

# SUR LA DISTRIBUTION DE L'ARITÉ DE LA RACINE D'UNE ARBORESCENCE HYPERQUATERNaire À $d$ DIMENSIONS

Gilbert Labelle et Louise Laforest, LACIM-UQAM

## 0. Introduction

La notion d'arborescence hyperquaternaire fait appel aux  $d$ -tuplets de permutations en considérant la  $L$ -espèce produit

$$T = L \times L \times \dots \times L \quad (d \text{ facteurs})$$

où  $L$  est l'espèce des ordres linéaires. Posons  $[n] = \{1, 2, \dots, n\}$ . Une arborescence hyperquaternaire (i.e. une  $T$ -structure) sur  $[n]$  consiste donc en une suite de  $d$  permutations quelconques des entiers de 1 à  $n$ . En langage géométrique [4, 5, 7, 11, 15, 19], une  $T$ -structure est une structure arborescente qui code l'« histoire » des apparitions successives de  $n$  points aléatoires  $P_1, P_2, \dots, P_d$  dans l'hypercube  $[0,1]^d$ . Pour  $i = 1, 2, \dots, d$ , chaque point  $P_i$  est contenu dans un sous-pavé de  $[0,1]^d$  qu'il subdivise en  $2^d$  sous-pavés disjoints (généralisant ainsi le contexte des arborescences binaires de recherche,  $d = 2$ .) On suppose que les points  $P_i$  sont indépendants et uniformément distribués; ce qui équivaut à dire que les  $T$ -structures sont équidistribuées avec probabilité  $1/n!^d$ .

Dans la figure 1, nous avons représenté schématiquement  $d$  permutations dans lesquelles le nombre 1 est mis en évidence pour faire ressortir le fait qu'il correspond au premier point aléatoire.

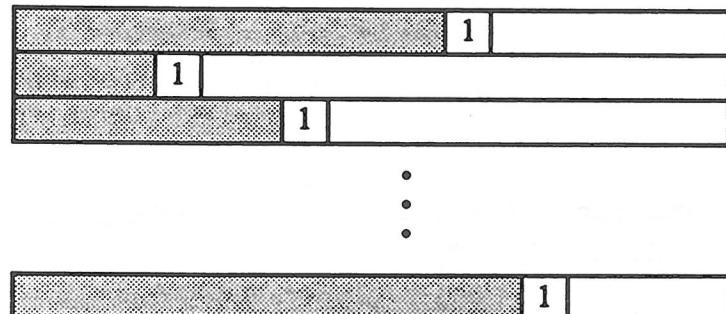


figure 1

Par exemple, dans le cas à deux dimensions, supposons que nous ayions les deux permutations

$$\sigma_1 = (3, 2, 4, 5, 1, 6) \quad \text{et} \quad \sigma_2 = (4, 5, 3, 1, 6, 2)$$

représentant respectivement les projections sur l'axe des  $x$  et l'axe des  $y$  des numéros des points constituant le nuage de points. La numérotation correspond à l'ordre d'apparition des points en question. On peut visualiser ce nuage de points dans la figure 2 :

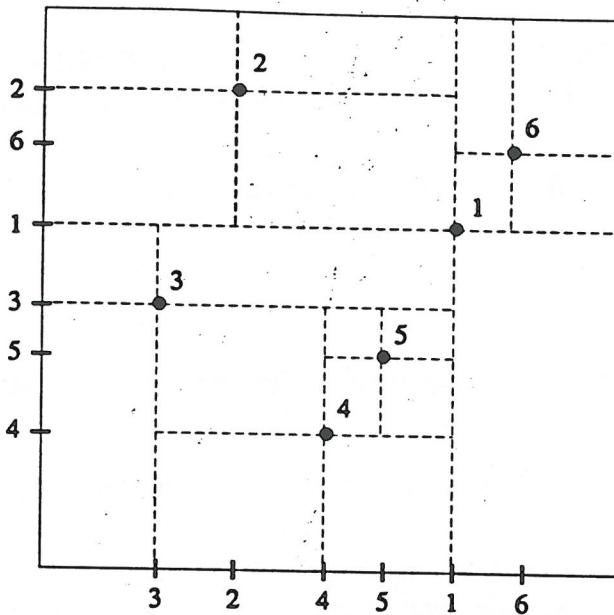


figure 2

Soit  $\pi_{n,k,d}$  la probabilité qu'une sous-arborescence donnée de la racine d'une arborescence hyperquaternaire aléatoire de  $n$  noeuds, à  $d$  dimensions ait  $k$  enfants. Plusieurs expressions, récursives et explicites, pour ces probabilités ont été obtenues dans [13,15]. La distribution (ou l'espérance) du nombre de noeuds ayant une arité donnée (i.e. nombre d'enfants donné) est fonction de cette probabilité  $\pi_{n,k,d}$  ainsi que de la probabilité  $p_{n,k,d}$  que la racine d'une arborescence hyperquaternaire de  $d$  dimensions, ayant  $n$  noeuds soit d'arité  $k$ . Pour  $k$  et  $d$  fixés, nous avons établi dans [13] que l'espérance  $e_{n,k,d}$  du nombre de noeuds ayant  $k$  enfants dans une arborescence hyperquaternaire aléatoire de  $n$  points satisfait le schéma de récurrence suivant:

