

Numéro d'ordre : 4180

Thèse présentée pour l'obtention du

*Doctorat de l'Université des Sciences et Technologies de Lille
spécialité Informatique*

Calcul symbolique non commutatif : analyse des constantes d'arbre de fouille

Christian COSTERMANS

Université des Sciences et Technologies de Lille

Présentée et soutenue publiquement le 05 juin 2008,
Devant le jury composé de

Présidente :

Brigitte VALLÉE, Directrice de recherche CNRS,

Université de Caen

Rapporteurs :

Jacky CRESSON, Professeur,

Université de Pau

Gérard DUCHAMP, Professeur,

Université Paris 13

Joris van der HOEVEN, Chargé de recherche CNRS, HDR,

Université Paris 11

Directeur :

HOANG NGOC MINH, Professeur,

Université Lille 2

Examineurs :

Marie-Claude VIANO, Professeur Émérite,

Université Lille 1

Sophie TISON, Professeur,

Université Lille 1

Numéro d'ordre : 4180

Thèse présentée pour l'obtention du

Doctorat de l'Université des Sciences et Technologies de Lille
spécialité Informatique

Calcul symbolique non commutatif : analyse des constantes d'arbre de fouille

Christian COSTERMANS

Université des Sciences et Technologies de Lille

Présentée et soutenue publiquement le 05 juin 2008,
Devant le jury composé de

Présidente :

Brigitte VALLÉE, Directrice de recherche CNRS,

Université de Caen

Rapporteurs :

Jacky CRESSON, Professeur,

Université de Pau

Gérard DUCHAMP, Professeur,

Université Paris 13

Joris van der HOEVEN, Chargé de recherche CNRS, HDR,

Université Paris 11

Directeur :

HOANG NGOC MINH, Professeur,

Université Lille 2

Examineurs :

Marie-Claude VIANO, Professeur Émérite,

Université Lille 1

Sophie TISON, Professeur,

Université Lille 1

Table des matières

Introduction	1
1 Arbres hyperquaternaires de points	5
1.1 Récurrence fondamentale pour les paramètres additifs	7
1.1.1 Probabilités d'éclatement	8
1.1.2 Moyenne des paramètres additifs	10
1.1.3 Arité de la racine	10
1.2 Séries génératrices et transformation d'Euler	13
1.2.1 Analyse de singularités	13
1.2.2 Récurrence fondamentale, forme intégrale	15
1.2.3 Transformation d'Euler	16
1.2.4 Analyse en moyenne des paramètres additifs	18
1.3 Étude de la distribution	19
1.3.1 Du nombre de feuilles	20
1.3.2 D'autres paramètres	21
2 Quelques éléments d'algèbre non commutative	23
2.1 Combinatoire des mots	24
2.2 Mots de Lyndon et théorème de Radford	26
2.3 Algèbres de Lie	27
2.3.1 Généralités	27
2.3.2 Formes crochetées et bases duales	28
2.4 Algèbres de Hopf	29
3 Fonctions polylogarithmes	33
3.1 Codage des polylogarithmes	33
3.2 Relation de mélange	34
3.3 Propriétés de la série génératrice	35
4 Des fonctions symétriques aux sommes harmoniques	39
4.1 Algèbre des sommes harmoniques	40
4.1.1 Fonctions symétriques	40
4.1.2 Fonctions quasi-symétriques	40

4.1.3	Sommes harmoniques multiples	42
4.1.4	Séries génératrices sur Y	44
4.2	Analyse asymptotique	46
4.2.1	Par la série génératrice	46
4.2.2	Par la formule d'Euler-MacLaurin	52
4.2.3	Constantes associées à des polyzêtas divergents	56
4.2.4	Application pour l'amélioration des algorithmes	59
4.2.5	Efficacités comparées	61
5	Applications	63
5.1	Aux arbres hyperquaternaires	65
5.1.1	Autour de la racine	65
5.1.2	Retour sur les paramètres additifs	69
5.2	Aux maxima pour des données multidimensionnelles	74
5.2.1	Introduction	74
5.2.2	Équivalent asymptotique de $\mathbb{V}ar(K_{n,d})$	76
5.2.3	Détermination du développement complet	78
	Conclusion	81
	A Valeurs des constantes κ_d	83
	B Expression exacte de $\mathbb{V}ar(K_{n,d})$	85
	C Développement asymptotique de $\mathbb{V}ar(K_{n,d})$	87
	Index	97

Introduction

"Les gens qui analysent les algorithmes ressentent une joie double. Ils expérimentent d'abord la beauté d'élégantes structures mathématiques qui entourent d'élégantes procédures informatiques. Puis ils reçoivent une récompense en retour lorsque leurs théories rendent possibles de résoudre d'autres problèmes plus rapidement et plus économiquement."

Ces lignes, qui constituent un avant-propos du livre "An introduction to the analysis of algorithms" ([25]), sont dûes à un certain D.E. Knuth, et apparaissent très appropriées vis-à-vis de la méthode générale qui nous a guidés au cours de ce mémoire, à savoir l'application de méthodes *symboliques* pour l'étude de constantes apparaissant dans des problèmes de fouille de données aléatoires. En effet, Knuth fut à l'origine de la popularisation des notations dites "en \mathcal{O} ," mais regrettait déjà en 1976 [38] la faible utilisation des notations en Ω et en Θ . Pourtant, dans certains cas, les calculs de complexité moyenne peuvent prendre une forme exacte. Un exemple classique et fondamental de suite particulière permettant de caractériser une certaine complexité est celui des nombres harmoniques d'ordre $r \geq 1$, $H_r(N)$ (ou *nombres harmoniques généralisés* si $r \geq 2$),

$$H_r(N) = \sum_{k=1}^N \frac{1}{k^r}.$$

Ainsi, l'algorithme de tri rapide Quicksort, sur N éléments générés aléatoirement nécessite, en moyenne [25], $2(N+1)(H_1(N+1) - 1)$ comparaisons. De plus, la variance de ce nombre de comparaisons est donné explicitement par $7N^2 - 4(N+1)H_2(N) - 2(N+1)H_1(N) + 13N$ [39]. Autre exemple, dans un arbre binaire de recherche aléatoire, le coût moyen d'une recherche infructueuse est $2H_1(N+1) - 2$ et la variance de ce coût est $2H_1(N+1) - 4H_2(N+1) + 2$.

Par ailleurs, Euler a utilisé sa formule sommatoire (également découverte par la

suite, et de manière indépendante, par Mac-Laurin) pour obtenir les développements

$$\sum_{n=1}^N \frac{1}{n} = \log N + \gamma - \sum_{j=1}^{k-1} \frac{B_j}{j} \frac{1}{N^j} + \mathcal{O}\left(\frac{1}{N^k}\right),$$

$$\sum_{n=1}^N \frac{1}{n^r} = \zeta(r) - \sum_{j=r-1}^{k-1} \frac{B_{j-r+1}}{j} \binom{j}{r-1} \frac{1}{N^j} + \mathcal{O}\left(\frac{1}{N^k}\right),$$

où les B_i sont les nombres de Bernoulli et γ (respectivement $\zeta(r), r \geq 2$) la constante d'Euler-Mac Laurin associée au nombre harmonique divergent $H_1(N)$ (respectivement au nombre harmonique convergent $H_r(N), r \geq 2$) [29]. Ces résultats nous permettent d'obtenir des équivalents (en réalité des développements aussi précis que souhaité) des moyennes et variances précédentes dans l'échelle, dite de Bertrand, des $\{N^\alpha \ln^\beta(N), \alpha \in \mathbb{Z}, \beta \in \mathbb{Z}\}$, échelle qui caractérise une complexité "linéaire", "quasi-linéaire", "quadratique", etc.

Récemment, et pour des données multidimensionnelles, l'application de stratégies du type *diviser pour régner* à des algorithmes et des structures de données hiérarchiques sur les arbres ont conduit plusieurs auteurs à l'étude des sommes harmoniques multiples, associées à un multi-indice (s_1, \dots, s_r) [26, 41]

$$H_{s_1, \dots, s_r}(N) = \sum_{N \geq n_1 > \dots > n_r > 0} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}$$

dont la fonction génératrice ordinaire, P_{s_1, \dots, s_r} , est une fonction polylogarithme, à un facteur $\frac{1}{1-z}$ près :

$$P_{s_1, \dots, s_r}(z) = \sum_{N \geq 1} H_{s_1, \dots, s_r}(N) z^N = \frac{\text{Li}_{s_1, \dots, s_r}(z)}{1-z}$$

avec

$$\text{Li}_{s_1, \dots, s_r}(z) = \sum_{n_1 > \dots > n_r > 0} \frac{z^{n_1}}{n_1^{s_1} \dots n_r^{s_r}}.$$

L'objet du chapitre 1 est de préciser en détail un contexte particulier dans lequel apparaissent ces objets, celui des arbres hyperquaternaires aléatoires, construits à partir de suites de points dans \mathbb{R}^d . Cette structure, définie par une décomposition récursive de l'espace, et les paramètres additifs qui lui sont liés (nombre de feuilles, de noeuds à 1 enfant, longueur de cheminement) ont déjà fait l'objet de nombreuses publications [45, 23, 41, 8], et les techniques utilisées (tout du moins pour des arbres en dimension $d \geq 3$) font appel à des techniques d'analyse complexe, plus précisément d'analyse de singularités, dont nous rappelons les grandes lignes en section 1.2.1 page 13.

Tandis que le chapitre 2 constitue un survol de la combinatoire des mots - qui va constituer l'ossature de notre approche symbolique des quantités rencontrées au chapitre précédent -, et nous permet de présenter les produits de mélange et les structures

algébriques associées, i.e. les *algèbres de mélange*, qui interviennent dans les *relations de mélange* vérifiées par les fonctions polylogarithmes, ou bien par les sommes harmoniques multiples, le chapitre 3 rentre précisément dans le vif de la démarche, en adoptant un codage symbolique pour les polylogarithmes. Ceci permet d'établir que la \mathbb{C} -algèbre engendrée par les polylogarithmes $\text{Li}_{s_1, \dots, s_r}(z)$ et par les fonctions logarithmes $\log^n(z)$, $n \geq 1$, est isomorphe à une \mathbb{C} -algèbre de mélange, sur l'alphabet à deux lettres $X = \{x_0, x_1\}$ qui est une algèbre libre et possède une base constituée par les mots de Lyndon (Théorème de Radford 2.2.3 page 26). Ce point crucial permet d'obtenir des algorithmes pour le calcul de la monodromie [60], du groupe de Galois différentiel [53], d'équations fonctionnelles de type Kummer [61] pour ces polylogarithmes, via leur série génératrice non commutative.

La description de l'algèbre engendrée par les sommes harmoniques multiples $H_s(N)$ (vues comme des fonctions du paramètre N) nous conduit, au chapitre 4, à un nouveau codage sur un alphabet infini $Y = \{y_i\}_{i \geq 1}$. En effet, à chaque multi-indice (s_1, \dots, s_r) , nous associons un mot $w = y_{s_1} \dots y_{s_r}$ sur l'alphabet Y . De la sorte, les polylogarithmes, les sommes harmoniques et leur fonction génératrice peuvent être indexées par des mots du monoïde Y^* :

$$H_w = H_{s_1, \dots, s_r}, \quad \text{Li}_w = \text{Li}_{s_1, \dots, s_r}, \quad P_w = P_{s_1, \dots, s_r}.$$

Comme précédemment, cette stratégie nous permet de prouver que la \mathbb{C} -algèbre de Hadamard des fonctions génératrices P_w des sommes harmoniques est isomorphe à une \mathbb{C} -algèbre de mélange, sur l'alphabet Y , qui possède également une base (Théorème 2.2.4 page 26). De plus, $\text{Li}_w(z)$ et $H_w(N)$ sont reliés par le point suivant, lorsque $w \in Y^* \setminus y_1 Y^*$: la limite de $\text{Li}_w(z)$, quand $z \rightarrow 1$, et celle de $H_w(N)$, quand $N \rightarrow \infty$, existent, et par le théorème d'Abel, sont égales :

$$\lim_{z \rightarrow 1} \text{Li}_w(z) = \lim_{N \rightarrow \infty} H_w(N) = \zeta(w),$$

$\zeta(w)$ étant le polyzêta

$$\zeta(w) = \zeta(s_1, \dots, s_r) = \sum_{n_1 > \dots > n_r > 0} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}.$$

Afin d'étudier les cas divergents, i.e. $w \in y_1 Y^*$, nous considérons les séries génératrices non commutatives

$$\Lambda(z) = \sum_{w \in Y^*} \text{Li}_w(z) w \quad \text{et} \quad H(N) = \sum_{w \in Y^*} H_w(N) w$$

et nous prouvons alors par des techniques "à la Hopf" le théorème suivant "à l'Abel" (Théorème 4.1.22 page 46)

$$\lim_{z \rightarrow 1} e^{-y_1 \text{Li}_{y_1}(z)} \Lambda(z) = \lim_{N \rightarrow \infty} \exp\left(\sum_{k \geq 1} H_{y_k}(N) \frac{(-y_1)^k}{k}\right) H(N) = S,$$

où S désigne la série génératrice non commutative des polyzêtas convergents. Ceci nous permet d'expliciter les constantes d'Euler généralisées associées à des polyzêtas divergents $\{\zeta_{\sqcup}(w)\}_{w \in y_1 Y^*}$, [14] et d'obtenir le développement asymptotique de $H_w(n)$. Nous prouvons que ces constantes appartiennent à la \mathbb{Q} -algèbre engendrée par la constante d'Euler γ et par les polyzêtas convergents $\{\zeta(w)\}_{w \in Y^* \setminus y_1 Y^*}$. Par ailleurs, pour obtenir le développement de $H_w(N)$, les propriétés structurelles de la série génératrice des $\{Li_w, w \in X^*\}$, et particulièrement son comportement lorsque $z \rightarrow 1$ donne également naissance à un algorithme propre [10]. En outre, l'existence d'une base pour l'algèbre de mélange sur Y , associé à l'isomorphie avec l'algèbre des sommes harmoniques donne également naissance à un troisième algorithme [11]. Nous sommes alors amenés à discuter chacune des procédures, à les comparer en termes de temps de calcul, et à conclure à l'amélioration significative apportée par le calcul explicite des constantes d'Euler généralisés.

Enfin, le chapitre 5 est consacré aux applications de cette approche symbolique. Dans un premier temps, la section 5.1 reprend les éléments dont nous étions partis au chapitre 1, et propose des simplifications pour le calcul de probabilités intervenant dans l'arité de la racine d'un arbre hyperquaternaire, puis pour les paramètres additifs qui en découlent. Ensuite, dans la section 5.2, nous nous intéressons à un autre contexte de données multidimensionnelles, celui du nombre de maxima au sein d'un ensemble de variables aléatoires indépendantes et identiquement distribués, à valeurs dans \mathbb{R}^d . Le problème du nombre de maxima intervient dans divers domaines [68], celui de la recherche de maxima en géométrie algorithmique, en théorie des graphes, en économétrie, optimisation multi-critère... Bien que cette variable diffère du problème précédent en ceci que le nombre de maxima ne dépend absolument pas de l'ordre d'arrivée des points, il s'avère que l'étude de la variance de cette variable fait apparaître des variations de sommes harmoniques dont la similarité avec le domaine précédent a déjà été notée par des spécialistes des arbres hyperquaternaires [44]. Notre approche permet alors de préciser la nature algébrique des coefficients intervenant dans son développement asymptotique (Théorème 5.2.11 page 80). De plus, nous donnons une valeur explicite pour la constante dominante de ce développement (Théorème 5.2.3 page 77). Des exemples de calcul de ces constantes, ainsi qu'une forme close de ces variances sont donnés en annexes.

