$\frac{1}{n^2} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}}$
{0, 1, 2}	4	$\frac{1}{nn!} \sum_{i+j+k=n-1} \frac{(i+j)!(i+k)!}{i!} = \frac{H_n}{n}$
{0, 1, 2, 3}	1	1

### 5.2 Tridimensional case

Orbit's representative	Size	$J_n[S]$
{}	1	0
{0}	8	$\frac{1}{n^3}$
{0, 1}	12	$\frac{1}{n^2}$
{0, 3}	12	$\frac{1}{n^3} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}} = \frac{1}{nn!^2} \sum_{i+j=n-1} i!j!$
{0, 7}	4	$\frac{1}{n^3} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}^2} = \frac{1}{nn!^2} \sum_{i+j=n-1} i!^2 j!^2$
{0, 1, 2}	24	$\frac{1}{nn!^2} \sum_{i+j+k=n-1} \frac{(i+j)!(i+k)!}{i!} = \frac{H_n}{n^2}$
{0, 1, 6}	24	$\frac{1}{nn!^2} \sum_{i+j+k=n-1} \frac{(i+j)!^2 (i+k)!k!}{i!}$
{0, 3, 5}	8	$\frac{1}{n^3} \sum_{i+j+k=n-1} \frac{\binom{n-1}{i,j,k}}{\binom{n-1}{i} \binom{n-1}{j} \binom{n-1}{k}}$
{0, 1, 2, 3}	6	$\frac{1}{n}$
{0, 1, 2, 4}	8	$\frac{1}{n^3} \sum_{i+j+k+l=n-1} \frac{\binom{n-1}{i,j,k,l}}{\binom{n-1}{i} \binom{n-1}{j} \binom{n-1}{k}}$
{0, 1, 2, 5}	24	$\frac{1}{n^3} \sum_{i+j+k+l=n-1} \frac{\binom{n-1}{i,j,k,l}}{\binom{n-1}{i+k} \binom{n-1}{k} \binom{n-1}{l}}$
{0, 1, 2, 7}	24	$\frac{1}{n^3} \sum_{i+j+k+l=n-1} \frac{\binom{n-1}{i,j,k,l}}{\binom{n-1}{i+j} \binom{n-1}{i+k} \binom{n-1}{l}}$
{0, 1, 6, 7}	6	$\frac{1}{n^3} \sum_{i+j+k+l=n-1} \frac{\binom{n-1}{i,j,k,l}}{\binom{n-1}{i+j}^2 \binom{n-1}{i+k}}$
{0, 3, 5, 6}	2	$\frac{1}{n^3} \sum_{i+j+k+l=n-1} \frac{\binom{n-1}{i,j,k,l}}{\binom{n-1}{i+j} \binom{n-1}{i+k} \binom{n-1}{i+l}}$

The remaining  $J_n[S]$  can be computed using (2.13).

## References

- [1] Z. D. Bai, L. Devroye, H. K. Hwang, T. H. Tsai. "Maxima in hypercubes". *Random Structures and Algorithms*, vol. 27, no. 3, p. 290-309, October 2005.
- [2] J. L. Bentley, H. T. Kung, M. Schkolnick, C. D. Thompson. "On the Average Number of Maxima in a Set of Vectors and Applications". *Journal of the ACM*, vol. 25, no. 4, p. 536-543. October 1978.
- [3] N. Bleistein and R. A. Handelsman. *Asymptotic Expansions of Integrals*. New-York: Holt, Rinehart and Winston. 1975.
- [4] B. W. Char, K. O. Geddes, G. H. Gonnet, B. L. Leong, M. B. Monagan, S. M. Watt. *Maple V Library Reference Manual*. Berlin: Springer-Verlag. 1991.
- [5] R. A. Finkel and J. L. Bentley. "Quad Trees: A Data Structure for Retrieval on Composite Keys". *Acta Informatica*, vol. 4, p. 1-9. 1974.
- [6] P. Flajolet, G. Gonnet, C. Puech, and J. M. Robson. Analytic variations on quadtrees. *Algorithmica* 10, p. 473-500, 1993.
- [7] P. Flajolet, G. Labelle, L. Lafortest, B. Salvy. "Hypergeometrics and the cost structure of quadtrees". *Random Structures and Algorithms*, vol. 7, no. 2, p. 117-144. 1995.
- [8] S. Janson, Asymptotic Degree Distribution in Random Recursive Trees. *Random Structures and Algorithms*. vol. 26, no. 1-2, January - March 2005, p. 69-83
- [9] G. Labelle and L. Lafortest, "Combinatorial Variations on Multidimensional Quadtrees". *Journal of Combinatorial Theory*, vol. 69, no. 1, p. 1-16. 1995.
- [10] G. Labelle and L. Lafortest, "Sur la distribution de l'arité de la racine d'une arborescence hyperquaternaire". *Discrete Mathematics* 139, no. 1-3, p. 287-302. 1995.
- [11] L. Lafortest, "Étude des arbres hyperquaternaires". Thèse de doctorat, Montréal: Université McGill. 1990.
- [12] H. M. Mahmoud and R. T. Smythe, "Asymptotic joint normality of outdegrees of nodes in random recursive trees". *Random Structures and Algorithms*, 3, p. 255-266. 1992.
- [13] H. M. Mahmoud, R. Smythe and J. Szymanski, "On the structure of plane-oriented recursive trees and their branches". *Random Structures and Algorithms*, 4, p. 151-176. 1993.
- [14] X. Provençal. "Une étude de l'arité de la racine d'un arbre hyperquaternaire". Mémoire de maîtrise, Montréal: Université du Québec à Montréal. 2004.
- [15] R. Smythe and H. M. Mahmoud, "A survey of recursive trees", *Theory Probab. Math. Statist.* 51. p. 1-27. 1996.