$$e_{n,k,d} = p_{n,k,d} + 2^d \sum_{i=0}^{n-1} \pi_{n,i,d} e_{i,k,d} \quad (1)$$

où

$$p_{n,k,d} = \text{probabilité que la racine d'une arborescence hyperquaternaire de } n \text{ points ait } k \text{ enfants } (0 \leq k \leq 2^d). \quad (2)$$

Le but du présent travail est de faire une étude combinatoire et analytique de ces probabilités fondamentales.

### 1. Probabilité que les $n$ points tombent dans une région donnée.

Nous allons d'abord regarder ce qui se passe dans le cas bidimensionnel. Il nous faut examiner les cas où  $k = 0, 1, 2, 3$  et  $4$ . Pour commencer, nous allons considérer la probabilité que les  $n$  points tombent dans un ensemble fixé de quadrants. On peut visualiser ces possibilités dans la figure 3. Les régions considérées sont les ensembles de carrés gris. Les quadrants sont numérotés de la façon suivante :

1	3	01	11
0	2	00	10

(3)

Les variables  $t_1$  et  $t_2$  correspondent respectivement aux coordonnées  $x$  et  $y$  du premier point aléatoire. Pour les besoins de l'illustration, nous avons choisi le point  $(\frac{1}{2}, \frac{1}{2})$  sans oublier que  $(t_1, t_2)$  est aléatoire et que  $t_1$  et  $t_2$  sont indépendants et distribués selon une loi  $U[0,1]$ . Pour un choix de quadrants donnés, l'intégrale correspondante donne la probabilité que les  $n-1$  autres points tombent dans les régions grises.

Pour les besoins de la présentation, nous allons introduire la notation suivante :

$$t^{<0>} = t \quad \text{et} \quad t^{<1>} = 1-t \quad (4)$$

où  $t$  est une variable réelle. Avec cette convention, on obtient une écriture plus uniforme des intégrales impliquées. Dans la figure 4 nous retrouvons les intégrales de la figure 3 exprimées avec cette nouvelle notation. La probabilité est maintenant notée  $J_n[S]$  où  $S$  est l'ensemble de mots binaires qui code les quadrants choisis.

On remarquera que les indices supérieurs dans les intégrants correspondent exactement aux mots binaires apparaissant dans l'ensemble  $S$ .

Dans le cas à trois dimensions, nous illustrons dans la figure 5 le cas de quatre octants gris. Nous avons dessiné deux des 70 façons possibles de choisir quatre octants parmi huit. Pour la configuration (a), voici l'intégrale qui correspond à la probabilité que les  $n-1$  autres points tombent dans les quatre octants choisis :

$$\int_0^1 \int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} t_3^{<0>} + t_1^{<0>} t_2^{<0>} t_3^{<1>} + t_1^{<0>} t_2^{<1>} t_3^{<0>} + t_1^{<1>} t_2^{<1>} t_3^{<1>})^{n-1} dt_1 dt_2 dt_3.$$

Dans le cas général à  $d$  dimensions, nous introduisons la notation suivante :

$$J_n[S] = \int_0^1 \dots \int_0^1 \left( \sum_{\varepsilon \in S} t_1^{<\varepsilon_1>} \dots t_d^{<\varepsilon_d>} \right)^{n-1} dt_1 \dots dt_d \quad (5)$$

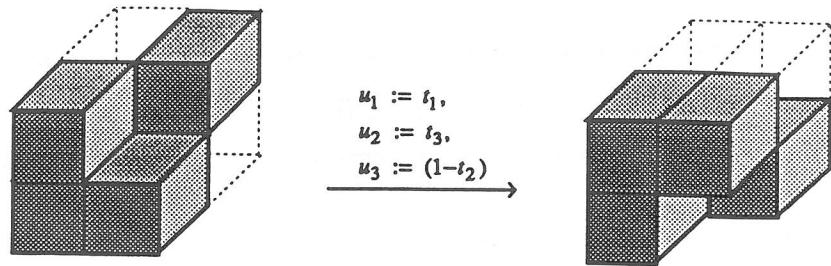
où  $S$  est un ensemble quelconque de mots binaires de longueur  $d$  codant la configuration des hyperoctants choisis.