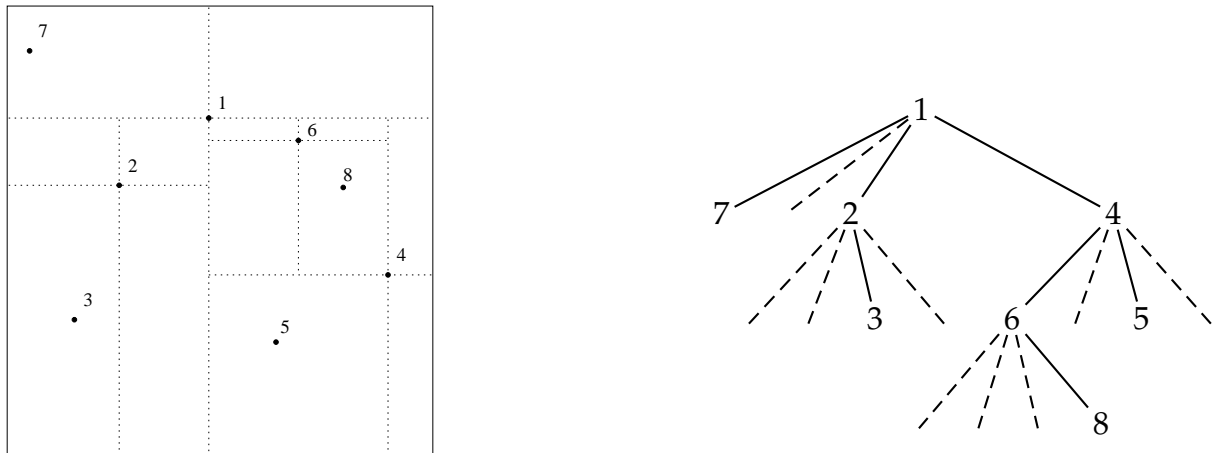
Chapitre 1

Arbres hyperquaternaires de points

Sommaire

1.1	Réurrence fondamentale pour les paramètres additifs	7
1.1.1	Probabilités d'éclatement	8
1.1.2	Moyenne des paramètres additifs	10
1.1.3	Arité de la racine	10
1.2	Séries génératrices et transformation d'Euler	13
1.2.1	Analyse de singularités	13
1.2.2	Réurrence fondamentale, forme intégrale	15
1.2.3	Transformation d'Euler	16
1.2.4	Analyse en moyenne des paramètres additifs	18
1.3	Étude de la distribution	19
1.3.1	Du nombre de feuilles	20
1.3.2	D'autres paramètres	21

Parmi les nombreux cadres possibles dans lesquels s'inscrit la fouille de données, celui d'un ensemble massif de données spatiales constitue un cadre propice à l'utilisation d'arbres de partitionnement de l'espace. Pour des données totalement ordonnées suivant une unique dimension, la structure la plus simple tant pour le codage que pour l'exploitation se nomme arbre binaire de recherche. Sa généralisation à des données modélisables comme des points de \mathbb{R}^d peut prendre plusieurs formes : arbre m -aires de recherche (où chaque noeud peut contenir jusqu'à $m - 1$ clés), arbres de type $k - d$ (arbres binaires basées sur une division de l'espace alternativement selon une des d dimensions), arbres hyperquaternaires - ces derniers pouvant eux-même être *de points* ou *de région*-. Ces types d'arbres appartiennent à la famille des arbres de partitionnement, mais différent précisément quant à la manière de partitionner l'espace. Il convient de préciser également qu'en 1998, Devroye [17] a défini une structure d'arbre-grille (*grid*)

FIG. 1.1 – Arbre quaternaire construit sur $n = 8$ points

tree) en plaçant les $m - 1$ premiers points dans la racine, ces $m - 1$ points divisant l'espace en, au plus, m^d régions (des grilles). Pour $m = 2$, cette structure se confond avec l'arbre hyperquaternaire, tandis que pour $d = 1$, elle se confond avec l'arbre de recherche m -aire.

Le principe de la construction de l'arbre quaternaire de points en dimension $d = 2$ est le suivant : on considère n points (X_1, \dots, X_n) répartis dans $[0, 1]^2$. Cette suite est représentée par un arbre quaternaire construit récursivement suivant les règles suivantes :

- Si $n = 0$, l'arbre est vide
- Si $n \geq 1$, le point X_1 constitue la racine de l'arbre, et les quatre sous-arbres de la racine sont construits récursivement à partir des ensembles de points constitués par les quadrants nord-ouest, nord-est, sud-ouest et sud-est déterminés à partir de X_1 .

La construction se généralise immédiatement à des points répartis dans $[0, 1]^d$, en remplaçant les quatre quadrants par les 2^d hyperquadrants induits par le point X_1 .

Ainsi, les arbres hyperquaternaires, ou quadtree, ou arbres quadrants (parfois nommés octrees dans le cas de la dimension $d = 3$) constituent, selon une définition de Samet, de 1984 [67], une structure de données hiérarchique basée sur une décomposition récursive de l'espace. En fait, cette structure fut introduite pour la première fois par Finkel et Bentley en 1974 [21], et un peu auparavant, sous l'appellation Q -tree, en 1971, par Klinger [37], sous la forme d'arbre quaternaire dit *de région*, afin de représenter des images binaires codées sous formes de pixels blanc (0) ou noir (1), que l'on va ensuite regrouper sous forme de plus grands carrés possibles ne contenant que des 0 ou que des 1. Les différents types d'arbre hyperquaternaires, *de points* ou *de régions*, sous différentes variantes passées en revue par H. Samet dans [50] trouvent d'ailleurs la majeure partie de leurs applications dans le domaine de la géométrie algorithmique [16], du traitement d'images, mais également dans des problèmes de clustering pour des données géographiques [63] ou biologiques [46]. Il est possible de se faire une idée des

différentes variantes dans la construction de ces arbres à travers des applets disponibles sur internet [6], et dues à F. Brabek et H. Samet.

1.1 Récurrence fondamentale pour les paramètres additifs

Une étude approfondie de l'arbre hyperquaternaire aléatoire *de points* a été réalisée par Laforest dans sa thèse, en 1990 [45]. L'étude probabiliste de cet arbre peut s'effectuer dans un cadre très général, celui de points répartis dans \mathbb{R}^d , à d coordonnées indépendantes et suivant une distribution continue. En fait, sans perte de généralité, cette étude se ramène au cas de variables aléatoires indépendantes, et uniformément réparties dans $[0, 1]^d$.

Notation. Soit T un arbre hyperquaternaire. Nous désignons par $|T|$ la "taille" de l'arbre, c'est-à-dire le nombre de noeuds de l'arbre T (correspondant au nombre de points à partir duquel est construit l'arbre). Nous noterons par ailleurs $\{T_i, 1 \leq i \leq 2^d\}$ les 2^d sous-arbres (éventuellement vides) enfants de la racine.

Notre objectif dans ce chapitre est d'étudier des paramètres additifs définis sur les arbres hyperquaternaires.

Nous désignons par paramètre additif une fonction f , associant un arbre à un nombre réel, telle que pour tout arbre hyperquaternaire T en dimension d quelconque, la relation suivante soit vérifiée

$$\begin{cases} f(T) = \sum_{k=1}^{2^d} f(T_k) + t_{|T|} & \text{si } |T| \geq 1, \\ f(\emptyset) = t_0, \end{cases} \quad (1.1)$$

$(t_n)_{n \geq 1}$ désignant une "suite-test".

Exemple 1.1.1. Quelques suites-test caractéristiques

- Si $t_0 = 0$, et $t_n = 1$ pour tout $n \geq 1$, alors $f(T) = |T|$ calcule le nombre de noeuds de l'arbre,
- Si $t_n = \delta_{n,1}$ (symbole de Kronecker), alors $f(T)$ correspond au nombre de feuilles de l'arbre,
- Si $t_n = n$, $f(T)$ mesure la longueur de cheminement de l'arbre (avec la convention qu'un arbre réduit à sa racine a une longueur de cheminement de 1.)

Ces paramètres très généraux correspondent en fait au coût d'algorithmes de type "diviser pour régner" dont la philosophie consiste à diviser un problème en sous-problèmes, à les résoudre récursivement, puis utiliser les solutions pour résoudre le problème de départ. L'archétype de cette famille d'algorithmes est l'algorithme de tri rapide *Quicksort*, algorithme dont Chern, Hwang et Tsai ont remarqué [9] que la mise

en application pouvait se traduire par la construction d'un arbre binaire de recherche, et qu'il y avait une forte correspondance entre le coût algorithmique d'un Quicksort et certains paramètres sur les arbres (plus précisément, cette remarque s'applique à une famille d'algorithmes regroupés sous le nom Quicksort, et pas uniquement le Quicksort *stricto sensu*).

C'est donc l'étude de ces quantités, de ces paramètres (nombre de feuilles, nombre de noeuds à 1, 2 fils, longueur de cheminement ...) qui va nous occuper à présent, étude pour laquelle nous allons présenter différents résultats, faisant appel à différentes branches de l'analyse. Le point de départ pour la conduite de cette étude, consiste à déterminer la récurrence fondamentale vérifiée par les paramètres additifs, récurrence impliquant la taille des sous-arbres d'un arbre hyperquaternaire.

1.1.1 Probabilités d'éclatement

L'étude réalisée par Laforest a mis en évidence la probabilité pour le i -ème sous-arbre issu de la racine ($1 \leq i \leq 2^d$) d'être de cardinal $|T_i|$ égal à k , probabilité en réalité indépendante de i , sous la forme suivante

$$\mathbb{P}(|T_i| = k) = \frac{1}{N} \sum_{N \geq i_1 \geq \dots \geq i_{d-1} \geq k+1} \frac{1}{i_1 \dots i_{d-1}}, \quad 0 \leq k \leq N-1, \quad (1.2)$$

pour un arbre construit à partir de N points indépendants, et uniformément distribués sur $[0, 1]^d$. En précisant la notation $|T_i|$ en $|T_i^{(d)}|$, d désignant la dimension de l'hypercube dans lequel vivent nos points, la formule précédente se déduit aisément de la relation de récurrence suivante :

$$\mathbb{P}(|T_i^{(d)}| = k) = \sum_{m=k}^{N-1} \frac{1}{m+1} \mathbb{P}(|T_i^{(d-1)}| = m), \quad (1.3)$$

obtenue en considérant la projection du nuage de points sur $[0, 1]^{d-1}$, et en se souvenant qu'en dimension 1, on dispose d'une distribution uniforme sur $\{0, 1, \dots, N\}$ pour la taille de chacun des deux sous-arbres, i.e.

$$\mathbb{P}(|T_i^{(1)}| = k) = \frac{1}{N}.$$

La formule de l'équation (1.2) admet une nuée d'expressions jumelles, dont voici quelques exemples, donnés par Labelle et Laforest dans [41]

Theorème 1.1.2 ([41]). *En dimension d , les probabilités $\pi_{n,k,d}$ qu'une arborescence aléatoire hyperquaternaire de n points possède exactement k noeuds dans un hyperquadrant donné sont fournies par les formules explicites et récursives suivantes, pour $d \geq 1$,*

$$0 \leq k \leq n - 1,$$

$$\pi_{n,k,1} = \frac{1}{n} \quad (1.4)$$

$$\pi_{n,k,d+1} = \sum_{m=k+1}^n \frac{\pi_{n,m-1,d}}{m} \quad (1.5)$$

$$\pi_{n,k,d} = \binom{n-1}{k} \int_0^1 \dots \int_0^1 \left(\prod_{i=1}^d u_i \right)^k \left(1 - \prod_{i=1}^d u_i \right)^{n-1-k} du_1 \dots du_d \quad (1.6)$$

$$\pi_{n,k,d} = \binom{n-1}{k} \sum_{i=0}^{n-1-k} \binom{n-1-k}{i} \frac{(-1)^i}{(k+i+1)^d} \quad (1.7)$$

L'expression (1.4) vient du fait déjà évoqué que dans un arbre binaire de recherche de taille n , la taille du sous-arbre gauche (ou droit) suit une loi uniforme sur $\{0, \dots, n-1\}$. L'expression (1.6) s'obtient en conditionnant la variable $|T_1|$ par X_1 , de sorte que

$$\mathbb{P}(|T_1| = k | X_1) = \binom{n-1}{k} \left(\prod_{i=1}^d u_i \right)^k \left(1 - \prod_{i=1}^d u_i \right)^{n-1-k}$$

(avec $X_1 = (u_1, \dots, u_d)$ et en supposant, sans perte de généralité, que T_1 correspond à l'hyperquadrant dont un sommet est $(0, \dots, 0)$). Il ne reste alors plus qu'à procéder à l'intégration sur $[0, 1]^d$, dont le calcul, après un changement de variables, fournit l'expression (1.7).

Aussi, introduisons dès à présent la notion de somme harmonique multiple, généralisant la définition des nombres harmoniques pour un multi-indice.

Définition 1.1.3. Soit $\mathbf{s} = (s_1, \dots, s_r) \in (\mathbb{N}^*)^r$, nous définissons

$$H_{\mathbf{s}}(N; k) = \sum_{N \geq n_1 > \dots > n_r \geq k} \frac{1}{n_1^{s_1} \dots n_r^{s_r}},$$

$$\underline{H}_{\mathbf{s}}(N; k) = \sum_{N \geq n_1 \geq \dots \geq n_r \geq k} \frac{1}{n_1^{s_1} \dots n_r^{s_r}},$$

et pour des raisons de simplicité, nous préférons la notation $H_{\mathbf{s}}(N)$ (respectivement $\underline{H}_{\mathbf{s}}(N)$) à la notation $H_{\mathbf{s}}(N; 1)$ (respectivement à $\underline{H}_{\mathbf{s}}(N; 1)$).

La définition 1.1.3 nous permet de rajouter une nouvelle forme à notre catalogue (non exhaustif) de formulations, à savoir

$$\pi_{n,k,d} = \frac{1}{n} \underline{H}_{\mathbf{s}_{d-1}^{(1)}}(n; k+1),$$

avec $\mathbf{s}_{d-1}^{(1)} = (1, 1, \dots, 1) \in \mathbb{R}^{d-1}$.

1.1.2 Moyenne des paramètres additifs

Notons f_n l'espérance du paramètre additif $f(T)$ associé à la suite-test $(t_n)_{n \geq 1}$, sous le modèle de distribution uniforme de n points dans $[0, 1]^d$. En tenant compte des propriétés de symétrie entre les différents sous-arbres, la récurrence générale (1.1) se transforme aisément en

$$f_n = t_n + 2^d \sum_{k=0}^{n-1} \pi_{n,k,d} f_k, \quad (1.8)$$

Démonstration. Comme les 2^d sous-arbres T_i sont indépendants et de même nature, on obtient, en calculant l'espérance des deux membres de l'équation (1.1)

$$\begin{aligned} \mathbb{E}(f(T)) &= \mathbb{E}(t_{|T|}) + 2^d \mathbb{E}(f(T_1)), \\ &= t_n + 2^d \sum_{k=0}^{n-1} \mathbb{E}(f(T_1) | |T_1| = k) P(|T_1| = k), \\ &= t_n + 2^d \sum_{k=0}^{n-1} \pi_{n,k,d} f_k. \end{aligned}$$

□

Un des premiers paramètres de ce type auxquels Labelle et Laforest se sont intéressés fut le nombre de noeuds ayant k enfants. L'étude de ce paramètre suppose préalablement la précision de la suite-test $(t_n)_{n \geq 0}$ associée, qui correspond simplement à la fonction indicatrice de l'événement "la racine est d'arité k ".

1.1.3 Arité de la racine

Reprenant les notations adoptées dans [42, 44], étant donné l'hypercube $[0, 1]^d$, un point initial X_1 sépare l'hypercube en 2^d hyperquadrants, et nous indexons par le mot binaire $\epsilon_1 \dots \epsilon_d \in \{0, 1\}^d$, l'hyperquadrant contenant le sommet $(\epsilon_1, \dots, \epsilon_d)$.

Notation. Soit S un ensemble de mots binaires codant des hyperquadrants, et n variables aléatoires indépendantes et uniformément distribués dans $[0, 1]^d$, nous désignons par $J_n[S]$ la probabilité que ces n points tombent dans l'ensemble S , et par $J_n^*[S]$ cette même probabilité, avec la contrainte supplémentaire que chaque hyperquadrant déterminé par S contienne au moins un point.

Exemple 1.1.4. En dimension 2, soit $X_1 = (t_1, t_2)$ aléatoire et $S = \{01, 11\}$ (codant les deux quadrants "nord"), alors

$$J_n[S] = \int_0^1 \int_0^1 (t_1(1-t_2) + (1-t_1)(1-t_2))^{n-1} dt_1 dt_2.$$






Région	Codage correspondant
	$S=\{\}$
	$S=\{00\}$
	$S=\{01\}$
	$S=\{10\}$
	$S=\{11\}$

FIG. 1.2 – Codage binaire en dimension $d = 2$

Plus généralement, avec les notations $t_i^{<0>} = t_i$ et $t_i^{<1>} = 1 - t_i$, en dimension d quelconque, la probabilité $J_n[S]$ est donnée par l'intégrale multiple suivante

$$J_n[S] = \int_0^1 \dots \int_0^1 f_S(t_1, \dots, t_d)^{n-1} dt_1 \dots dt_d, \quad \text{où } f_S(t_1, \dots, t_d) = \sum_{\epsilon \in S} \prod_{i=1}^d t_i^{\epsilon_i} \quad (1.9)$$

Proposition 1.1.5. La probabilité $p_{n,k,d}$ que la racine d'un arbre hyperquaternaire aléatoire de n points soit d'arité k est donné par

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

Démonstration. Les probabilités $J_n[S]$ et $J_n^*[S]$ se déduisent l'une de l'autre simplement par la relation

$$J_n[S] = \sum_{S' \subseteq S} J_n^*[S'], \quad \text{d'où } J_n^*[S] = \sum_{S' \subseteq S} (-1)^{|S|-|S'|} J_n[S'].$$

Par conséquent

$$\begin{aligned}
 p_{n,k,d} &= \sum_{|S|=k} J_n^*[S] \\
 &= \sum_{|S|=k} \sum_{S' \subseteq S} (-1)^{|S|-|S'|} J_n[S'] \\
 &= \sum_{j=0}^k \sum_{|S|=k} \sum_{S' \subseteq S, |S'|=j} (-1)^{k-j} J_n[S'] \\
 &= \sum_{j=0}^k \sum_{|S'|=j} (-1)^{k-j} J_n[S'] \sum_{|S|=k, S' \subseteq S} 1,
 \end{aligned}$$

le nombre de sur-ensembles de cardinal k d'un ensemble S' fixé de cardinal j étant de $\binom{2^d - j}{k - j}$, le résultat s'ensuit. \square

En fait, on peut remarquer que des choix différents de S peuvent aboutir au même résultat, par exemple, en dimension 2, $J_n[\{00\}] = J_n[\{01\}] = J_n[\{10\}] = J_n[\{11\}] = \frac{1}{n^2}$. En effet, ceci est le cas dès que les différents ensembles appartiennent à la même orbite sous l'action du groupe hyperoctaédral. Cette remarque permet de limiter les calculs en ne considérant qu'un seul représentant de l'orbite.