Région	Probabilité	Codage
	$\int_0^1 \int_0^1 0^{n-1} dt_1 dt_2$	{ }
	$\int_0^1 \int_0^1 (t_1 t_2)^{n-1} dt_1 dt_2$	{00}
	$\int_0^1 \int_0^1 (t_1 (1-t_2))^{n-1} dt_1 dt_2$	{01}
	$\int_0^1 \int_0^1 ((1-t_1) t_2)^{n-1} dt_1 dt_2$	{10}
	$\int_0^1 \int_0^1 ((1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{11}
	$\int_0^1 \int_0^1 (t_1 t_2 + t_1 (1-t_2))^{n-1} dt_1 dt_2$	{00,01}
	$\int_0^1 \int_0^1 (t_1 t_2 + (1-t_1) t_2)^{n-1} dt_1 dt_2$	{00,10}
	$\int_0^1 \int_0^1 (t_1 (1-t_2) + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{01,11}
	$\int_0^1 \int_0^1 ((1-t_1) t_2 + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{10,11}
	$\int_0^1 \int_0^1 (t_1 t_2 + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{00,11}
	$\int_0^1 \int_0^1 (t_1 (1-t_2) + (1-t_1) t_2)^{n-1} dt_1 dt_2$	{01,10}
	$\int_0^1 \int_0^1 (t_1 t_2 + t_1 (1-t_2) + (1-t_1) t_2)^{n-1} dt_1 dt_2$	{00,01,10}
	$\int_0^1 \int_0^1 (t_1 t_2 + t_1 (1-t_2) + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{00,01,11}
	$\int_0^1 \int_0^1 (t_1 t_2 + (1-t_1) t_2 + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{00,10,11}
	$\int_0^1 \int_0^1 (t_1 (1-t_2) + (1-t_1) t_2 + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2$	{01,10,11}
	$\int_0^1 \int_0^1 1^{n-1} dt_1 dt_2$	{00,01,10,11}

figure 3

Région	Probabilité $J_n[S]$	$S$
	$\int_0^1 \int_0^1 0^{n-1} dt_1 dt_2$	{}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>})^{n-1} dt_1 dt_2$	{00}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<1>})^{n-1} dt_1 dt_2$	{01}
	$\int_0^1 \int_0^1 (t_1^{<1>} t_2^{<0>})^{n-1} dt_1 dt_2$	{10}
	$\int_0^1 \int_0^1 (t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{11}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{00,01}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<1>} t_2^{<0>})^{n-1} dt_1 dt_2$	{00,10}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{01,11}
	$\int_0^1 \int_0^1 (t_1^{<1>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{10,11}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{00,11}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<0>})^{n-1} dt_1 dt_2$	{01,10}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{00,01,10}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{00,01,11}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<1>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{00,10,11}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{01,10,11}
	$\int_0^1 \int_0^1 (t_1^{<0>} t_2^{<0>} + t_1^{<0>} t_2^{<1>} + t_1^{<1>} t_2^{<0>} + t_1^{<1>} t_2^{<1>})^{n-1} dt_1 dt_2$	{00,01,10,11}

figure 4



$$J_n[S] = J_n[\{000, 001, 010, 111\}]$$

(a)

$$J_n[S'] = J_n[\{000, 001, 011, 110\}]$$

(b)

figure 5

**Proposition 1:** La probabilité  $p_{n,k,d}$  qu'une arborescence hyperquaternaire aléatoire de  $n$  points possède exactement  $k$  enfants est donné par la formule

$$p_{n,k,d} = \sum_{v=0}^k (-1)^{k-v} \binom{2^d - v}{k-v} \sum_{|S|=v} J_n[S] \quad (6)$$

où les  $J_n[S]$  sont définis par l'équation (5).

**Preuve :** Pour un ensemble  $S$  d'hyperoctants donnés, définissons  $J_n^*[S]$  comme étant la probabilité que les hyperoctants déterminés par  $S$  contiennent au moins un point chacun. On a évidemment que

$$J_n[T] = \sum_{S \subseteq T} J_n^*[S]. \quad (7)$$

Par inversion de Möbius, ou le principe d'inclusion-exclusion, on en déduit que

$$J_n^*[S] = \sum_{T \subseteq S} (-1)^{|S|-|T|} J_n[T]. \quad (8)$$

Nous obtenons donc la suite d'expressions suivantes pour la probabilité cherchée :

$$\begin{aligned} p_{n,k,d} &= \sum_{|S|=k} J_n^*[S] = \sum_{|S|=k} \sum_{T \subseteq S} (-1)^{|S|-|T|} J_n[T] \\ &= \sum_{v=0}^k \sum_{\substack{T, S \\ |T|=v, T \subseteq S, |S|=k}} (-1)^{k-v} J_n[T] = \sum_{v=0}^k \sum_{|T|=v} \sum_{\substack{T \subseteq S \\ |S|=k}} (-1)^{k-v} J_n[T] \\ &= \sum_{v=0}^k \sum_{|T|=v} (-1)^{k-v} J_n[T] \sum_{\substack{|S|=k \\ T \subseteq S}} 1 = \sum_{v=0}^k (-1)^{k-v} \binom{2^d - v}{k-v} \sum_{|T|=v} J_n[T]. \end{aligned}$$

■

## 2. Action des groupes hyperoctaédraux sur les intégrales $J_n[S]$ .

On peut remarquer que plusieurs intégrales  $J_n[S]$  de la proposition précédente prennent la même valeur en effectuant un changement de variables simple comme dans l'exemple de la figure 5.

Dans cette figure, nous pouvons constater que le deuxième volume représenté par  $S' = \{000,001,011,110\}$  peut être obtenu du premier, représenté par  $S = \{000,001,010,111\}$ , en effectuant une rotation de  $90^\circ$  vers la droite. L'intégrale  $J_n[S']$  peut s'obtenir de  $J_n[S]$  par le changement de variable donné par

$$u_1 := t_1^{<0>} , u_2 := t_3^{<0>} , u_3 := t_2^{<1>} \quad (9)$$

ou

$$u_1 := t_1 , u_2 := t_3 , u_3 := (1-t_2).$$

Les intégrales  $J_n[S]$  et  $J_n[S']$  sont donc égales. On remarquera que les variables sont permutées et/ou « inversées ».

Dans le cas général, les changements de variables sont codés par des couples  $(\alpha, \sigma)$  tels que

$$\alpha = \alpha_1 \alpha_2 \dots \alpha_d \in 2^d , \quad \sigma = \sigma(1) \sigma(2) \dots \sigma(d) \in \mathfrak{S}_d \quad (10)$$

où  $2^d$  est l'ensemble des mots binaires de longueur  $d$  et où  $\mathfrak{S}_d$  est le groupe symétrique de degré  $d$ .

Le changement de variables associé au couple  $(\alpha, \sigma)$  s'effectue de la façon suivante : Pour  $i = 1, 2, \dots, d$ ,

$$\boxed{\begin{aligned} J_n[S] &= \int_0^1 \dots \int_0^1 \left( \sum_{\varepsilon \in S} t_1^{<\varepsilon_1>} \dots t_d^{<\varepsilon_d>} \right)^{n-1} dt_1 \dots dt_d \\ u_i &:= t_{\sigma^{-1}(i)}^{<\alpha_i>} \\ J_n[S'] &= \int_0^1 \dots \int_0^1 \left( \sum_{\varepsilon \in S} t_1^{<\varepsilon_1>} \dots t_d^{<\varepsilon_d>} \right)^{n-1} dt_1 \dots dt_d. \end{aligned}} \quad (11)$$

Tous les changements de variables codés par tous les  $2^d d!$  couples possibles  $(\alpha, \sigma)$  seront appelés « changements de variables admissibles ».

**Proposition 2 :** L'ensemble de tous les changements de variables admissibles forme un groupe isomorphe au groupe hyperoctaédral  $B_d$ . L'action des changements de variables sur les intégrales  $J_n[S]$  est isomorphe à l'action du groupe hyperoctaédral sur les ensembles  $S$  de mots binaires de longueur  $d$  et est décrite comme suit :

$$B_d \times \wp(2^d) \rightarrow \wp(2^d) \quad (12)$$

où

$$(\alpha, \sigma)S = S' = \alpha + S \circ \sigma^{-1} = \{ w' \mid w = w_1 w_2 \dots w_d \in S, w'_i = \alpha_i + w_{\sigma^{-1}(i)}, i = 1, \dots, d \},$$

la somme désignant la somme binaire (ou exclusif) et  $\circ$  représentant la composition.

**Preuve :** On vérifie que le composé de deux changements de variables  $(\alpha, \sigma)$  et  $(\bar{\alpha}, \bar{\sigma})$ , noté  $(\bar{\alpha}, \bar{\sigma})(\alpha, \sigma)$ , est le changement admissible  $(\bar{\alpha}, \bar{\sigma})$  donné par  $\bar{\alpha} = \bar{\alpha} + \alpha \circ \bar{\sigma}^{-1}$  et  $\bar{\sigma} = \bar{\sigma} \circ \sigma$ . Ces opérations sont justement celles du groupe hyperoctaédral  $B_d = 2^d \times \mathfrak{S}_d$ . Il est facile de voir que  $S'$  donné par  $\alpha + w \circ \sigma^{-1}$  est égal à l'ensemble codant la deuxième intégrale de (11). ■

On remarquera que le changement de variables donné dans la figure 5 correspond au couple  $(\alpha, \sigma) = (001, 132)$  et que

$$(\alpha, \sigma)S = (\alpha, \sigma)\{000, 001, 010, 111\} = \{000, 001, 011, 110\} = S'.$$

Si deux ensembles  $S$  et  $S'$  appartiennent à la même orbite de l'action (12) alors les intégrales correspondantes  $J_n[S]$  et  $J_n[S']$  sont égales. On est naturellement amenés à rechercher un système de représentants de ces orbites ainsi que la cardinalité de chacune d'elle afin de regrouper les intégrales de valeur égale dans (6).