On peut également remarquer que remplacer un ensemble S par son complémentaire revient, dans la formule (1.9), à remplacer $f_S(t_1, \dots, t_d)$ par son complémentaire à 1. Par conséquent, \bar{S} désignant le complémentaire de S ,

$$J_n[S] = \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} J_k[\bar{S}]. \quad (1.10)$$

Proposition 1.1.6. *L'arité moyenne $E_{n,d}$ de la racine d'un arbre hyperquaternaire en dimension d à n points vaut*

$$E_{n,d} = 2^d \left(1 - \frac{1}{n} \mathbb{H}_{\mathbb{S}_{d-1}^{(1)}}(n) \right).$$

Démonstration. L'arité de la racine d'un arbre T s'exprime simplement comme la variable aléatoire $\sum_{i=1}^{2^d} \mathbb{1}_{\{T_i \neq 0\}}$, chaque indicatrice suivant une loi de Bernoulli de paramètre $1 - \pi_{n,0,d} = 1 - \frac{1}{n} \mathbb{H}_{\mathbb{S}_{d-1}^{(1)}}(n)$. Le passage à l'espérance est immédiat. \square

La proposition 1.1.5 nous permet de calculer la distribution de l'arité de la racine en fonction des intégrales $J_n[S]$. Dès lors, la connaissance du comportement asymptotique

de l'un suppose la connaissance du comportement asymptotique de l'autre. Labelle et Laforest proposent dans [42] d'exprimer l'intégrale $J_n[S]$ comme la transformée de Laplace en n , $\int_0^{+\infty} e^{-nt} \Phi(t) dt$, d'une certaine fonction Φ définie de la manière suivante :

$$\Phi(t) = \omega(1 - e^{-t}), \quad \text{où} \quad \omega(x) = \frac{d}{dx} \int_{[0,1]^d} \mathbb{1}_{\{f_S(t_1, \dots, t_d) > 1-x\}} dt_1 \dots dt_d. \quad (1.11)$$

En dimension 2, les probabilités $J_n[S]$ prennent les formes suivantes [44] :

Représentant de l'orbite	Taille de l'orbite	$J_n[S]$
{00}	4	$\frac{1}{n^2}$
{00,01}	4	$\frac{1}{n}$
{00,11}	2	$\frac{1}{n^2} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}}$
{00,01,10}	4	$\frac{H_1(n)}{n}$
{00,01,10,11}	1	1

TAB. 1.1 – Valeur de $J_n[S]$ en dimension $d = 2$

et le seul terme pour lequel le passage à l'asymptotique n'est pas immédiat est celui relatif à l'ensemble {00, 11}. Nous allons précisément voir qu'il est possible de contourner les calculs "à la Laplace" via des techniques d'analyse de singularités.

1.2 Séries génératrices et transformation d'Euler

Les techniques que nous allons introduire maintenant ont été développés par Flajolet, Sedgewick, Odlyzko et d'autres depuis la fin des années 80. On pourra se référer à l'article de Flajolet et Odlyzko de 1988 [24] ou au chapitre VI du livre de Flajolet [22].

1.2.1 Analyse de singularités

Partons d'un exemple simple, en considérant le cas $d = 1$ et notons notre fonctionnelle $f(T) = X_n$, pour $|T| = n > 0$. On peut alors reformuler l'équation (1.8) ainsi :

$$f_n = t_n + \frac{2}{n} \sum_{k=0}^{n-1} f_k,$$

avec la condition initiale $f_0 = t_0$.

Le problème qui se pose est le suivant : la suite (t_n) étant connue, comment déterminer le comportement asymptotique de f_n ?

Pour cela, on considère les fonctions génératrices $f(z) = \sum_{n \geq 0} f_n z^n$ et $t(z) = \sum_{n \geq 0} t_n z^n$.

L'équation reliant f_n à t_n se transfère alors, via des manipulations classiques, en une équation reliant $f(z)$ à $t(z)$:

$$f(z) = t(z) + 2 \int_0^z f(w) \frac{dw}{1-w}$$

équation qui se résout, par dérivation, et par méthode de variation de la constante, en :

$$f(z) = (1-z)^{-2} \int_0^z t'(w)(1-w)^2 dw$$

en supposant, sans perte de généralité, que $f_0 = t_0 = 0$.

Sous cette forme, on peut traiter un premier exemple directement.

Si $t_n = \binom{n+\alpha}{\alpha}$, i.e. $t(z) = (1-z)^{-\alpha-1}$, alors

- Si $\alpha \neq 1$

$$f(z) = \frac{\alpha+1}{\alpha-1} [(1-z)^{-\alpha-1} - (1-z)^{-2}]$$

$$f_n = \frac{\alpha+1}{\alpha-1} \left[\binom{n+\alpha}{\alpha} - (n+1) \right]$$

- Si $\alpha = 1$, alors $f(z) = \frac{2}{(1-z)^2} \log \frac{1}{1-z}$ et $f_n = 2(n+1)(H_1(n+1) - 1)$.

Or, d'après la formule de Stirling, $\binom{n+\alpha}{\alpha} \sim \frac{n^\alpha}{\Gamma(\alpha+1)}$.

On peut donc résumer ces résultats ainsi :

$$t_n = \binom{n+\alpha}{\alpha}, \alpha > 1 \implies f_n \sim \frac{\alpha+1}{\alpha-1} \frac{n^\alpha}{\Gamma(\alpha+1)}$$

$$t_n = n+1 \implies f_n \sim 2n \log n$$

$$t_n = \binom{n+\alpha}{\alpha}, 0 < \alpha < 1 \implies f_n \sim \frac{1+\alpha}{1-\alpha} n$$

Dans cet exemple, le choix de t_n nous a conduit à une expression explicite de $t(z)$ et donc de $f(z)$. Mais si $t_n = \sqrt{n}$ ou $t_n = \ln n$, nous n'aurons plus de formule explicite, et nous devons alors nous contenter d'équivalents asymptotiques, au voisinage de 1, pour $t(z)$.

L'exemple traité nous permet d'esquisser le raisonnement général qui va prévaloir dans les paragraphes suivants :

- Trouver une relation entre f_n et t_n
- En déduire une relation entre $f(z)$ et $t(z)$
- Chercher un développement asymptotique pour $f(z)$
- Conclure sur le comportement asymptotique de f_n .

Exemple 1.2.1. Reprenons l'expression de la probabilité $J_n[\{00, 11\}]$ vue au tableau 1.1.3

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}} = \sum_{i=0}^{n-1} \frac{1}{2^i(n-i)} = [z^n] \frac{2}{2-z} \log \left(\frac{1}{1-z} \right).$$

Or, comme

$$\frac{2}{2-z} \log \left(\frac{1}{1-z} \right) = 2 \log \left(\frac{1}{1-z} \right) - 2(1-z) \log \left(\frac{1}{1-z} \right) + \mathcal{O} \left((1-z)^2 \log \left(\frac{1}{1-z} \right) \right),$$

on en déduit

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}} = \frac{2}{n} + \frac{2}{n^2} + \mathcal{O} \left(\frac{1}{n^3} \right).$$

1.2.2 Récurrence fondamentale, forme intégrale

Définissons les fonctions génératrices ordinaires

$$f(z) = \sum_{n \geq 0} f_n z^n, \quad \text{et} \quad t(z) = \sum_{n \geq 0} t_n z^n. \quad (1.12)$$

Theorème 1.2.2 ([23]). Les fonctions génératrices f et t des suites $(f_n)_{n \geq 0}$ et $(t_n)_{n \geq 0}$ vérifiant la récurrence (1.8) vérifient la relation

$$f(z) = t(z) + 2^d J^{d-1} I f(z),$$

les opérateurs intégraux I et J étant définis par

$$I g(z) = \int_0^z \frac{g(t)}{1-t} dt, \quad \text{et} \quad J g(z) = \int_0^z \frac{g(t)}{t(1-t)} dt.$$

Démonstration. Par récurrence sur la dimension d : si $d = 1$, alors $\pi_{n,k,1} = \frac{1}{n}$, et la relation est immédiate à vérifier. Supposons maintenant que $\sum_{n \geq 1} \sum_{k=0}^{n-1} \pi_{n,k,d} f_k z^n = J^{d-1} I f(z)$,

et posons $\alpha(z) = \sum_{n \geq 1} \sum_{k=0}^{n-1} \pi_{n,k,d+1} f_k z^n$.

Dans l'expression de $\pi_{n,k,d}$ fournie par (1.2), il est facile de vérifier, en séparant les cas où le plus grand indice est égal à n et celui où il est inférieur ou égal à $n - 1$, que

$$n\pi_{n,k,d+1} = (n-1)\pi_{n-1,k,d+1} + \pi_{n,k,d},$$

ainsi

$$\sum_{n \geq 1} n \sum_{k=0}^{n-1} \pi_{n,k,d+1} f_k z^n = \sum_{n \geq 1} (n-1) \sum_{k=0}^{n-2} \pi_{n-1,k,d+1} f_k z^n + \sum_{n \geq 1} \sum_{k=0}^{n-1} \pi_{n,k,d} f_k z^n,$$

soit, en traduisant à l'aide de fonctions génératrices

$$\begin{aligned} z \frac{d}{dz} \alpha(z) &= z^2 \frac{d}{dz} \alpha(z) + J^{d-1} I f(z), \text{ d'où} \\ \frac{d}{dz} \alpha(z) &= \frac{1}{z(1-z)} J^{d-1} I f(z) \\ \alpha(z) &= J^d I f(z). \end{aligned}$$

□

1.2.3 Transformation d'Euler

La transformation d'Euler est l'application involutive définie par $f \mapsto f^*$, avec

$$f^*(z) = \frac{1}{1-z} f\left(\frac{z}{z-1}\right).$$

Son équivalent sur les suites est l'endomorphisme de $\mathbb{R}^{\mathbb{N}}$, $(f_n)_n \mapsto (f_n^*)_n$, avec

$$f_n^* = \sum_{k=0}^n (-1)^k \binom{n}{k} f_k.$$

En particulier, si $f(z) = \sum_{n \geq 0} f_n z^n$, alors $f^*(z) = \sum_{n \geq 0} f_n^* z^n$.

Exemple 1.2.3. Si $t_n = \delta_{n,1}$ alors $t_n^* = -n$. De manière équivalente,

$$t(z) = z \implies t^*(z) = -\frac{z}{(1-z)^2}.$$

Exemple 1.2.4. En considérant l'équation (1.10) dans le cas de la dimension $d = 3$ avec $S = \{1, 2, 3, 4, 5, 6, 7\}$ (en confondant la notation binaire et sa traduction décimale), nous obtenons

$$J_n[S] = \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} J_k[\{0\}] = -\sum_{k=1}^n (-1)^k \binom{n}{k} \frac{1}{k^2}.$$

Par conséquent, en définissant la fonction dilogarithme $\text{Li}_2(z) = \sum_{n \geq 0} \frac{z^n}{n^2}$, pour $|z| \leq 1$, il apparaît que

$$J_n[S] = -[z^n] \frac{1}{1-z} \text{Li}_2\left(\frac{z}{z-1}\right).$$

L'intérêt de cette transformation réside dans la propriété suivante : alors que l'espérance f_n d'un paramètre additif satisfait la récurrence (1.8) nécessitant les n termes précédents, la récurrence vérifiée par sa transformée d'Euler est une récurrence d'ordre 1 :

$$f_n^* = t_n^* - t_{n-1}^* + \left(1 - \left(\frac{2}{n}\right)^d\right) f_{n-1}^*. \quad (1.13)$$

Démonstration. En partant de (1.8), et en utilisant l'expression (1.7) de $\pi_{n,k,d}$, il vient

$$\begin{aligned} f_N^* &= t_N^* + 2^d \sum_{n=0}^N (-1)^n \binom{N}{n} \sum_{k=0}^{n-1} \pi_{n,k,d} f_k \\ &= t_N^* + 2^d \sum_{n=0}^N (-1)^n \binom{N}{n} \sum_{k=0}^{n-1} \binom{n-1}{k} \sum_{i=0}^{n-1-k} \binom{n-1-k}{i} \frac{(-1)^i}{(k+i+1)^d} f_k \\ &= t_N^* + 2^d \sum_{k=0}^{N-1} (-1)^k f_k \sum_{i=k}^{N-1} \frac{(-1)^i}{(i+1)^d} \sum_{n=i+1}^N (-1)^n \binom{N}{n} \binom{n-1}{k} \binom{n-1-k}{i-k}, \end{aligned}$$

la somme sur n se révélant indépendante de $N \geq i+1$, il suffit de la calculer en $N = i+1$, d'où

$$\begin{aligned} f_N^* &= t_N^* - 2^d \sum_{k=0}^{N-1} (-1)^k f_k \sum_{i=k}^{N-1} \frac{1}{(i+1)^d} \binom{i}{k} \\ &= t_N^* - 2^d \sum_{i=0}^{N-1} \frac{1}{(i+1)^d} \sum_{k=0}^i (-1)^k \binom{i}{k} f_k \\ &= t_N^* - \sum_{i=0}^{N-1} \left(\frac{2}{i+1}\right)^d f_i^*. \end{aligned}$$

Le résultat s'ensuit par soustraction. \square

Cette simplification de la récurrence a pour conséquence une expression explicite de f_n^* , pour $n \geq 2$ à savoir

$$f_n^* = [n]_d! \sum_{j=2}^n \frac{t_j^* - t_{j-1}^*}{[j]_d!}, \quad (1.14)$$

où $[j]_d!$ représente le d -analogue de $j!$, i.e.

$$[j]_d! = [3]_d \dots [j]_d, \quad \text{avec} \quad [j]_d = 1 - \frac{2^d}{j^d},$$

et avec la convention $[2]_d! = 1$.

Autre conséquence directe, la simplification a également lieu pour l'équation différentielle vérifiée par $f^*(z)$:

$$\left(z \frac{d}{dz}\right)^d \left((1-z)(f^*(z) - t^*(z)) \right) + 2^d z f^*(z) = 0. \quad (1.15)$$

Comme la transformation d'Euler est involutive, on en déduit l'expression de $f_n = f_n^{**}$

$$f_n = \sum_{k=0}^n (-1)^k \binom{n}{k} f_k^* = t_0 + n((2^d - 1)t_0 + t_1) + \sum_{k=2}^n \binom{n}{k} (-1)^k [k]_d! \sum_{j=2}^k \frac{t_j^* - t_{j-1}^*}{[j]_d!}.$$

Dans le cas particulier $t_n = \delta_{n,1}$, correspondant à $t_j^* - t_{j-1}^* = -1$, nous verrons que nous pouvons exprimer cette quantité sous la forme

$$f_n = n - \sum_{k=1}^n (-1)^k \binom{n}{k} \sum_{j=1}^n (-2)^{kd} (H_{\mathbf{s}_k^{(d)}}(j) - H_{\mathbf{t}_k^{(d)}}(j)),$$

avec $\mathbf{s}_k^{(d)} = (d, \dots, d) \in \mathbb{R}^k$ et $\mathbf{t}_k^{(d)} = (d, \dots, d, d-1) \in \mathbb{R}^k$.

1.2.4 Analyse en moyenne des paramètres additifs

Le point-clé de l'analyse réalisée dans [23] réside dans l'extrapolation de la quantité $\sum_{j=2}^n \frac{t_j^* - t_{j-1}^*}{[j]_d!}$, vue comme une fonction du paramètre n , en une certaine fonction $\xi(s)$, coïncidant aux entiers positifs avec cette quantité.

Pour cela, on peut étendre dans un premier temps le d -analogue de la fonction factorielle à une variable complexe s sous la forme

$$[s]_d! = \prod_{j=1}^{+\infty} \frac{[j+2]_d}{[j+s]_d},$$

ou encore à l'aide de la fonction Γ ,

$$\frac{[s]_d!}{[+\infty]_d!} = \prod_{\omega^d=2^d} \frac{\Gamma(s+1-\omega)}{\Gamma(s+1)}.$$

- Si $t_n = n$, comme $t_n^* = -\delta_{n,1}$, l'équation (1.14) se résume à $f_n^* = [n]_d!$ pour $n \geq 2$, $f_0^* = 0$ et $f_1^* = -1$, d'où on tire une expression explicite pour la série génératrice

$$f^*(t) = -t + t^2 {}_{d+1}F_d \left(\begin{matrix} 3 - \omega_1, \dots, 3 - \omega_d, 1 \\ 3, \dots, 3 \end{matrix} \middle| t \right).$$

Une analyse assez fine, via une représentation sous forme intégrale — une intégrale de Mellin inverse — permet de trouver le comportement de la fonction f^* au voisinage de $-\infty$, donc de celui de f au voisinage de 1 et, par analyse de singularité, permet de conclure que

$$f_n \sim \frac{2}{d} n \log(n) \quad (1.16)$$

- Si $t_n = \delta_{n,1}$, comme $t_n^* = -n$ et par conséquent

$$f_n = n - \sum_{k=2}^n \binom{n}{k} (-1)^k \sum_{j=2}^k \frac{[k]_d!}{[j]_d!}.$$

Le point-clé de l'analyse consiste alors à extrapoler $\nu_k = \sum_{j=2}^k \frac{1}{[j]_d!}$ en

$$\xi(s) = \frac{s-1}{[+\infty]_d!} + \sum_{j=2}^{+\infty} \left(\frac{1}{[j]_d!} - \frac{1}{[s+j-1]_d!} \right).$$

Il vient alors, toujours par des techniques de représentation sous forme intégrale, et de calcul de résidus,

$$f(z) \sim \frac{2}{d} \xi'(1) \frac{1}{(1-z)^2}, \text{ quand } z \rightarrow 1$$

$$f_n \sim \frac{2}{d} \xi'(1) n.$$

1.3 Étude de la distribution

Nous signalons enfin des résultats concernant les moments d'ordre supérieur des différents paramètres additifs étudiés dans la littérature, établissant dans le cas du nombre de feuilles l'apparition d'un phénomène de "changement de phase", c'est-à-dire la convergence vers une loi limite pour des dimensions d inférieures à un certain seuil, et l'inexistence de loi limite au-delà de cette dimension critique, ou au contraire, comme dans le cas de la longueur de cheminement, l'existence d'une loi limite, quelle que soit la dimension considérée.