Revenons à notre exemple à trois dimensions et quatre octants gris. La table 1 nous donne les 70 possibilités de choisir quatre octants parmi huit, classées en orbites. Pour chacune, on donne la taille, un représentant ainsi que les octants choisis. On a directement, en utilisant les informations de la table 1 que

$$\begin{aligned} \sum_{|S|=4} J_n[S] &= \sum_{|S|=4} \int_0^1 \int_0^1 \int_0^1 \left( \sum_{\varepsilon \in S} t_1^{<\varepsilon_1>} t_2^{<\varepsilon_2>} t_3^{<\varepsilon_3>} \right)^{n-1} dt_1 dt_2 dt_3 \\ &= 6J_n[\{000, 001, 010, 011\}] + 24J_n[\{000, 001, 010, 111\}] + 8J_n[\{000, 001, 010, 100\}] + \\ &\quad 24J_n[\{000, 001, 010, 101\}] + 6J_n[\{000, 001, 110, 111\}] + 2J_n[\{000, 011, 101, 110\}]. \end{aligned}$$

Dans les tables 2 et 3, générées par un programme maple [3], nous retrouvons la taille et un représentant pour chaque orbite dans le cas des dimensions  $d = 2$  et  $3$ . Ici, nous avons remplacé le mot binaire par son écriture en base 10. Le représentant choisi est l'ensemble minimal : c'est l'ensemble donnant le nombre minimal si on considère ses éléments, un fois triés en ordre croissant, comme un nombre écrit à la base  $2^d$ . Les cardinalités des orbites, pour chaque  $d$  fixé, sont en concordance avec les tables données dans [8] dans le contexte de la classification des fonctions booléennes.

Cas où  $d = 3$  et  $k = 4$

```

Nombre d'orbites :      6
orbite   1  taille    6  représentant  {000, 001, 010, 011}
orbite   2  taille   24  représentant  {000, 001, 010, 111}
orbite   3  taille    8  représentant  {000, 001, 010, 100}
orbite   4  taille   24  représentant  {000, 001, 010, 101}
orbite   5  taille    6  représentant  {000, 001, 110, 111}
orbite   6  taille    2  représentant  {000, 011, 101, 110}

l'ensemble des orbites :

{ {000, 001, 010, 011}, {000, 001, 100, 101}, {000, 010, 100, 110},
  {001, 011, 101, 111}, {010, 011, 110, 111}, {100, 101, 110, 111} }

{ {000, 001, 010, 111}, {000, 001, 011, 110}, {000, 001, 100, 111},
  {000, 001, 101, 110}, {000, 010, 011, 101}, {000, 010, 100, 111},
  {000, 010, 101, 110}, {000, 011, 100, 101}, {000, 011, 100, 110},
  {000, 011, 101, 111}, {000, 011, 110, 111}, {000, 101, 110, 111},
  {001, 010, 011, 100}, {001, 010, 100, 101}, {001, 010, 100, 110},
  {001, 010, 101, 111}, {001, 010, 110, 111}, {001, 011, 100, 111},
  {001, 011, 101, 110}, {001, 011, 110, 111}, {010, 011, 100, 111},
  {010, 011, 101, 110}, {010, 010, 101, 111}, {011, 100, 101, 110} }

{ {000, 001, 010, 100}, {000, 001, 011, 101}, {000, 010, 011, 110},
  {000, 100, 101, 110}, {001, 010, 011, 111}, {001, 100, 101, 111},
  {010, 100, 110, 111}, {011, 101, 110, 111} }

{ {000, 001, 010, 101}, {000, 001, 010, 110}, {000, 001, 011, 100},
  {000, 001, 011, 111}, {000, 001, 101, 110}, {000, 001, 101, 111},
  {000, 010, 011, 100}, {000, 010, 011, 111}, {000, 010, 100, 101},
  {000, 010, 110, 111}, {000, 100, 101, 111}, {001, 011, 100, 101},
  {001, 010, 011, 101}, {001, 010, 011, 110}, {001, 011, 011, 100},
  {001, 011, 110, 111}, {001, 100, 101, 110}, {001, 101, 110, 111},
  {010, 011, 100, 110}, {010, 011, 101, 111}, {010, 100, 101, 110},
  {010, 101, 110, 111}, {011, 100, 101, 111}, {011, 100, 110, 111} }

{ {000, 001, 110, 111}, {000, 010, 101, 111}, {000, 011, 100, 111},
  {001, 010, 101, 110}, {001, 011, 100, 110}, {010, 011, 100, 101} }

{ {000, 011, 101, 110}, {001, 010, 100, 111} }

```

table 1

$d = 2$		
Nombre d'hyperoctants	Taille de l'orbite	Représentant de l'orbite
$k = 0$	1	{}
$k = 1$	4	{0}
$k = 2$	2	{0, 3}
	4	{0, 1}
$k = 3$	4	{0, 1, 2}
$k = 4$	1	{0, 1, 2, 3}

table 2

$d = 3$		Taille de l'orbite	Représentant de l'orbite
Nombre d'hyperoctants			
$k = 0$		1	{}
$k = 1$		8	{0}
$k = 2$		12	{0, 1}
		12	{0, 3}
		4	{0, 7}
$k = 3$		24	{0, 1, 2}
		24	{0, 1, 6}
		8	{0, 3, 5}
$k = 4$		6	{0, 1, 2, 3}
		8	{0, 1, 2, 4}
		24	{0, 1, 2, 5}
		24	{0, 1, 2, 7}
		6	{0, 1, 6, 7}
		2	{0, 3, 5, 6}
$k = 5$		24	{0, 1, 2, 3, 4}
		8	{0, 1, 2, 4, 7}
		24	{0, 1, 2, 5, 6}
$k = 6$		12	{0, 1, 2, 3, 4, 5}
		12	{0, 1, 2, 3, 4, 7}
		4	{0, 1, 2, 5, 6, 7}
$k = 7$		8	{0, 1, 2, 3, 4, 5, 6}
$k = 8$		1	{0, 1, 2, 3, 4, 5, 6, 7}

table 3

### 3. Analyse asymptotique des intégrales $J_n[S]$ .

Les valeurs exactes des intégrales  $J_n[S]$  ne semblent pas pouvoir s'exprimer sous des formes closes ou être calculables par des schémas récursifs simples. Nous présentons ici l'ébauche d'une approche permettant de les estimer asymptotiquement lorsque  $n \rightarrow \infty$ . Pour un ensemble  $S$  de mots binaires de longueur  $d$ , posons

$$f_S(t_1, t_2, \dots, t_d) = \sum_{\varepsilon \in S} t_1^{<\varepsilon_1>} t_2^{<\varepsilon_2>} \dots t_d^{<\varepsilon_d>}. \quad (13)$$

Puisque  $f_S(t_1, t_2, \dots, t_d)$  est la probabilité qu'un point aléatoire tombe dans l'ensemble d'hyperoctants déterminés par  $S$  et par le point  $(t_1, t_2, \dots, t_d)$ , on a évidemment

$$(t_1, t_2, \dots, t_d) \in [0,1]^d \Rightarrow 0 \leq f_S(t_1, t_2, \dots, t_d) \leq 1. \quad (14)$$

Pour estimer l'intégrale

$$J_n[S] = \int_0^1 \dots \int_0^1 (f_S(t_1, t_2, \dots, t_d))^{n-1} dt_1 \dots dt_d, \quad (15)$$

il faut donc d'abord déterminer l'ensemble des points de l'hypercube unité maximisant  $f_S$ .

Posons à cet effet

$$M_S = \{(t_1, t_2, \dots, t_d) \in [0,1]^d \mid f_S(t_1, t_2, \dots, t_d) = 1\}, \quad (16)$$

et pour tout mot binaire  $\alpha = \alpha_1 \alpha_2 \dots \alpha_d \in 2^d$ , dénotons par

$$\alpha^* = (1 - \alpha_1, 1 - \alpha_2, \dots, 1 - \alpha_d) \in [0,1]^d \quad (17)$$

le sommet de l'hypercube correspondant au complément binaire du mot  $\alpha$ . La proposition suivante décrit complètement la structure de l'ensemble  $M_S$ .

**Proposition 3 :** L'ensemble  $M_S = \{(t_1, t_2, \dots, t_d) \in [0,1]^d \mid f_S(t_1, t_2, \dots, t_d) = 1\}$  est le plus petit ensemble  $M \subseteq [0,1]^d$  satisfaisant les deux conditions suivantes :

$$a) \quad \alpha \in S \Rightarrow \alpha^* \in M, \quad (18)$$

$$b) \quad 0 \leq t \leq 1, \quad (t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_d) \in M, \quad (t_1, \dots, t_{i-1}, 1, t_{i+1}, \dots, t_d) \in M$$

$$\Rightarrow (t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_d) \in M. \quad (19)$$