1.3.1 Du nombre de feuilles

De la récurrence fondamentale (1.1), dans le cas où $t_n = \delta_{n,1}$, il ressort non seulement la conséquence (1.8) pour l'espérance du nombre de feuilles, mais également la même récurrence vérifiée par l'ensemble des moments de la variable $X_n = f(T)$, T arbre hyperquaternaire tel que $|T| = n$. Plus précisément, en notant $A_n = \mathbb{E}[X_n^j]$, et $B_n = \delta_{n,1}$, il vient

$$A_n = B_n + 2^d \sum_{k=0}^{n-1} \pi_{n,k,d} A_k. \quad (1.17)$$

On peut alors utiliser l'approche dite de "transfert des moments", qui permet d'obtenir des informations sur le comportement asymptotique de A_n connaissant le comportement de la suite-test B_n , ici $(\delta_{n,1})_{n \geq 0}$. Cette approche s'est révélée efficace dans de nombreux problèmes de nature récursive, comme par exemple dans l'étude de fonctionnelles additives sur les arbres m -aires de recherche [36].

Theorème 1.3.1 ([8]). *Soit $(A_n)_n$ définie par la récurrence (1.17) avec $A_0 = 0$ et $(B_n)_n$ une suite donnée, alors*

$$\exists K \in \mathbb{R} \quad A_n \sim K \cdot n \iff B_n = o(n) \text{ et } \sum_{n \geq 1} \left| \frac{B_n}{n^2} \right| < +\infty.$$

Ce résultat peut être affiné si on dispose d'une meilleure connaissance de $(B_n)_n$. En effet, en notant α et β les réels tels que $2 \exp\left(\frac{2i\pi}{d}\right) = \alpha + 1 + i\beta$,

Theorème 1.3.2 ([8]). *si $B_n = o(n^\alpha)$, alors il existe des constantes $K_1, K_2 \in \mathbb{R}$,*

$$A_n = Kn + K_1 n^{\alpha+i\beta} + K_2 n^{\alpha-i\beta} + o(n^\alpha + n^\varepsilon).$$

Grâce à ces éléments, on peut arriver à la conclusion suivante :

Proposition 1.3.3 ([8]). *Soit $M_{n,k} = \mathbb{E} \left[\left(X_n - \mu_d n - \frac{\mu_d}{2^d - 1} \right)^k \right]$.*

– Pour $1 \leq d \leq 8$, il existe une constante $\sigma_d \in \mathbb{R}$ telle que

$$\begin{cases} M_{n,2k} & \sim \frac{(2k)!}{2^k k!} \sigma_d^{2k} n^{2k} \\ M_{n,2k-1} & = o\left(n^{k-\frac{1}{2}}\right). \end{cases}$$

– Si $d \geq 9$,

$$M_{n,2k} \sim G_k(\beta \log n) n^{k\alpha},$$

avec $\alpha = 2 \cos\left(\frac{2\pi}{d}\right) - 1$ et $\beta = 2 \sin\left(\frac{2\pi}{d}\right)$, et G_k désignant une fonction périodique bornée.

Les moments (décalés) de la variable X_n convergent donc vers les moments d'une loi normale et, comme ces moments définissent une unique distribution, on peut en déduire :

Theorème 1.3.4 ([8]). *Si $1 \leq d \leq 8$, il existe des constantes μ_d et σ_d telles que*

$$\frac{X_n - \mu_d n}{\sigma_d \sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

où $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi, et $\mathcal{N}(0, 1)$ la loi normale centrée réduite.

Si $d \geq 9$, alors la suite $\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\text{Var}[X_n]}}$ ne converge pas vers une loi limite fixée.

1.3.2 D'autres paramètres

Nous continuons à désigner par X_n la variable aléatoire $f(T)$, T arbre hyperquaternaire tel que $|T| = n$. Chern et al. formulent dans [8] un résultat de changement de phase pour une suite-test $(t_n)_n$ assez générale :

Theorème 1.3.5 ([8]). *Pour une suite $(t_n)_{n \geq 0}$ telle que $t_n = \mathcal{O}(\sqrt{n}(\ln n)^{-1/2-\epsilon})$ et non identiquement égal à 1, alors pour $1 \leq d \leq 8$, il existe des constantes μ'_d et σ'_d telles que*

$$\frac{X_n - \mu'_d n}{\sigma'_d \sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Si $d \geq 9$, alors la suite $\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\text{Var}[X_n]}}$ ne converge pas vers une loi limite fixée.

Un cas particulier qui prolonge celui de l'étude du nombre de feuilles, et qui rentre dans le cadre de ce théorème, est le cas $t_n = \delta_{n,b}$. En effet, dans ce cas, quand $n \rightarrow +\infty$,

$$\mathbb{E}[X_n] \sim \lambda_b \cdot n,$$

où les constantes λ_b , qui dépendent de la dimension d , sont appelées "constantes universelles" et admettent une expression explicite pour $d = 2$ [41] :

$$\lambda_k = 3 - 18k + 6k(3k + 1)(\zeta(2) - 6H_2(k)).$$

Ces constantes permettent de calculer directement les constantes de linéarité pour des suites $(t_n)_n$ telles que $\mathbb{E}[X_n]$ soit asymptotiquement linéaire. Celles-ci feront l'objet d'une discussion en section 5.1.2.1.

Enfin, nous signalons un exemple de cas ne rentrant pas dans le cadre précédent d'une suite-test "faible", celui de la longueur de cheminement, où le phénomène de changement de phase n'apparaît pas.

Theorème 1.3.6 ([34, 62]). Si $t_n = n$, alors $\frac{X_n - \mathbb{E}[X_n]}{n}$ converge en loi, quand $n \rightarrow +\infty$, vers une variable non-gaussienne X_0 .

Remarque 1.3.7. La variable X_0 peut être caractérisée, en dimension $d = 1$ comme solution de l'équation en distribution [34]

$$X_0 \stackrel{\mathcal{L}}{=} UX_0 + (1 - U)X_0^* + 2U \log(U) + 2(1 - U) \log(1 - U) + 1,$$

où $\stackrel{\mathcal{L}}{=}$ désigne une égalité en loi, U une variable uniforme sur $[0, 1]$, $X_0^* \stackrel{\mathcal{L}}{=} X_0$, et les variables U , X_0 et X_0^* sont indépendantes.

En dimension supérieure, X_0 vérifie encore une équation en distribution [62]. De plus, l'espérance de la variable limite admet l'équivalent asymptotique, lorsque $n \rightarrow +\infty$, donnée en (1.16).

Chapitre 2

Quelques éléments d'algèbre non commutative

Sommaire

2.1	Combinatoire des mots	24
2.2	Mots de Lyndon et théorème de Radford	26
2.3	Algèbres de Lie	27
2.3.1	Généralités	27
2.3.2	Formes crochétées et bases duales	28
2.4	Algèbres de Hopf	29

Nous venons de présenter une structure de données, l'arbre hyperquaternaire aléatoire de points, qui nous a conduit à la définition des sommes harmoniques multiples, intervenant dans le calcul de la taille des sous-arbres issus de la racine, et plus généralement dans la récurrence fondamentale régissant les paramètres additifs. Par ailleurs, nous avons pu constater que ces paramètres avaient été précédemment étudiés via des techniques d'analyse de singularités.

L'approche que nous souhaitons proposer maintenant pour l'étude des différents objets introduits au chapitre précédent est une approche *formelle*, fondée sur l'algèbre non commutative et les séries génératrices non commutatives. Nous allons préciser progressivement, dans les deux prochains chapitres, la manière dont nous entendons nous servir de cet outillage. Le présent chapitre est composé pour l'essentiel de théorie sur la combinatoire des mots - qui nous servira par la suite pour le *codage* des sommes harmoniques et de leurs séries génératrices - et sur les propriétés des *produits de mélanges*, produits étroitement liés à des *relations de mélange* vérifiés par les sommes harmoniques ou par leurs séries génératrices.

2.1 Combinatoire des mots

Un *alphabet* est un ensemble de lettres, fini $X = \{x_1, \dots, x_k\}$ ou infini $Y = \{y_i, i \geq 0\}$. Un *mot* est une suite finie $w = x_{i_1} \dots x_{i_k}$ de lettres. Le *mot vide*, noté ϵ est le mot ne contenant aucune lettre. La longueur d'un mot $w = x_{i_1} \dots x_{i_k}$, noté $|w|$ est la longueur de la suite de lettres constituant w , i.e. l'entier k . Le nombre d'occurrences de la lettre x_i dans le mot w est noté $|w|_{x_i}$. Si chaque lettre est associée à un entier nommé *poids*, nous appellerons *poids* d'un mot la somme des poids respectifs des lettres qui le composent. L'ensemble des mots sur un alphabet X est noté X^* . On note de plus X^+ l'ensemble des mots non vides.

Exemple 2.1.1. Soit $Y = \{y_i, i \geq 1\}$, et $w = y_1^2 y_3 y_2 y_1 \in Y^*$, le poids de chaque lettre étant défini par son indice, w a pour poids 8. De plus, $|w| = 5$, et $|w|_{y_1} = 3$.

L'ensemble des polynômes sur l'alphabet X est noté $\mathbb{C}\langle X \rangle$. Une série formelle non commutative sur X à coefficients dans \mathbb{C} est une application S de X^* dans \mathbb{C} :

$$S : w \in X^* \mapsto \langle S|w \rangle \in \mathbb{C},$$

$\langle S|w \rangle$ désignant le coefficient de w dans S . Par abus de notation, on écrira simplement $S = \sum_{w \in X^*} \langle S|w \rangle w$. L'ensemble des séries formelles sur X à coefficients dans \mathbb{C} est noté $\mathbb{C}\langle\langle X \rangle\rangle$.

Le produit de concaténation, associatif mais non commutatif, de $u = x_{i_1} \dots x_{i_k}$ et $v = x_{j_1} \dots x_{j_q}$ est le mot $w = uv = x_{i_1} \dots x_{i_k} x_{j_1} \dots x_{j_q}$. Un mot u est un facteur de $v \in X^*$, s'il existe deux mots $(w_1, w_2) \in (X^*)^2$ tel que $v = w_1 u w_2$. Si w_1 (resp. w_2) est le mot vide, alors u est un facteur gauche (resp. droit) de v .

Définition 2.1.2. Soit $a, b \in X$ et $u, v \in X^*$. Le shuffle (ou produit de battage) de $u = au'$ et $v = bv'$ est le polynôme de $\mathbb{C}\langle X \rangle$ défini récursivement par

$$\epsilon \sqcup u = u \sqcup \epsilon = u \quad \text{et} \quad u \sqcup v = a(u' \sqcup v) + b(u \sqcup v'). \quad (2.1)$$

Définition 2.1.3. Soit $u, v \in Y^*$. Le stuffle (ou produit de battage contractant) de $u = y_i u'$ et $v = y_j v'$ est le polynôme de $\mathbb{C}\langle Y \rangle$ défini par

$$\epsilon \sqcup u = u \sqcup \epsilon = u \quad \text{et} \quad u \sqcup v = y_i(u' \sqcup v) + y_j(u \sqcup v') + y_{i+j}(u' \sqcup v'). \quad (2.2)$$

De manière similaire, le minus-stuffle de u et v est le polynôme de $\mathbb{C}\langle Y \rangle$ défini par

$$\epsilon \sqcup u = u \sqcup \epsilon = u \quad \text{et} \quad u \sqcup v = y_i(u' \sqcup v) + y_j(u \sqcup v') - y_{i+j}(u' \sqcup v'). \quad (2.3)$$

Exemple 2.1.4.

$$\begin{aligned} y_1 \sqcup y_2 &= y_1 y_2 + y_2 y_1, \\ y_1 \sqcup y_2 &= y_1 y_2 + y_2 y_1 + y_3 \\ \text{et} \quad y_1 \sqcup y_2 &= y_1 y_2 + y_2 y_1 - y_3. \end{aligned}$$

Hoffman définit dans [31] un produit dit de *quasi-mélange* \star , ayant pour élément neutre ϵ , et tel que pour toutes lettres y_i et y_j de Y et tous mots u' et v' de Y^*

$$y_i u' \star y_j v' = y_i(u' \star v) + y_j(u \star v') + [y_i, y_j](u' \star v'),$$

où l'application $[\cdot, \cdot] : \bar{Y}^2 \rightarrow \bar{Y}$ (avec $\bar{Y} = Y \cup \{0\}$) vérifie

- $\forall a \in Y, [a, 0] = 0,$
- $\forall (a, b) \in \bar{Y}^2, [a, b] = [b, a]$
- $\forall (a, b, c) \in \bar{Y}^3, [[a, b], c] = [a, [b, c]]$
- Ou bien $[a, b] = 0$ ou bien pour tous a et b , $[a, b]$ a pour poids la somme des poids de a et b .

Les produits \sqcup et \sqcup peuvent donc être regardés comme des cas particuliers de produits de *quasi-mélange*. Afin de couvrir les trois produits de mélange évoqués dans ce mémoire, nous tordons le produit de *quasi-mélange* en un produit que nous continuons à noter \star , d'élément neutre ϵ , et tel que

$$y_i u' \star y_j v' = y_i(u' \star v) + y_j(u \star v') + \lambda [y_i, y_j](u' \star v'), \quad \text{avec } \lambda \in \{-1, 1\}. \quad (2.4)$$

Proposition 2.1.5. *Le produit \star est commutatif et associatif.*

Démonstration. Pour prouver que $w_1 \star w_2 = w_2 \star w_1$, nous procédons par récurrence sur $|w_1| + |w_2|$. L'hypothèse de récurrence est prouvée par la neutralité de ϵ lorsque $|w_1| + |w_2| \leq 1$ et l'hérédité découle de (2.4).

De la même manière, nous prouvons que $w_1 \star (w_2 \star w_3) = (w_1 \star w_2) \star w_3$ par récurrence sur $|w_1| + |w_2| + |w_3|$. Encore une fois, l'initialisation est prouvée par la neutralité de ϵ lorsque $|w_1| + |w_2| + |w_3| \leq 1$. Puis, le calcul de $y_i w_1 \star (y_j w_2 \star y_k w_3)$ donne

$$\begin{aligned} & y_i(w_1 \star y_j(w_2 \star y_k w_3)) + y_j(y_i w_1 \star (w_2 \star y_k w_3)) + \lambda [y_i, y_j](w_1 \star (w_2 \star y_k w_3)) \\ & + y_i(w_1 \star y_k(y_j w_2 \star w_3)) + y_k(y_i w_1 \star (y_j w_2 \star w_3)) + \lambda [y_i, y_k](w_1 \star (y_j w_2 \star w_3)) \\ & + \lambda y_i(w_1 \star [y_j, y_k](w_2 \star w_3)) + \lambda [y_j, y_k](y_i w_1 \star (w_2 \star w_3)) + \lambda^2 [y_i, [y_j, y_k]](w_1 \star (w_2 \star w_3)) \end{aligned}$$

D'autre part, le calcul de $(y_i w_1 \star y_j w_2) \star y_k w_3$ donne

$$\begin{aligned} & y_i((w_1 \star y_j w_2) \star y_k w_3) + y_k(y_i(w_1 \star y_j w_2) \star w_3) + \lambda [y_i, y_k]((w_1 \star y_j w_2) \star w_3) \\ & + y_j((y_i w_1 \star w_2) \star y_k w_3) + y_k(y_j(y_i w_1 \star w_2) \star w_3) + \lambda [y_j, y_k]((y_i w_1 \star w_2) \star w_3) \\ & + \lambda [y_i, y_j]((w_1 \star w_2) \star y_k w_3) + \lambda y_k([y_i, y_j](w_1 \star w_2) \star w_3) + \lambda^2 [[y_i, y_j], y_k]((w_1 \star w_2) \star w_3) \end{aligned}$$

En soustrayant les deux égalités et en utilisant l'hypothèse de récurrence, on aboutit à :

$$\begin{aligned} & y_i(w_1 \star y_j(w_2 \star y_k w_3)) + y_i(w_1 \star y_k(y_j w_2 \star w_3)) + y_k(y_i w_1 \star (y_j w_2 \star w_3)) \\ & + \lambda y_i(w_1 \star [y_j, y_k](w_2 \star w_3)) - y_i((w_1 \star y_j w_2) \star y_k w_3) - y_k(y_i(w_1 \star y_j w_2) \star w_3) \\ & - y_k(y_j(y_i w_1 \star w_2) \star w_3) - \lambda y_k([y_i, y_j](w_1 \star w_2) \star w_3), \end{aligned}$$

qui peut encore être simplifié en utilisant (2.4) en

$$\begin{aligned} & y_i(w_1 \star (y_j w_2 \star y_k w_3)) + y_k(y_i w_1 \star (y_j w_2 \star w_3)) \\ & - y_i((w_1 \star y_j w_2) \star y_k w_3) - y_k((y_i w_1 \star y_j w_2) \star w_3), \end{aligned}$$

expression réduite à zéro par l'hypothèse de récurrence. \square

2.2 Mots de Lyndon et théorème de Radford

Par définition, un *mot de Lyndon* est un mot $l \in X^+$ strictement inférieur à chacun de ses facteurs droits stricts [66] (pour l'ordre lexicographique) i.e. pour tous $u, v \in X^+$, $l = uv \Rightarrow l < v$. L'ensemble des mots de Lyndon sur X est noté $\mathcal{Lyn}X$.

Exemple 2.2.1. Pour $X = \{x_0, x_1\}$ muni de l'ordre $x_0 < x_1$ les mots de Lyndon de longueur inférieure ou égale à 5 sont (donnés dans l'ordre lexicographique croissant)

$$\{x_0, x_0^4x_1, x_0^3x_1, x_0^3x_1^2, x_0^2x_1, x_0^2x_1x_0x_1, x_0^2x_1^2, x_0^2x_1^3, x_0x_1, x_0x_1x_0x_1^2, x_0x_1^2, x_0x_1^3, x_0x_1^4, x_1\}.$$

Pour $Y = \{y_i, i \geq 1\}$, muni de l'ordre $y_i < y_j$ si $i > j$, les mots de Lyndon de poids inférieur ou égal à 5 sont

$$\{y_5, y_4, y_4y_1, y_3, y_3y_2, y_3y_1, y_3y_1^2, y_2, y_2^2y_1, y_2y_1, y_2y_1^2, y_2y_1^3, y_1\}.$$

Remarque 2.2.2. Nous verrons plus tard que l'ordre adopté sur l'alphabet Y n'est pas arbitraire, mais permet d'identifier bijectivement les mots de Lyndon sur X différents de x_0 et ceux sur Y via la projection

$$\pi_Y(x_0^{s_1-1}x_1 \dots x_0^{s_r-1}x_1) = y_{s_1} \dots y_{s_r}.$$

Théorème 2.2.3 (Radford, [65]). Soit $C_1 = \mathbb{C} \oplus (x_0\mathbb{C}\langle X \rangle x_1)$ l'ensemble des polynômes dits "convergents" sur X . Alors,

$$(\mathbb{C}\langle X \rangle, \sqcup) \simeq \mathbb{C}[\mathcal{Lyn}X] = C_1[x_0, x_1].$$

Théorème 2.2.4 ([49]). Soit $C_2 = \mathbb{C} \oplus (\mathbb{C}\langle Y \rangle \setminus y_1\mathbb{C}\langle Y \rangle) \simeq C_1$, l'ensemble des polynômes dits "convergents" sur Y . Alors,

$$(\mathbb{C}\langle Y \rangle, \sqcup) \simeq (\mathbb{C}\langle Y \rangle, \sqcup) \simeq \mathbb{C}[\mathcal{Lyn}Y] = C_2[y_1].$$

Exemple 2.2.5.

$$\begin{aligned} x_1x_0x_1x_0x_1 + 2x_0x_1^2x_0x_1 + 2x_0x_1x_0x_1^2 &= \frac{1}{2}x_0x_1 \sqcup x_0x_1 \sqcup x_1 - 2x_0^2x_1^2 \sqcup x_1 \in \mathbb{C}[\mathcal{Lyn}X] \\ &= x_0x_1x_0x_1 \sqcup x_1 \in C_1[x_0, x_1] \\ y_2y_4y_1 + y_2y_1y_4 + y_1y_2y_4 + y_2y_5 + y_3y_4 &= y_4 \sqcup y_2 \sqcup y_1 - y_4y_2 \sqcup y_1 - y_6 \sqcup y_1 \in \mathbb{C}[\mathcal{Lyn}Y] \\ &= y_2y_4 \sqcup y_1 \in C_2[y_1] \end{aligned}$$

Remarque 2.2.6. L'isomorphisme signalé au théorème 2.2.4 entre les espaces vectoriels C_1 et C_2 n'est autre que l'application π_Y évoquée à la remarque 2.2.2.

Signalons enfin deux décompositions fondamentales relatives aux mots de Lyndon ([47]).

Proposition 2.2.7. Soit X un alphabet muni d'un ordre, et soit $l \in \mathcal{Lyn}X$, un mot de longueur $|l| \geq 2$, alors l admet une factorisation unique, dite factorisation standard, sous la forme $l = uv$, avec $u, v \in \mathcal{Lyn}X$, et le mot v étant le plus long possible.

Exemple 2.2.8. Avec l'ordre $x_0 < x_1$ sur l'alphabet $X = \{x_0, x_1\}$, le mot de Lyndon $x_0x_1x_0x_1x_0x_1^2$ se décompose sous la forme d'un produit de deux mots de Lyndon de deux façons, $x_0x_1x_0x_1 \cdot x_0x_1^2$ et $x_0x_1 \cdot x_0x_1x_0x_1^2$, mais, avec la condition sur la longueur de v , seule cette dernière expression correspond à la factorisation standard.

Proposition 2.2.9. Tout mot $w \in X^+$ peut s'exprimer, de manière unique, comme un produit décroissant de mots de Lyndon,

$$w = l_1^{\alpha_1} \dots l_k^{\alpha_k}, \text{ avec } l_i \in \mathcal{Lyn}X, \text{ et } l_1 > \dots > l_k.$$

Par exemple, $x_1^2x_0x_1x_0x_1^2x_0^2x_1x_0^3 = (x_1)^2 \cdot x_0x_1x_0x_1^2 \cdot x_0^2x_1 \cdot (x_0)^3$, avec $x_1 > x_0x_1x_0x_1^2 > x_0^2x_1 > x_0$.

2.3 Algèbres de Lie

2.3.1 Généralités

L'ensemble des monômes de Lie est défini par récurrence : les lettres de l'alphabet X sont des monômes de Lie et le crochet de Lie $[a, b] = ab - ba$ de deux monômes de Lie a et b est un monôme de Lie. K désignant un corps fixé, un polynôme de Lie est une combinaison K -linéaire de monômes de Lie. L'ensemble des polynômes de Lie forme alors une algèbre de Lie, notée $Lie\langle X \rangle$, l'algèbre de Lie libre sur X . Une série est une série de Lie si et seulement si toutes ses composantes homogènes sont des polynômes de Lie. L'ensemble des séries de Lie est noté $Lie\langle\langle X \rangle\rangle$.

Exemple 2.3.1. Sur $X = \{x_0, x_1\}$, les éléments

$$\begin{aligned} [x_0, x_1] &= x_0x_1 - x_1x_0, \\ [x_1, [x_0, x_1]] &= [x_1, x_0x_1 - x_1x_0] = 2x_1x_0x_1 - x_1^2x_0 - x_0x_1^2, \end{aligned}$$

sont des monômes de Lie. Le polynôme $[x_0, [x_0, x_1]] + 3[[x_1, x_1], x_0]$ est un polynôme homogène de Lie de degré 3.

Définition 2.3.2. Soit Δ le morphisme d'algèbre de $K\langle\langle X \rangle\rangle$ (muni du produit de concaténation) dans $K\langle\langle X \rangle\rangle \otimes K\langle\langle X \rangle\rangle$ défini pour une lettre x par $\Delta(x) = x \otimes 1 + 1 \otimes x$.

Une série $S \in K\langle\langle X \rangle\rangle$ est dite primitive si elle vérifie $\Delta(S) = S \otimes 1 + 1 \otimes S$.

Une série $S \in K\langle\langle X \rangle\rangle$ est dite group-like si elle vérifie $\Delta(S) = S \otimes S$.

Une série $S \in K\langle\langle X \rangle\rangle$ vérifie le critère de Friedrichs si pour tout couple $(w_1, w_2) \in (X^*)^2$, $\langle S|w_1 \rangle \langle S|w_2 \rangle = \langle S|w_1 \sqcup w_2 \rangle$

Theorem 2.3.3 ([66]). *Soit $S \in K\langle\langle X \rangle\rangle$.*

$$\begin{aligned} S \in \text{Lie}\langle\langle X \rangle\rangle &\Leftrightarrow S \text{ est primitive} \\ &\Leftrightarrow e^S \text{ est group-like} \\ &\Leftrightarrow e^S \text{ vérifie le critère de Friedrichs.} \end{aligned}$$

Une série $S \in K\langle\langle X \rangle\rangle$ est une exponentielle de Lie si et seulement s'il existe une série de Lie L telle que $S = e^L$.

2.3.2 Formes crochetées et bases duales

Dans le cas de l'algèbre de Lie libre $\text{Lie}\langle X \rangle$, nous allons maintenant nous fixer une base, la base de Lyndon, un choix caractérisé par une construction explicite, et donc une implémentation efficace pour les algorithmes que nous aurons à mener au sein de cette algèbre.

Theorem 2.3.4 ([51]). *Soit $\{\check{S}_i\}_{i \geq 1}$ une base de $\text{Lie}\langle X \rangle$ totalement ordonnée et $\mathcal{B} = \{\check{S}_{i_1}^{j_1} \dots \check{S}_{i_n}^{j_n}, j_k \geq 0 \text{ et } \check{S}_{i_1} > \dots > \check{S}_{i_n}\}$. Alors \mathcal{B} est une base de $K\langle X \rangle$, appelée base de Poincaré-Birkhoff-Witt (PBW), associée à $\{\check{S}_i\}_{i \geq 1}$.*

Soit $\mathcal{B}^* = \{\mathcal{S}_B, B \in \mathcal{B}\}$ la base duale de \mathcal{B} , i.e. vérifiant pour tous B_1, B_2 éléments de la base \mathcal{B} , $\langle \mathcal{S}_{B_1} | B_2 \rangle = \delta_{B_1, B_2}$, où $\delta_{B_1, B_2} = 1$ si $B_1 = B_2$, et 0 sinon.

Explicitons plus avant la construction de la base de PBW pour un choix particulier des $\{\check{S}_i\}_{i \geq 1}$: la forme crochetée \check{S}_l d'un mot de Lyndon l (de factorisation standard $l = uv$ si $|l| \geq 2$), est définie récursivement par

$$\begin{cases} \check{S}_l = [\check{S}_u, \check{S}_v] \\ \check{S}_x = x \text{ pour toute lettre } x \in X, \end{cases}$$

le crochet $[\check{S}_u, \check{S}_v]$ désignant le crochet de Lie. La base de Lyndon, qui constitue une base de $\text{Lie}\langle X \rangle$, est l'ensemble des formes crochetées des mots de Lyndon.

Exemple 2.3.5. *Si $X = \{x_0, x_1\}$, le mot de Lyndon $x_0^3 x_1$ a pour forme crochetée*

$$\begin{aligned} \check{S}_{x_0^3 x_1} &= [\check{S}_{x_0}, \check{S}_{x_0^2 x_1}] \\ &= [x_0, [x_0, \check{S}_{x_0 x_1}]] \\ &= [x_0, [x_0, [x_0, x_1]]] \end{aligned}$$

La base de Poincaré-Birkhoff-Witt $\mathcal{B} = \{\check{S}_w; w \in X^*\}$ et sa base duale $\mathcal{B}^* = \{\mathcal{S}_w; w \in X^*\}$ s'obtiennent alors en posant [66]

$$\begin{cases} \check{S}_w = \check{S}_{l_1}^{\alpha_1} \check{S}_{l_2}^{\alpha_2} \dots \check{S}_{l_k}^{\alpha_k}, \\ \mathcal{S}_w = \frac{\mathcal{S}_{l_1}^{\sqcup \alpha_1} \sqcup \dots \sqcup \mathcal{S}_{l_k}^{\sqcup \alpha_k}}{\alpha_1! \alpha_2! \dots \alpha_k!}, \end{cases}$$

l	\mathcal{S}_l	\mathcal{S}_l
x_0	x_0	x_0
x_1	x_1	x_1
x_0x_1	$[x_0, x_1]$	x_0x_1
$x_0^2x_1$	$[x_0, [x_0, x_1]]$	$x_0^2x_1$
$x_0x_1^2$	$[[x_0, x_1], x_1]$	$x_0x_1^2$
$x_0^3x_1$	$[x_0, [x_0, [x_0, x_1]]]$	$x_0^3x_1$
\vdots	\vdots	\vdots
$x_0^3x_1^3$	$[x_0, [x_0, [[[x_0, x_1], x_1], x_1]]]$	$x_0^3x_1^3$
$x_0^2x_1x_0x_1^2$	$[x_0, [[x_0, x_1], [[x_0, x_1], x_1]]]$	$3x_0^3x_1^3 + x_0^2x_1x_0x_1^2$
$x_0^2x_1^2x_0x_1$	$[[x_0, [[x_0, x_1], x_1]], [x_0, x_1]]$	$6x_0^3x_1^3 + 3x_0^2x_1x_0x_1^2 + x_0^2x_1^2x_0x_1$
$x_0^2x_1^4$	$[x_0, [[[[x_0, x_1], x_1], x_1], x_1]]$	$x_0^2x_1^4$
$x_0x_1x_0x_1^3$	$[[x_0, x_1], [[[x_0, x_1], x_1], x_1]]$	$4x_0^2x_1^4 + x_0x_1x_0x_1^3$
$x_0x_1^5$	$[[[[[x_0, x_1], x_1], x_1], x_1], x_1]]$	$x_0x_1^5$

TAB. 2.1 – Mots de Lyndon, formes crochetées et base duale

où les l_i sont les facteurs intervenants dans la décomposition de la proposition 2.2.9. Enfin la base duale est complètement déterminée en précisant que, pour $l \in \text{Lyn}X$,

$$\mathcal{S}_l = x\mathcal{S}_w, \quad \text{où } l = xw, x \in X, w \in X^*.$$

Lemme 2.3.6. Pour tout $w \in x_0X^*x_1$, $\mathcal{S}_w \in x_0\mathbb{Z}\langle X \rangle x_1$.

2.4 Algèbres de Hopf

Le produit de mélange \star permet de définir une application linéaire,

$$st : \mathbb{C}\langle Y \rangle \otimes \mathbb{C}\langle Y \rangle \rightarrow \mathbb{C}\langle Y \rangle,$$

définie pour w_1 et $w_2 \in Y^*$ par

$$st(w_1 \otimes w_2) = w_1 \star w_2,$$

et étendue par linéarité aux polynômes. Alors, l'application linéaire

$$\begin{aligned} \mathbf{1} : \mathbb{C} &\rightarrow \mathbb{C}\langle Y \rangle \\ k &\mapsto \mathbf{1}(k) = k \end{aligned}$$

constitue un élément *unité*. Donc $(\mathbb{C}\langle Y \rangle, st, \mathbf{1})$ constitue une \mathbb{C} -algèbre associative et graduée. Cette algèbre est connue comme *algèbre de mélange*.

Nous définissons un coproduit, pour le produit \star , $\Delta : \mathbb{C}\langle Y \rangle \rightarrow \mathbb{C}\langle Y \rangle \otimes \mathbb{C}\langle Y \rangle$ par l'opération classique de "déconcaténation"

$$\Delta(w) = \sum_{uv=w} u \otimes v.$$

$e : S \in \mathbb{C}\langle Y \rangle \mapsto e(S) = \langle S | \epsilon \rangle \in \mathbb{C}$ apparaît comme un élément unité pour le coproduit Δ , et donc $(\mathbb{C}\langle Y \rangle, \Delta, e)$ est une cogèbre (non cocommutative).

Pour un mot $w = y_{s_1} \dots y_{s_r}$, nous pouvons définir l'action d'une composition $I = (i_1, \dots, i_l)$ de l'entier r (i.e. une suite finie d'entiers ayant pour somme r) par

$$I[w] = \lambda^l [y_{s_1}, \dots, y_{s_{i_1}}] \dots [y_{s_{(i_1+\dots+i_{l-1}+1)}}, \dots, y_{s_r}].$$

Exemple 2.4.1. Soit $w = y_1^5 y_6 y_2$ et $I = (1, 2, 4)$ une composition de 7 alors

$$I[w] = \lambda^3 [y_1][y_1, y_1][y_1, y_1, y_6, y_2].$$

Si le produit \star considéré est le produit stuffle \sqcup , alors $I[w] = y_1 y_2 y_{10}$.

Dans le cas du produit \sqcup , nous aurons $I[w] = -y_1 y_2 y_{10}$.

Alors, la bigèbre $(\mathbb{C}\langle Y \rangle, st, \mathbb{1}, \Delta, e)$ se révèle en fait être une algèbre de Hopf, appelée *algèbre de factorisation* dont l'antipode a est donnée par [31]

$$a(w) = (-1)^r \sum_{I \in \text{Comp}(r)} I[y_{s_r} \dots y_{s_1}], \quad (2.5)$$

$\text{Comp}(r)$ désignant l'ensemble des compositions de r .

Remarque 2.4.2. Dans le cas du produit \sqcup , l'antipode se résume à

$$a_{\sqcup}(w) = (-1)^{|w|} \widetilde{w},$$

avec $\widetilde{y_{s_1} \dots y_{s_r}} = y_{s_r} \dots y_{s_1}$ (fonction miroir).