**Preuve :** Pour tout  $T \subseteq 2^d$ , désignons par  $\chi_T(x_1, x_2, \dots, x_d)$  la fonction caractéristique de l'ensemble  $T$  (ainsi, pour tout mot binaire  $\alpha = \alpha_1 \alpha_2 \dots \alpha_d$  on a  $\alpha \in T$  si et seulement si  $\chi_T(\alpha_1, \alpha_2, \dots, \alpha_d) = 1$ ). Il est facile de vérifier que

$$f_S(\alpha_1, \alpha_2, \dots, \alpha_d) = \chi_S(1 - \alpha_1, 1 - \alpha_2, \dots, 1 - \alpha_d)$$

pour tout mot binaire  $\alpha_1 \alpha_2 \dots \alpha_d$ . Ainsi, les sommets de l'hypercube  $[0,1]^d$  qui appartiennent à  $M_S$  coincident avec les  $\alpha^*$  où  $\alpha$  parcourt  $S$ . Remarquons que  $f_S(t_1, t_2, \dots, t_d)$  est de degré  $\leq 1$  en chaque variable  $t_i$ . En fait, on peut écrire

$$f_S(t_1, \dots, t_i, \dots, t_d) = t_i f_S(t_1, \dots, 1, \dots, t_d) + (1 - t_i) f_S(t_1, \dots, 0, \dots, t_d). \quad (20)$$

Supposons maintenant que

$$f_S(t_1, \dots, 0, \dots, t_d) = f_S(t_1, \dots, 1, \dots, t_d) = 1, \quad (21)$$

alors par (20) on en déduit que  $f_S(t_1, \dots, t_i, \dots, t_d) = 1$ . Ainsi, la propriété b) est nécessairement satisfaite par  $M_S$ . Pour établir la minimalité de  $M_S$ , on raisonne comme suit : Prenons un point  $p = (t_1, t_2, \dots, t_d) \in M_S$  et considérons l'ensemble  $K_p = \{i \mid 1 \leq i \leq d, 0 < t_i < 1\}$ . En faisant appel à (20) pour chaque  $i \in K_p$ , on détermine facilement l'ensemble  $S_p$  des sommets de l'hypercube  $[0,1]^d$  dont la fermeture convexe contient  $p$ . On a évidemment  $S_p \subseteq M_S$ . Ainsi, le point  $p$  provient nécessairement d'applications de a) et /ou de b). ■

L'analyse asymptotique des  $J_n[S]$  peut être faite en approximant les ensembles  $M_S$  et en appliquant la méthode classique de Laplace [10] aux représentations intégrales suivantes :

**Proposition 4 :** Pour tout ensemble  $S$  de mots binaires de longueur  $d$ , posons

$$M_S(x) = \{(t_1, \dots, t_d) \in [0,1]^d \mid f_S(t_1, \dots, t_d) \geq 1 - x\}, \quad (22)$$

alors

$$J_n[S] = \int_0^1 \dots \int_0^1 (f_S(t_1, \dots, t_d))^{n-1} dt_1 \dots dt_d \quad (23)$$

$$= \int_0^1 (1-x)^{n-1} \omega(x) dx \quad (24)$$

$$= \int_0^1 e^{-nt} \varphi(t) dt = (\mathcal{L}\varphi)(s)|_{s=n} \text{ (transformée de Laplace),} \quad (25)$$

où

$$\omega(x) = \frac{d\mu(M_S(x))}{dx}, \quad \varphi(t) = \omega(1-e^{-t}). \quad (26)$$

**Preuve :** Pour un entier  $N > 0$ , prenons  $x_i = i/N$ ,  $i = 0, 1, \dots, N-1$ . On a évidemment,

$$\begin{aligned} J_n[S] &= \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} (1-x_i)^{n-1} \mu(M_S(x_{i+1}) \setminus M_S(x_i)) \\ &= \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} (1-x_i)^{n-1} \omega(x_i) \Delta x_i. \end{aligned}$$

D'où l'on tire la représentation intégrale (24). La transformée de Laplace (25) découle du changement de variable  $x = 1 - e^{-t}$ . ■

Le comportement asymptotique de  $\omega(x)$  pour  $x \rightarrow 0$  ou de  $\varphi(t)$  pour  $t \rightarrow 0$  détermine donc complètement le comportement asymptotique des  $J_n[S]$  pour  $n \rightarrow \infty$ .