Si le coproduit, pour tous les produits \star , est identique, en partant maintenant du produit de concaténation sur l'alphabet X , le coproduit singularise un quasi-mélange particulier, le produit \sqcup . En effet, l'algèbre de Hopf dite *de décomposition* est la bigèbre $(\mathbb{C}\langle X \rangle, \cdot, \mathbb{1}, \Delta', e)$ munie de l'antipode a_{\sqcup} précédemment définie, et où le coproduit Δ' est défini par transposition du produit \sqcup :

$$\forall (S, P_1, P_2) \in (\mathbb{C}\langle X \rangle)^3 \quad \langle \Delta'(S) | P_1 \otimes P_2 \rangle = \langle S | P_1 \sqcup P_2 \rangle.$$

Cette algèbre est alors non commutative, mais cocommutative. En fait, en notant sh l'application

$$sh : w_1 \otimes w_2 \in \mathbb{C}\langle X \rangle \otimes \mathbb{C}\langle X \rangle \mapsto w_1 \sqcup w_2 \in \mathbb{C}\langle X \rangle,$$

les algèbres $(\mathbb{C}\langle X \rangle, \cdot, 1, \Delta', e, a_{\sqcup})$ et $(\mathbb{C}\langle X \rangle, sh, 1, \Delta, e, a_{\sqcup})$ sont des algèbres de Hopf duales l'une de l'autre. Et, de plus, d'après le théorème de Cartier-Milner-Moore, $\mathbb{C}\langle X \rangle$ est isomorphe, comme algèbre de Hopf cocommutative graduée, à l'algèbre enveloppante de l'algèbre de Lie $Lie\langle X \rangle$. Par conséquent,

Theorème 2.4.3 (Schützenberger,[66]). *La série diagonale, considérée comme une série sur X^* , à coefficients dans l'algèbre de mélange $(\mathbb{C}\langle X \rangle, \sqcup)$, peut être factorisée en un produit infini d'exponentielles de Lie, indexées par les mots de Lyndon, pris dans l'ordre décroissant*

$$\sum_{w \in X^*} w \otimes w = \prod_{l \in \mathcal{L}yn X}^{\downarrow} e^{S_l \otimes \tilde{S}_l}$$

Soit S une exponentielle de Lie, en considérant l'image de $\sum_{w \in X^*} w \otimes w$, par l'application $w \otimes w \mapsto \langle S | w \rangle w$, nous obtenons

Theorème 2.4.4.

$$S = \prod_{l \in \mathcal{L}yn X}^{\downarrow} e^{\langle S | S_l \rangle \tilde{S}_l}$$

Chapitre 3

Fonctions polylogarithmes

Sommaire

3.1	Codage des polylogarithmes	33
3.2	Relation de mélange	34
3.3	Propriétés de la série génératrice	35

Nous pouvons maintenant commencer à appliquer les résultats théoriques du chapitre précédent à l'algèbre des *polylogarithmes*, fonctions stables par intégration par rapport aux formes différentielles $\frac{dz}{z}$ et $\frac{dz}{1-z}$, des 1-formes déjà rencontrées au chapitre 1, section 1.2.2. Les fonctions de cette algèbre vérifient en outre l'équation de Drinfel'd [18], d'où l'on peut déduire que leur série génératrice forme une exponentielle de Lie. De là, découlent la relation de mélange vérifiée par les polylogarithmes, ainsi que la factorisation par les mots de Lyndon.

3.1 Codage des polylogarithmes

Considérons les deux formes différentielles suivantes $\omega_0(z) = \frac{dz}{z}$ et $\omega_1(z) = \frac{dz}{1-z}$. Le polylogarithme $\text{Li}_s(z)$ est défini pour un multi-indice $s = (s_1, \dots, s_r)$ et pour un complexe z tel que $|z| < 1$ par la formule

$$\text{Li}_s(z) = \sum_{n_1 > \dots > n_r} \frac{z^{n_1}}{n_1^{s_1} \dots n_r^{s_r}}, \quad (3.1)$$

et correspond à l'intégrale itérée par rapport à ω_0 et ω_1 , le long du chemin d'intégration $0 \rightsquigarrow z$,

$$\text{Li}_s = \int_{0 \rightsquigarrow z} \omega_0^{s_1-1} \omega_1 \dots \omega_0^{s_r-1} \omega_1. \quad (3.2)$$

Soit $X = \{x_0, x_1\}$. Incités par l'égalité (3.2), nous identifierons dorénavant le multi-indice $\mathbf{s} = (s_1, \dots, s_r)$ avec le mot $w = x_0^{s_1-1}x_1 \dots x_0^{s_r-1}x_1$ élément de X^*x_1 . Nous obtenons ainsi un isomorphisme de concaténation de la \mathbb{C} -algèbre des multi-indices vers la sous-algèbre $\mathbb{C}\langle X \rangle_{x_1} \subset \mathbb{C}\langle X \rangle$. Ainsi, le polylogarithme $\text{Li}_{\mathbf{s}}(z)$ défini par la formule (3.1) peut également être indexé par $w \in X^*x_1$. Par ailleurs, la forme d'intégrale itérée permet d'affirmer pour des mots $w \in X^*x_1$,

$$\text{Li}_{x_0w}(z) = \int_0^z \text{Li}_w(t) \frac{dt}{t} \quad (3.3)$$

$$\text{Li}_{x_1w}(z) = \int_0^z \text{Li}_w(t) \frac{dt}{1-t}. \quad (3.4)$$

Pour étendre la définition des polylogarithmes à X^* , nous posons alors, pour tout $n \in \mathbb{N}$, $\text{Li}_{x_0^n}(z) = \frac{\log^n(z)}{n!}$, et utilisons les équations (3.3) et (3.4).

Exemple 3.1.1.

$$\text{Li}_{x_1x_0^2}(z) = \int_0^z \frac{\log^2(t)}{2!} \frac{dt}{1-t}$$

Enfin, l'extension aux polynômes de $\mathbb{C}\langle X \rangle$ se fait par linéarité.

Proposition 3.1.2. Soit $w = x_0^{s_1-1}x_1 \dots x_0^{s_r-1}x_1 \in X^*x_1$. Alors, la limite $\lim_{z \rightarrow 1} \text{Li}_w(z)$ existe si et seulement si $s_1 > 1$. Dans ce cas,

$$\lim_{z \rightarrow 1} \text{Li}_w(z) = \sum_{n_1 > \dots > n_r} \frac{1}{n_1^{s_1} \dots n_r^{s_r}} \stackrel{\text{def}}{=} \zeta(w),$$

limite connue sous le nom de polyzêta (ou Multiple Zeta Value [70]).

Les polyzêtas sont définies seulement pour les mots convergents et étendus linéairement à $C_1 = \mathbb{C} \oplus x_0\mathbb{C}\langle X \rangle_{x_1}$.

3.2 Relation de mélange

Soit $\mathcal{C} = \mathbb{C} \left[z, \frac{1}{z}, \frac{1}{1-z} \right]$ et $\text{LI}_{\mathcal{C}} = \mathcal{C}[\{\text{Li}_w\}_{w \in X^*}]$. La série génératrice des polylogarithmes, $L = \sum_{w \in X^*} \text{Li}_w w$, vérifie l'équation différentielle de Drinfel'd [18]

$$dL = (x_0\omega_0 + x_1\omega_1)L \quad \text{avec la condition } L(\varepsilon) = e^{x_0 \log \varepsilon} + \mathcal{O}(\sqrt{\varepsilon}) \quad \text{pour } \varepsilon \rightarrow 0^+ \quad (3.5)$$

Ceci permet de prouver ([60]) que L peut s'écrire comme l'exponentielle d'une série de Lie, ce qui est équivalent ([66]) à

Proposition 3.2.1 ([59]). *Pour $u, v \in X^*$, $\text{Li}_{u \sqcup v} = \text{Li}_u \text{Li}_v$. Donc, pour u, v appartenant à $x_0 X^* x_1$, $\zeta(u \sqcup v) = \zeta(u)\zeta(v)$.*

Exemple 3.2.2. *Prolongeons l'exemple 2.2.5,*

$$\begin{aligned} \text{Li}_{x_1 x_0 x_1 x_0 x_1} + 2 \text{Li}_{x_0 x_1^2 x_0 x_1} + 2 \text{Li}_{x_0 x_1 x_0 x_1^2} &= \frac{1}{2} \text{Li}_{x_0 x_1} \text{Li}_{x_0 x_1} \text{Li}_{x_1} - 2 \text{Li}_{x_0^2 x_1^2} \text{Li}_{x_1} \\ &= \text{Li}_{x_0 x_1 x_0 x_1} \text{Li}_{x_1}, \end{aligned}$$

ou, de manière équivalente

$$\begin{aligned} \text{Li}_{1,2,2} + 2 \text{Li}_{2,1,2} + 2 \text{Li}_{2,2,1} &= \frac{1}{2} \text{Li}_2^2 \text{Li}_1 - 2 \text{Li}_{3,1} \text{Li}_1 \\ &= \text{Li}_{2,2} \text{Li}_1. \end{aligned}$$

La proposition 3.2.1 nous présente donc l'application $L : w \mapsto \text{Li}_w$, comme un morphisme. En réalité, on dispose du résultat suivant,

Theorème 3.2.3 ([60]). *L'application $L : w \mapsto \text{Li}_w$ est un isomorphisme de $(\mathbb{C}\langle X \rangle, \sqcup)$ sur $(\mathbb{C}[\{\text{Li}_w\}_{w \in X^*}], \cdot)$.*

Démonstration. Démontrons par récurrence sur k que les fonctions $\{\text{Li}_w, w \in X^*, |w| \leq k\}$ sont \mathbb{C} -linéairement indépendantes. Cela étant acquis si $k = 0$, supposons le résultat vrai pour $k = K - 1$. Alors

$$\sum_{w \in X^* / |w| \leq K} \lambda_w \text{Li}_w = 0 \iff \lambda_\epsilon + \sum_{x \in X} \sum_{|u| < K} \lambda_{ux} \text{Li}_{ux} = 0.$$

Soit $\rho \in \mathbb{C}$ et $x \in X$, $\langle L e^{\rho x} | ux \rangle = \text{Li}_{ux} + \rho \text{Li}_u + \dots$. Appliquons alors l'opérateur $\langle L(e^{\rho x} - Id) | \cdot \rangle$ à la combinaison linéaire précédente :

$$\forall x \in X, \quad \rho \sum_{|u|=K-1} \lambda_{ux} \text{Li}_u + \sum_{|u| < K-1} \mu_u \text{Li}_u = 0.$$

D'après l'hypothèse de récurrence, nous obtenons le résultat énoncé pour $k = K$. \square

Remarque 3.2.4. *De la \mathbb{C} -indépendance linéaire, on peut déduire l'indépendance linéaire des polylogarithmes Li_w sur n'importe quelle algèbre de fonctions sans monodromie, notamment l'algèbre \mathcal{C} .*

3.3 Propriétés de la série génératrice

Ayant déjà précisé que la série L était une exponentielle de Lie, il en découle alors une factorisation de celle-ci dans la base de Lyndon,

$$L(z) = e^{x_1 \log \frac{1}{1-z}} \left[\prod_{l \in \mathcal{Lyn}X \setminus \{x_0, x_1\}} e^{\text{Li}_{S_l}(z) \check{S}_l} \right] e^{x_0 \log z}. \quad (3.6)$$

Pour tout $l \in \mathcal{Lyn}X \setminus \{x_0, x_1\}$, nous savons que $S_l \in x_0 X^* x_1$. Par conséquent, nous pouvons considérer la série régularisée L_{reg} ainsi que son évaluation en $z = 1$ ([60]),

$$L_{\text{reg}} = \prod_{l \in \mathcal{Lyn}X \setminus \{x_0, x_1\}} e^{\text{Li}_{S_l} \check{S}_l} \quad \text{et} \quad Z = L_{\text{reg}}(1). \quad (3.7)$$

Soit σ le morphisme de monoïde vérifiant $\sigma(x_0) = -x_1, \sigma(x_1) = -x_0$, nous avons alors ([61])

$$L(z) = \sigma[L(1-z)]Z = e^{x_0 \log z} \sigma[L_{\text{reg}}(1-z)] e^{-x_1 \log(1-z)} Z. \quad (3.8)$$

Exemple 3.3.1.

$$\begin{aligned} \text{Li}_1(z) &= -\log(1-z) \\ \text{Li}_2(z) &= -\text{Li}_2(1-z) + \log(1-z) \text{Li}_1(1-z) + \zeta(2) \\ \text{Li}_{2,1}(z) &= -\text{Li}_3(1-z) + \log(1-z) \text{Li}_2(1-z) - \frac{1}{2} \log^2(1-z) \text{Li}_1(1-z) \\ &\quad - \zeta(2) \text{Li}_1(1-z) + \zeta(3) \end{aligned}$$

Remarque 3.3.2. Cette relation entre $L(z)$ et $L(1-z)$ nous donne un accès direct au développement (éventuellement singulier) d'une fonction $\text{Li}_w(z)$, pour $w \in X^* x_1$ au voisinage de $z = 1$.

Exemple 3.3.3.

$$\begin{aligned} \text{Li}_2(z) &= -(1-z) + \log(1-z)(1-z) - \frac{(1-z)^2}{4} + \frac{\log(1-z)(1-z)^2}{2} \\ &\quad + \zeta(2) + \mathcal{O}(\log(1-z)(1-z)^3). \end{aligned}$$

En conséquence de (3.6) et (3.8), nous obtenons

$$L(z) \underset{z \rightarrow 0}{\sim} \exp(x_0 \log z) \quad \text{et} \quad L(z) \underset{z \rightarrow 1}{\sim} \exp\left(x_1 \log \frac{1}{1-z}\right) Z,$$

où l'équivalence doit être comprise comme une équivalence coefficient par coefficient.

Définition 3.3.4. Soit $\zeta_{\sqcup} : (\mathbb{C}\langle X \rangle, \sqcup) \rightarrow (\mathbb{C}, \cdot)$ le morphisme d'algèbre (i.e. pour $u, v \in X^*$, $\zeta_{\sqcup}(u \sqcup v) = \zeta_{\sqcup}(u) \zeta_{\sqcup}(v)$) vérifiant pour tout mot convergent $w \in x_0 X^* x_1$, $\zeta_{\sqcup}(w) = \zeta(w)$, et tel que $\zeta_{\sqcup}(x_0) = \zeta_{\sqcup}(x_1) = 0$.

Exemple 3.3.5. Toujours dans le prolongement de l'exemple 2.2.5,

$$\begin{aligned}\zeta_{\sqcup}(x_1x_0x_1x_0x_1) + 2\zeta_{\sqcup}(x_0x_1^2x_0x_1) + 2\zeta_{\sqcup}(x_0x_1x_0x_1^2) &= \zeta_{\sqcup}(x_0x_1x_0x_1)\zeta_{\sqcup}(x_1) \\ &= \zeta(x_0x_1x_0x_1) \times 0 = 0,\end{aligned}$$

donc

$$\begin{aligned}\zeta_{\sqcup}(x_1x_0x_1x_0x_1) + 2\zeta(2, 1, 2) + 2\zeta(2, 2, 1) &= 0, \\ \text{d'où } \zeta_{\sqcup}(x_1x_0x_1x_0x_1) &= -2\zeta(2, 1, 2) - 2\zeta(2, 2, 1).\end{aligned}$$

Alors, la série génératrice non commutative $Z_{\sqcup} = \sum_{w \in X^*} \zeta_{\sqcup}(w)$ w vérifie $Z_{\sqcup} = Z$ ([58]). Par conséquent, Z_{\sqcup} est l'unique exponentielle de Lie vérifiant $\langle Z_{\sqcup} | w \rangle = \zeta(w)$, pour tout $w \in x_0X^*x_1$ et $\langle Z_{\sqcup} | x_0 \rangle = \langle Z_{\sqcup} | x_1 \rangle = 0$ et $\langle Z_{\sqcup} | w \rangle = \zeta(w)$.

De plus, en définissant $P_w(z) = \frac{\text{Li}_w(z)}{1-z}$ et $P(z) = \sum_{w \in X^*} P_w(z)$ $w = \frac{L(z)}{1-z}$, série génératrice non commutative définie sur LI_C , nous déduisons de la factorisation (3.6) de la série L que

$$P(z) = e^{-(x_1+1)\log(1-z)} L_{\text{reg}}(z) e^{x_0 \log z}. \quad (3.9)$$

Nous déduisons également de la relation (3.8) entre $L(z)$ et $L(1-z)$ celle entre $P(z)$ et $P(1-z)$

$$P(1-z) = \frac{1-z}{z} (\sigma[P](z)) Z \quad (3.10)$$

Exemple 3.3.6.

$$\begin{aligned}P_{x_0x_1^2}(1-z) &= \frac{1-z}{z} \left(-P_{x_0^2x_1}(z) + \log(z)P_{x_0x_1}(z) - \log^2(z)P_{x_1}(z) + \frac{\zeta(3)}{1-z} \right) \\ P_{2,1}(z) &= -\frac{z}{1-z}P_3(1-z) + \frac{z \log(1-z)}{1-z}P_2(1-z) \\ &\quad - \frac{1}{2} \frac{z \log^2(1-z)}{1-z}P_1(1-z) + \frac{\zeta(3)}{1-z}.\end{aligned}$$

Remarque 3.3.7. Nous comprendrons mieux l'intérêt de l'introduction de cette fonction P_w dans le chapitre suivant, où nous préciserons notamment son développement en série entière...

Chapitre 4

Des fonctions symétriques aux sommes harmoniques

Sommaire

4.1 Algèbre des sommes harmoniques	40
4.1.1 Fonctions symétriques	40
4.1.2 Fonctions quasi-symétriques	40
4.1.3 Sommes harmoniques multiples	42
4.1.4 Séries génératrices sur Y	44
4.2 Analyse asymptotique	46
4.2.1 Par la série génératrice	46
4.2.2 Par la formule d'Euler-MacLaurin	52
4.2.3 Constantes associées à des polyzêtas divergents	56
4.2.4 Application pour l'amélioration des algorithmes	59
4.2.5 Efficacités comparées	61

Nous avons introduit une algèbre de fonctions spéciales, l'algèbre des polylogarithmes, ainsi qu'une famille de nombres spéciaux, à savoir les valeurs prises par les polylogarithmes Li_w en $z = 1$, lorsque celles-ci existent. Le présent chapitre vise à répondre à plusieurs questions : d'abord, comment appliquer l'algèbre non commutative aux sommes harmoniques introduites au chapitre 1, qui apparaissent, à l'issue du chapitre précédent, comme des "polyzêtas tronqués". Puis, quel est le lien entre ces sommes harmoniques, les polylogarithmes, et les polyzêtas ? Dans le cas où le mot w est convergent, nous verrons que ce lien est immédiat et découle du théorème d'Abel. Pour tous les autres cas, nous sommes amenés à établir un théorème "à l'Abel" pour préciser ce lien. Enfin, nous proposons plusieurs algorithmes, dont nous analyserons l'efficacité, pour calculer le développement asymptotique de ces sommes harmoniques ; les applications aux structures de données multidimensionnelles seront précisés au chapitre suivant.

4.1 Algèbre des sommes harmoniques

4.1.1 Fonctions symétriques

Soit $\{t_i\}_{i \in \mathbb{N}_+}$ un ensemble infini de variables. Nous définissons les fonctions symétriques (modifiées) $\lambda_k^{(r)}$ et les sommes de puissances $\psi_k^{(r)}$ par

$$\lambda_k^{(r)}(\underline{t}) = \sum_{n_1 > \dots > n_k > 0} t_{n_1}^r \dots t_{n_k}^r \quad \text{et} \quad \psi_k^{(r)}(\underline{t}) = \sum_{n > 0} t_n^{rk}.$$

Remarque 4.1.1. Dans le cas où $r = 1$, nous retrouvons les fonctions symétriques élémentaires classiques [48].

Ces fonctions peuvent être vues comme les coefficients des fonctions génératrices

$$\lambda^{(r)}(\underline{t}|z) = \sum_{k > 0} \lambda_k^{(r)}(\underline{t}) z^k = \prod_{i \geq 1} (1 + t_i^r z) \quad \text{et} \quad \psi^{(r)}(\underline{t}|z) = \sum_{k > 0} \psi_k^{(r)}(\underline{t}) z^{k-1} = \sum_{i \geq 1} \frac{t_i^r}{1 - t_i^r z}.$$

De plus, les susdites fonctions satisfont à l'identité de Newton

$$\frac{d}{dz} \log \lambda^{(r)}(\underline{t}|z) = \psi^{(r)}(\underline{t}|z). \quad (4.1)$$

Nous rappelons enfin la formule de Waring (dans laquelle on adopte la convention $\lambda_0^{(r)} \equiv 1$) :

$$\lambda_k^{(r)} = \frac{(-1)^k}{k!} \sum_{\substack{s_1, \dots, s_k > 0 \\ s_1 + \dots + s_k = k}} \frac{k!}{s_1! \dots s_k!} \left(-\frac{\psi_1^{(r)}}{1} \right)^{s_1} \dots \left(-\frac{\psi_k^{(r)}}{k} \right)^{s_k} \quad (4.2)$$

4.1.2 Fonctions quasi-symétriques

Dorénavant, adoptant un deuxième codage, nous associons le multi-indice $\mathbf{s} = (s_1, \dots, s_r)$ également au mot $w = y_{s_1} \dots y_{s_r}$ défini sur l'alphabet $Y = \{y_i\}_{i \in \mathbb{N}_+}$. Nous rappelons que son poids est l'entier $\sum_{i=1}^r s_i$.

Définition 4.1.2. Soit $w = y_{s_1} \dots y_{s_r} \in Y^*$. Les fonctions quasi-symétriques monômiales F_w et G_w , de profondeur $r = |w|$ et de degré $s_1 + \dots + s_r$, sont définies par

$$F_w(\underline{t}) = \sum_{n_1 > \dots > n_r > 0} t_{n_1}^{s_1} \dots t_{n_r}^{s_r} \quad \text{et} \quad G_w(\underline{t}) = \sum_{n_1 \geq \dots \geq n_r > 0} t_{n_1}^{s_1} \dots t_{n_r}^{s_r}. \quad (4.3)$$

Gessel [28] a démontré que les fonctions $F_w, w \in Y^*$ sont linéairement indépendantes, et plus précisément que ces fonctions quasi-symétriques monômiales forment une base de l'espace des fonctions quasi-symétriques.

Remarque 4.1.3. Notons dès maintenant qu'il existe des relations explicites entre les fonctions F_w et G_w . Par souci de simplicité dans l'expression de ces relations, nous utilisons des notations multi-indices plutôt qu'une indexation par des mots. Rappelons également que $\text{Comp}(n)$ désigne l'ensemble des compositions de n , i.e. les suites finies (i_1, \dots, i_p) d'entiers strictement positifs ayant pour somme n . Si $I = (i_1, \dots, i_p)$ (respectivement $J = (j_1, \dots, j_q)$) est une composition de n (respectivement de p) alors $J \circ I = (i_1 + \dots + i_{j_1}, i_{j_1+1} + \dots + i_{j_1+j_2}, \dots, i_{p-j_q+1} + \dots + i_p)$ est une composition de n . Alors, si s désigne une composition de longueur $|s|$, nous avons ([32])

$$G_s = \sum_{J \in \text{Comp}(|s|)} F_{J \circ s} \tag{4.4}$$

$$F_s = \sum_{J \in \text{Comp}(|s|)} (-1)^{|J|-|s|} G_{J \circ s}, \tag{4.5}$$

la seconde formule étant déduite de la première par inversion de Möbius. Par exemple,

$$\begin{aligned} G_{(1,2,1)} &= F_{(1,2,1)} + F_{(3,1)} + F_{(1,3)} + F_{(4)}, \quad \text{et} \\ F_{(1,2,1)} &= G_{(1,2,1)} - G_{(3,1)} - G_{(1,3)} + G_{(4)}. \end{aligned}$$

Plus généralement, pour une composition de longueur r , le passage d'une forme à l'autre donne naissance à $\text{Card}(\text{Comp}(r)) = 2^{r-1}$ termes.

En particulier, $F_{y_1^k} = \lambda_k^{(1)}$ et $F_{y_k} = G_{y_k} = \psi_k^{(1)}$. De plus, l'intégration de l'équation différentielle (4.1) nous amène sans difficulté à la relation suivante

$$\sum_{k \geq 0} F_{y_1^k} z^k = \exp \left[- \sum_{k \geq 1} F_{y_k} \frac{(-z)^k}{k} \right] \quad \left(\text{ou} \quad \sum_{k \geq 0} G_{y_1^k} z^k = \exp \left[\sum_{k \geq 1} G_{y_k} \frac{z^k}{k} \right] \right). \tag{4.6}$$

Par linéarité, les définitions de F_w et G_w sont étendues aux polynômes de $\mathbb{C}\langle Y \rangle$.

Si u (resp. v) est un mot de Y^* , de longueur r (resp. s) et de degré p (resp. q), $F_{u \sqcup v}$ et $G_{u \sqcup v}$ sont des fonctions quasi-symétriques polynômiales de profondeur $r + s$ et de degré $p + q$, et de plus

$$F_u F_v = F_{u \sqcup v} \quad \text{et} \quad G_u G_v = G_{u \sqcup v}. \tag{4.7}$$

L'identité (4.2) se traduit alors en

$$y_1^k = \frac{(-1)^k}{k!} \sum_{\substack{s_1, \dots, s_k \geq 0 \\ s_1 + \dots + k s_k = k}} \frac{k!}{s_1! \dots s_k!} \frac{(-y_1)^{\sqcup s_1}}{1^{s_1}} \sqcup \dots \sqcup \frac{(-y_k)^{\sqcup s_k}}{k^{s_k}} \tag{4.8}$$

$$= \frac{1}{k!} \sum_{\substack{s_1, \dots, s_k \geq 0 \\ s_1 + \dots + k s_k = k}} \frac{k!}{s_1! \dots s_k!} \frac{y_1^{\sqcup s_1}}{1^{s_1}} \sqcup \dots \sqcup \frac{y_k^{\sqcup s_k}}{k^{s_k}}. \tag{4.9}$$

Exemple 4.1.4.

$$\begin{aligned}
y_1^2 &= \frac{1}{2} (y_1^{\boxplus 2} - y_2) \\
y_1^3 &= \frac{1}{6} (y_1^{\boxplus 3} - 3y_1 \boxplus y_2 + 2y_3) \\
y_1^4 &= \frac{1}{4!} (y_1^{\boxplus 4} - 6y_1^{\boxplus 2} \boxplus y_2 + 8y_1 \boxplus y_3 + 3y_2^{\boxplus 2} - 6y_4).
\end{aligned}$$

4.1.3 Sommes harmoniques multiples

Définition 4.1.5. Pour tout $w = y_{s_1} \dots y_{s_r} \in Y^*$, définissons les applications H_w et \underline{H}_w de \mathbb{N} dans \mathbb{Q} par

$$\begin{aligned}
H_w(N) &= \sum_{N \geq n_1 > \dots > n_r > 0} \frac{1}{n_1^{s_1} \dots n_r^{s_r}} \\
\underline{H}_w(N) &= \sum_{N \geq n_1 \geq \dots \geq n_r > 0} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}.
\end{aligned}$$

Nous posons de plus les conventions $H_\epsilon = \underline{H}_\epsilon \equiv 1$, ainsi que $H_w(0) = \underline{H}_w(0) = 0$. Ces définitions s'étendent par linéarité aux polynômes de $\mathbb{C}\langle Y \rangle$.

Exemple 4.1.6. Avec cette définition,

$$\sum_{1 \leq i_{d-1} \leq \dots \leq i_1 \leq n} \frac{1}{i_1 \dots i_{d-1}} = \underline{H}_{y_1^{d-1}}(n).$$

La constante suivante, qui apparaît dans l'étude de la variance du nombre de maxima dans un hypercube de dimension d ([1])

$$\kappa_d = \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \sum_{i_1 \dots i_{t-1} j_1 \dots j_{d-2-t}}^{(**)} \frac{1}{i_1 \dots i_{t-1} j_1 \dots j_{d-2-t}},$$

avec la somme $(**)$ calculée sur tous les indices vérifiant

$$1 \leq i_1 \leq \dots \leq i_{t-1} \leq l \quad \text{et} \quad 1 \leq j_1 \leq \dots \leq j_{d-2-t} \leq l,$$

se réécrit

$$\kappa_d = \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \underline{H}_{y_1^{t-1}}(l) \underline{H}_{y_1^{d-2-t}}(l).$$

Pour $N \geq 1$ et $w \in Y^*$, la somme $H_w(N)$ (resp. $\underline{H}_w(N)$) peut s'obtenir en évaluant la fonction quasi-symétrique F_w (resp. G_w) en $t_i = 1/i$, si $1 \leq i \leq N$ et $t_i = 0$ sinon ([30]). En particulier, les relations (4.4) se spécialisent en

$$\underline{H}_s = \sum_{J \in \text{Comp}(r)} H_{J \circ s} \quad \text{and} \quad H_s = \sum_{J \in \text{Comp}(r)} (-1)^{l(J)-r} \underline{H}_{J \circ s}. \quad (4.10)$$

Par ailleurs, les équations (4.7) se spécialisent en

Proposition 4.1.7 ([30]). *Pour tous $u, v \in Y^*$, $H_{u \sqcup v} = H_u H_v$ et $\underline{H}_{u \sqcup v} = \underline{H}_u \underline{H}_v$.*

Proposition 4.1.8. *Soit $w = y_s w' \in Y^*$.*

$$H_w(N) = \sum_{l=|w|}^N \frac{H_{w'}(l-1)}{l^s} \quad \text{et} \quad \underline{H}_w(N) = \sum_{l=1}^N \frac{\underline{H}_{w'}(l)}{l^s}.$$

$$H_w(N+1) - H_w(N) = \frac{H_{w'}(N)}{(N+1)^s} \quad \text{et} \quad \underline{H}_w(N+1) - \underline{H}_w(N) = \frac{\underline{H}_{w'}(N+1)}{(N+1)^s}$$

Cette proposition permet une formulation récursive (sur la longueur du mot) de $H_w(N)$, et a pour conséquence :

Théorème 4.1.9. *Pour tout $w = y_s w' \in Y^*$, $H_w(N)$ et $\underline{H}_w(N)$ convergent lorsque $N \rightarrow +\infty$ si et seulement si $s > 1$.*

Démonstration. La minoration évidente $H_w(N) \geq H_{w'}(|w| - 1) \sum_{l=|w|}^N l^{-s}$ prouve la di-

vergence de $H_w(N)$ si $s = 1$, i.e. $w \in y_1 Y^*$. La preuve de la convergence lorsque $w \in Y^* \setminus y_1 Y^*$ nécessite une majoration correcte de $H_{w'}(l)$, pour laquelle nous utilisons le lemme suivant : pour tout $w \in Y^*$, il existe une constante K et un entier α tels que, pour tout $l > 1$, $H_w(l) \leq K \ln^\alpha l$. Ce lemme peut se démontrer à partir de la proposition 4.1.8, et par récurrence sur la longueur de w .

Ce lemme, associé à une dernière utilisation de la proposition 4.1.8, assure la convergence de $H_w(N)$.

La même preuve s'applique pour $\underline{H}_w(N)$. □

Lorsqu'elle existe, la limite de $H_w(N)$ n'est autre que $\sum_{n_1 > \dots > n_r > 0} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}$, c'est-à-dire $\zeta(w)$. La limite de $\underline{H}_w(N)$ est quant à elle notée $\underline{\zeta}(w)$ ("polyzêta large"). Les valeurs $\zeta(w)$ et $\underline{\zeta}(w)$ sont bien entendus reliées par les équations (4.10).

De plus, le passage à la limite de la proposition 4.1.7 donne

Proposition 4.1.10. *Pour tous $u, v \in Y^*$, $\zeta(u \sqcup v) = \zeta(u) \zeta(v)$, et $\underline{\zeta}(u \sqcup v) = \underline{\zeta}(u) \underline{\zeta}(v)$.*

Remarque 4.1.11. *Cette proposition, additionnée à la proposition 3.2.1 page 35, confère à la \mathbb{Q} -algèbre engendrée par les polyzêtas convergents $\{\zeta(w)\}_{w \in x_0 X^* x_1}$, algèbre que nous noterons \mathcal{Z} , une double structure de mélange, permettant notamment d'établir de manière automatique une table des relations entre polyzêtas. Par exemple,*

$$\begin{aligned} x_0 x_1 \sqcup x_0 x_1 = 2x_0 x_1 x_0 x_1 + 4x_0^2 x_1^2 &\implies \zeta(2)^2 = 2\zeta(2, 2) + 4\zeta(3, 1), \\ y_2 \sqcup y_2 = 2y_2^2 + y_4 &\implies \zeta(2)^2 = 2\zeta(2, 2) + \zeta(4), \end{aligned}$$

d'où l'on déduit que $\zeta(4) = 4\zeta(3, 1)$.

Cette algèbre \mathcal{Z} a été étudiée dans [59, 5, 19], et l'on conjecture qu'elle forme une algèbre libre.

4.1.4 Séries génératrices sur Y

Proposition 4.1.12. *Soit $w = y_{s_1} \dots y_{s_r} \in Y^*$, et f_w la série génératrice ordinaire des $(H_w(N))_{N \geq 0}$, i.e. $f_w(z) = \sum_{N \geq 0} H_w(N) z^N$, alors pour tout z tel que $|z| < 1$,*

$$f_w(z) = \frac{\text{Li}_w(z)}{1 - z}.$$

Démonstration.

$$(1 - z)f_w(z) = H_w(0) + \sum_{N \geq 1} \left(H_w(N) - H_w(N - 1) \right) z^N,$$

or, d'après la proposition 4.1.8, $H_w(N) - H_w(N - 1) = \frac{H_w'(N - 1)}{N^{s_1}}$, donc

$$(1 - z)f_w(z) = H_w(0) + \sum_{N \geq 1} \frac{H_w'(N - 1)}{N^{s_1}} z^N = \text{Li}_w(z).$$

□

Cette proposition nous permet alors d'étendre la définition de f_w à des mots $w \in X^*$. En fait, cette fonction génératrice f_w n'est autre que la fonction P_w introduit à la fin du chapitre 3, notation que nous reprenons maintenant. Ainsi,

Corollaire 4.1.13. *Pour tous $u, v \in X^*$ et pour tout $z \in \mathbb{C}$ tel que $|z| < 1$,*

$$P_u(z)P_v(z) = \frac{P_{u \sqcup v}(z)}{1 - z}$$

Exemple 4.1.14. Comme $x_1 \sqcup x_0 x_1 = x_1 x_0 x_1 + 2x_0 x_1^2$, nous obtenons

$$P_{x_1 x_0 x_1}(z) = (1 - z)P_{x_1}(z)P_{x_0 x_1}(z) - 2P_{x_0 x_1^2}(z).$$

Sur l'alphabet Y , on dispose du résultat - immédiat - suivant (le produit de Hadamard \odot de deux séries étant simplement un produit coefficient par coefficient)

Proposition 4.1.15. Pour tous $u, v \in Y^*$, $P_u \odot P_v = P_{u \sqcup v}$.