**Exemple :** En dimension  $d = 2$ , considérons l'ensemble  $S = \{01, 10, 11\}$ . Alors  $M_S = \{(0, *), (*, 0)\}$  où l'étoile (\*) signifie de prendre toutes les valeurs possibles pour cette composante.



figure 6

L'ensemble  $M_S(x)$  est délimité par un arc d'hyperbole (voir figure 6). On vérifie facilement qu'on a alors  $\omega(x) = -\log x$  et

$$J_n[S] = \int_0^1 (1-x)^{n-1} (-\log x) dx \sim \frac{\log n}{n} + \frac{\gamma}{n} + \frac{1}{2n^2} - \frac{1}{12n^3} + \frac{1}{120n^5} - \dots,$$

où  $\gamma$  désigne la constante d'Euler. En procédant de façon analogue pour les autres intégrales  $J_n[S]$  et en utilisant les représentants de la section 2 on trouve les valeurs asymptotiques suivantes pour les probabilités  $p_{n,k,d}$  en dimension  $d = 2$ , lorsque  $n$  tend vers l'infini :

$$p_{n,0,2} = 0^{n-1} \sim 0$$

$$p_{n,1,2} = \frac{4}{n^2}$$

$$p_{n,2,2} \sim \frac{4}{n} - \frac{8}{n^2} + \frac{4}{n^3} + \frac{12}{n^4} + \frac{52}{n^5} + \frac{300}{n^6} + \dots$$

$$p_{n,3,2} \sim \frac{4 \log n}{n} - \frac{8 - 4\gamma}{n} + \frac{6}{n^2} - \frac{25}{3n^3} - \frac{24}{n^4} - \frac{3119}{30n^5} - \frac{600}{n^6} - \dots$$

$$p_{n,4,2} \sim 1 - \frac{4 \log n}{n} + \frac{4 - 4\gamma}{n} - \frac{2}{n^2} + \frac{13}{3n^3} + \frac{12}{n^4} + \frac{1559}{30n^5} + \frac{300}{n^6} + \dots$$

Des développements analogues sont possibles pour les dimensions supérieures.

## Références

- [1] F.Bergeron, P.Flajolet, B.Salvy, "Varieties of Increasing Trees," Rapport de recherche no. 176, mars 1992, LACIM/UQAM, Montréal (Québec).
- [2] B.C.Berndt, *Ramanujan's notebook, Part I*, Springer-Verlag, 1985.
- [3] B.W. Char, K.O.Geddes, G.H. Gonnet, B.L. Leong, M.B. Monagan, S.M. Watt, *MAPLE V Language Reference Manual*, Springer-Verlag, 1991.
- [4] L.Devroye, L.Laforest, "An Analysis of Random d-dimentional quadtrees," *SIAM J. Comput.*, 19 (5) 1990, pp. 821-832.
- [5] R.A. Finkel, J.L. Bentley, "Quad trees: a data structure for retrieval on composite keys," *Acta Informatica*, vol. 4, pp. 1-9, 1974.
- [6] P.Flajolet, communication par courrier-électronique, avril 1991.
- [7] P.Flajolet, G.Gonnet, C.Puech, J.M.Robson, "Analytic Variations on Quad-Trees," 1991 (à paraître).
- [8] E. Gebhardt, "Anzahlbestimmung und Konstruktion von Schaltfunktionen," Diplomarbeit, Mathematik, Aachen, 1977.
- [9] R. L. Graham, D. E. Knuth, O. Patashnik, *Concrete Mathematics, A Foundation for Computer Science*, Addison-Wesley, mai 1989.
- [10] P. Henrici, "Applied and Computational Complex Analysis," vol. 2, John Wiley & Sons, 1977.
- [11] M.Hoshi, P. Flajolet, "Page Usage in Quadtree Indexes," 1991, (à paraître).
- [12] G.Labelle, "Une nouvelle démonstration des formules d'inversion de Lagrange," *Adv. in Math.*, 42(3), 1981, pp. 217-247.
- [13] G.Labelle, L.Laforest, "Variations combinatoires autour des arborescences hyperquaternaires," Publications du LACIM/UQAM, Montréal, (à paraître) 1992.
- [14] G.Labelle, L.Laforest, "Étude asymptotique du nombre moyen de noeuds à un enfant dans un arbre quaternaire," Rapport de recherche no. 134, octobre 1990, LACIM/UQAM, Montréal (Québec).
- [15] L. Laforest, *Étude des arbres hyperquaternaires*, Thèse de doctorat, Université McGill, Juin 1990, Publications du LACIM/UQAM, Montréal, 1990.
- [16] P. Leroux, X.G.Viennot, Combinatorial Resolution of Systems of Differential Equations I : Ordinary Differential Equations, dans *combinatoire énumérative*, Montréal (Québec), 1985, Proceedings, G.Labelle et P.Leroux éditeurs, Lectures Notes in Math., no. 1234, Springer-Verlag, Heidelberg (1986), pp 210-245.
- [17] L.Lewin, *Polylogarithms and Associated Functions*, North-Holland, New York, 1981.
- [18] H.M. Mahmoud, "The expected distribution of degrees in random binary trees," *The Computer Journal*, vol. 29, pp. 36-37, 1986.
- [19] H. Samet, "The quadtree and related hierarchical data structures," *Computing Surveys*, vol. 16, pp. 187-260, 1984.
- [20] J. Wimp, *Computation with recurrence relations*, Pitman Advanced Publishing Program, Pitman Publ. Ltd., 1984.