Theorème 4.1.16 ([54]). L'application $P : u \mapsto P_u$ est un isomorphisme de l'algèbre $(\mathbb{C}\langle Y \rangle, \sqcup)$ sur l'algèbre de Hadamard $(\mathbb{C}\{P_w\}_{w \in Y^*}, \odot)$.

De plus, l'application $H : u \mapsto H_u = (H_u(N))_{N \geq 0}$ (resp. $\underline{H} : u \mapsto \underline{H}_u = (\underline{H}_u(N))_{N \geq 0}$) est un isomorphisme de $(\mathbb{C}\langle Y \rangle, \sqcup)$ (resp. $(\mathbb{C}\langle Y \rangle, \sqcup)$) sur l'algèbre $(\mathbb{C}\{H_w\}_{w \in Y^*}, \cdot)$ (resp. $(\mathbb{C}\{\underline{H}_w\}_{w \in Y^*}, \cdot)$).

Démonstration. Le fait que P, H , et \underline{H} soient des morphismes provient des propositions 4.1.7 page 43 et 4.1.15. Le fait que le morphisme soit bijectif provient quant à lui de la \mathbb{C} -indépendance linéaire des polylogarithmes $\{Li_w, w \in Y^*\}$ prouvé au Théorème 3.2.3 page 35. \square

Considérons les séries génératrices noncommutatives

$$\Lambda = \sum_{w \in Y^*} Li_w w \quad \text{et} \quad H = \sum_{w \in Y^*} H_w w.$$

Soit également π_Y la projection de $\mathbb{C}\langle\langle X \rangle\rangle$ sur $\mathbb{C}\langle\langle Y \rangle\rangle$ annulant les monômes finissant par la lettre x_0 (extension de la projection π_Y introduite à la remarque 2.2.2). Alors

$$\Lambda(z) = \pi_Y L(z) \underset{z \rightarrow 1}{\sim} \exp\left(y_1 \log \frac{1}{1-z}\right) \pi_Y Z. \quad (4.11)$$

Lemme 4.1.17. Soit $\text{Mono}(z) = e^{-(x_1+1)\log(1-z)}$. Alors

$$\pi_Y \text{Mono} = \sum_{k \geq 0} P_{y_1^k} y_1^k \quad \text{et} \quad \pi_Y \text{Mono}^{-1} = \sum_{k \geq 0} P_{y_1^k} (-y_1)^k.$$

Comme le coefficient de z^N dans le développement de Taylor de $P_{y_1^k}$ est $H_{y_1^k}(N)$ alors

Lemme 4.1.18. Soit $\text{Const} = \sum_{k \geq 0} H_{y_1^k} y_1^k$. Alors

$$\text{Const} = \exp\left[-\sum_{k \geq 1} H_{y_1^k} \frac{(-y_1)^k}{k}\right] \quad \text{et} \quad \text{Const}^{-1} = \exp\left[\sum_{k \geq 1} H_{y_1^k} \frac{(-y_1)^k}{k}\right].$$

Proposition 4.1.19. $P(z) \underset{z \rightarrow 0}{\sim} e^{x_0 \log z}$ et $P(z) \underset{z \rightarrow 1}{\sim} \text{Mono}(z)Z$.

Démonstration. De $P(z) = e^{-(x_1+1)\log(1-z)}L_{\text{reg}}(z)e^{x_0\log z}$, nous pouvons tirer le comportement de $P(z)$ au voisinage de 0. Alors, de la formule (3.9), nous obtenons le comportement de $P(z)$ au voisinage de 1. \square

Corollaire 4.1.20. Soit $\Pi(z) = \pi_Y P(z) = \sum_{w \in Y^*} P_w(z) w$. Alors $\Pi(z) \underset{z \rightarrow 1}{\sim} \pi_Y \text{Mono}(z) \pi_Y Z$.

De là, nous tirons les coefficients de P_w et, à l'aide du lemme 4.1.17 page précédente, nous obtenons

Corollaire 4.1.21. $H(N) \underset{N \rightarrow \infty}{\sim} \text{Const}(N) \pi_Y Z$.

Combiné au lemme 4.1.18 page précédente, et à l'équation (4.11), ce corollaire nous conduit également directement au théorème "à l'Abel" suivant,

Theorème 4.1.22 ([55, 56]).

$$\lim_{z \rightarrow 1} \exp\left(-y_1 \log \frac{1}{1-z}\right) \Lambda(z) = \lim_{N \rightarrow \infty} \exp\left(\sum_{k \geq 1} H_{y_k}(N) \frac{(-y_1)^k}{k}\right) H(N) = \pi_Y Z,$$

où la limite doit être comprise comme une limite prise coefficient par coefficient.

4.2 Analyse asymptotique

4.2.1 Par la série génératrice

4.2.1.1 Opérations sur les fonctions génératrices P_w

Pour une fonction $f(z) = \sum_{n \geq 0} a_n z^n$, nous rappelons la notation $[z^n]f(z) = a_n$. Comme multiplications et divisions par z agissent par un simple décalage sur $[z^n]f(z)$, il nous suffit d'étudier l'effet de la multiplication ou de la division par $1 - z$.

$$[z^n](1-z)P_w(z) = H_w(n) - H_w(n-1). \quad (4.12)$$

$$[z^n] \frac{P_w(z)}{1-z} = \sum_{k=0}^n H_w(k) \quad (4.13)$$

$$= \begin{cases} (n+1)H_w(n) - H_{y_{s-1}w'}(n) & \text{si } w = y_s w', \text{ avec } s \neq 1 \\ (n+1)H_w(n) - \sum_{j=1}^n H_{w'}(j-1) & \text{si } w = y_1 w'. \end{cases} \quad (4.14)$$

et, plus généralement,

Proposition 4.2.1.

$$[z^n](1-z)^k P_w(z) = \sum_{j=0}^k \binom{k}{j} (-1)^j H_w(n-j) \quad \text{et} \quad [z^n] \frac{P_w(z)}{(1-z)^k} = \sum_{n \geq j_1 \geq \dots \geq j_k \geq 0} H_w(j_k).$$

4.2.1.2 Opérations sur les coefficients de Taylor H_w

Nous cherchons maintenant comment la multiplication ou la division de $H_w(N)$ par N agit sur P_w .

- Un cas particulier : $w = \epsilon$

Le cas $w = \epsilon$, correspondant à $H_\epsilon \equiv 1$, peut être traité par la

Proposition 4.2.2. *Pour tout $q \in \mathbb{Z}$,*

$$n^q = \begin{cases} [z^n] \text{Li}_{-q}(z) & \text{si } q < 0, \\ [z^n](1-z)^{-1} & \text{si } q = 0, \\ [z^n] \frac{z}{1-z} N_q \left(\frac{1}{1-z} \right) & \text{si } q > 0, \end{cases}$$

où le polynôme N_q est défini par la récurrence

$$N_0(X) = 1, \quad \text{et} \quad N_q(X) = X \left(\sum_{j=0}^{q-1} (-1)^{q-1-j} \binom{q}{j} N_j(X) \right).$$

Exemple 4.2.3.

$$\begin{aligned} n &= [z^n] \left(\frac{z}{(1-z)^2} \right) = [z^n] \left(\frac{1}{(1-z)^2} - \frac{1}{1-z} \right), \\ n^2 &= [z^n] \left(\frac{2z}{(1-z)^3} - \frac{z}{(1-z)^2} \right) = [z^n] \left(\frac{2}{(1-z)^3} - \frac{3}{(1-z)^2} + \frac{1}{1-z} \right). \end{aligned}$$

- Comment diviser par n^k ?

Soit $w = y_{s_1} \cdots y_{s_r}$ et $w' = y_{s_2} \cdots y_{s_r}$ le suffixe de w , de longueur $r-1$. L'expression $n^{-k} H_w(n)$, k entier strictement positif, peut être identifiée comme

$$\begin{aligned} n^{-k} H_w(n) &= n^{-k} H_w(n-1) + n^{-s_1-k} H_{w'}(n-1) \\ &= [z^n] \text{Li}_{y_k w + y_{s_1+k} w'}(z) \\ &= [z^n] [(1-z) P_{y_k w + y_{s_1+k} w'}(z)]. \end{aligned}$$

- Comment multiplier par n^k ?

Afin d'étudier l'effet de la multiplication par n^k , $k > 0$, nous introduisons l'opérateur d'Euler $\theta = z\partial/\partial z$. Alors pour tout entier k ,

$$n^k H_w(n) = [z^n] \theta^k P_w(z). \quad (4.15)$$

Ainsi, il nous suffit de calculer $\theta^k P_w(z)$. Comme dans l'article [53], nous introduisons les notations

Définition 4.2.4. Pour tout $w = x_{i_1} \cdots x_{i_k}$ et toute composition $\mathbf{r} = (r_1, \dots, r_k)$, soit $\tau_{\mathbf{r}}(w)$ défini par $\tau_{\mathbf{r}}(w) = \tau_{r_1}(x_{i_1}) \cdots \tau_{r_k}(x_{i_k})$ avec,

$$\tau_0(x_0) = x_0, \quad \tau_0(x_1) = \frac{z}{1-z}x_1,$$

$$\text{et, pour } r \in \mathbb{N}^*, \quad \tau_r(x_0) = \theta^r x_0 = 0 \quad \text{et} \quad \tau_r(x_1) = \theta^r \frac{zx_1}{1-z}.$$

Nous définissons alors le degré de \mathbf{r} par $\deg(\mathbf{r}) = k$ et son poids (modifié) wgt par $\text{wgt}(\mathbf{r}) = k + r_1 + \cdots + r_k$.

En appliquant successivement l'opérateur θ à L , nous obtenons

Lemme 4.2.5. $\theta^l L = A_l L$, avec

$$A_l(z) = \sum_{\text{wgt}(\mathbf{r})=l} \sum_{w \in X^{\deg(\mathbf{r})}} \prod_{i=1}^{\deg(\mathbf{r})} \binom{\sum_{j=1}^i r_j + j - 1}{r_i} \tau_{\mathbf{r}}(w).$$

Démonstration. C'est une conséquence de la relation de récurrence vérifiée par A_l , i.e. $A_0(z) = 1$, et pour tout $l \in \mathbb{N}$, $A_{l+1}(z) = [\tau_0(x_0) + \tau_0(x_1)]A_l(z) + \theta A_l(z)$. \square

Ce lemme nous permet d'isoler l'expression de $\theta^l \text{Li}_w$, pour tout mot $w \in X^*$.

Exemple 4.2.6.

$$\begin{aligned} A_0(z) &= 1, \\ A_1(z) &= x_0 + \frac{z}{1-z}x_1, \\ A_2(z) &= x_0^2 + \frac{z}{1-z}x_0 \sqcup x_1 + \frac{z^2}{(1-z)^2}x_1^2 + \frac{z}{(1-z)^2}x_1. \end{aligned}$$

Donc, pour $w = x_0^2 x_1$,

$$\begin{aligned} \theta \text{Li}_{x_0^2 x_1} &= \left\langle \left(x_0 + \frac{z}{1-z}x_1 \right) L \mid x_0^2 x_1 \right\rangle \\ &= \text{Li}_{x_0 x_1}, \\ \theta^2 \text{Li}_{x_0^2 x_1} &= \left\langle \left(x_0^2 + \frac{z}{1-z}x_0 \sqcup x_1 + \frac{z^2}{(1-z)^2}x_1^2 + \frac{z}{(1-z)^2}x_1 \right) L \mid x_0^2 x_1 \right\rangle \\ &= \text{Li}_{x_1}. \end{aligned}$$

Lemme 4.2.7. Soit \perp l'opérateur linéaire sur $\mathbb{Z}[X]$ défini par $\perp X^n = (n+1)X^{n+1} + nX^n$ et $\{B_l\}_{l \in \mathbb{N}} \in \mathbb{Z}[X]$ définis par $B_0(X) = 1$ et $B_{l+1}(X) = \perp B_l(X)$. Alors

$$\theta^l \left(\frac{1}{1-z} \right) = \frac{1}{1-z} B_l \left(\frac{z}{1-z} \right).$$

Remarquons que le terme dominant de B_l , $l \geq 1$, est $l!X^l$ et son terme de queue est X .

Exemple 4.2.8.

$$\begin{aligned} B_0(X) &= 1, \\ B_1(X) &= X, \\ B_2(X) &= 2X^2 + X, \\ B_3(X) &= 6X^3 + 6X^2 + X. \end{aligned}$$

L'utilisation de la formule de Leibniz nous amène alors à

$$\begin{aligned} \theta^k P_w(z) &= \sum_{j=0}^k \binom{k}{j} \theta^{k-j} \text{Li}_w(z) \theta^j \frac{1}{1-z} \\ &= \sum_{j=0}^k \binom{k}{j} B_j \left(\frac{z}{1-z} \right) \frac{1}{1-z} \theta^{k-j} \text{Li}_w(z). \end{aligned}$$

Grâce au lemme 4.2.5 page précédente, nous pouvons extraire le coefficient $\theta^l \text{Li}_w$ de w dans $\theta^l L$: celui-ci s'écrit comme une combinaison \mathcal{C} -linéaire de Li_v , avec $|v| \leq |w| - 1$. Nous en déduisons ainsi l'expression de $\theta^k P_w$.

Exemple 4.2.9. Pour $w = x_0^2 x_1$ et $k = 2$,

$$\begin{aligned} \theta^2 P_{x_0^2 x_1}(z) &= \sum_{j=0}^2 \binom{2}{j} B_j \left(\frac{z}{1-z} \right) \frac{1}{1-z} \theta^{2-j} \text{Li}_{x_0^2 x_1}(z) \\ &= \frac{1}{1-z} \text{Li}_{x_1}(z) + 2 \frac{z}{1-z} \frac{1}{1-z} \text{Li}_{x_0 x_1}(z) + \left(2 \left(\frac{z}{1-z} \right)^2 + \frac{z}{1-z} \right) \times \\ &\quad \frac{z}{1-z} \text{Li}_{x_0^2 x_1}(z) \\ &= P_{x_1}(z) + \frac{2z}{1-z} P_{x_0 x_1}(z) + \frac{z^2 + z}{(1-z)^2} P_{x_0^2 x_1}(z). \end{aligned}$$

$$\text{Ainsi,} \quad n^2 H_3(n) = [z^n] \left(P_1(z) + \frac{2z}{1-z} P_2(z) + \frac{z^2 + z}{(1-z)^2} P_3(z) \right).$$

4.2.1.3 Développement singulier de P_w - Développement asymptotique de H_w

Notation. Au cours de cette partie, nous écrirons

$$f_n \sim \sum_{i=0}^{\infty} g_i(n) \quad \text{pour } n \rightarrow +\infty,$$

pour une échelle de fonctions donnée $(g_i)_{i \in \mathbb{N}}$ — i.e. vérifiant $g_{i+1}(n) = o(g_i(n))$, pour tout i — pour exprimer le fait que

$$f_n = \sum_{i=0}^I g_i(n) + \mathcal{O}(g_{I+1}(n)), \quad \text{pour tout } I \geq 0.$$

De la même manière, étant donnée une échelle de fonctions $(h_i)_{i \in \mathbb{N}}$ autour de $z = 1$ — i.e. vérifiant $h_{i+1}(1-z) = o(h_i(1-z))$, quand $z \rightarrow 1$, pour tout i — nous écrivons

$$g(z) \sim \sum_{i=0}^{\infty} h_i(1-z) \quad \text{quand } z \rightarrow 1,$$

pour résumer l'égalité

$$g(z) = \sum_{i=0}^I h_i(1-z) + \mathcal{O}(h_{I+1}(1-z)) \quad \text{pour tout } I \geq 0.$$

Le développement asymptotique des nombres harmoniques H_{y_r} est connu, via la formule sommatoire d'Euler-MacLaurin, faisant apparaître les nombres de Bernoulli $\{B_k\}_{k \geq 0}$,

$$\begin{aligned} H_{y_1}(N) &\sim \log N + \gamma - \sum_{k=1}^{+\infty} \frac{B_k}{k} \frac{1}{N^k}, \\ H_{y_r}(N) &\sim \zeta(r) - \sum_{k=r-1}^{+\infty} \frac{B_{k-r+1}}{k} \binom{k}{r-1} \frac{1}{N^k}, \quad \text{pour } r > 1, \end{aligned} \quad (4.16)$$

De plus, pour $w = y_1^k$, nous savons déjà que l'expression de $[z^N]P_{y_1^k}(z) = H_{y_1^k}(N)$ est donnée par la formule (4.8), d'où nous déduisons le développement lorsque $N \rightarrow +\infty$:

Exemple 4.2.10. De l'exemple 4.1.4 page 42, il découle

$$\begin{aligned} H_{y_1^2}(N) &= \frac{1}{2} \log^2(N) + \gamma \log(N) + \frac{1}{2}(\gamma^2 - \zeta(2)) + \frac{1}{2} \frac{\log(N) + \gamma + 1}{N} - \frac{\log(N)}{12N^2} + \mathcal{O}\left(\frac{1}{N^2}\right), \\ H_{y_1^3}(N) &= \frac{1}{6} \log^3(N) + \frac{1}{2} \gamma \log^2(N) + \frac{1}{2}(\gamma^2 - \zeta(2)) \log(N) - \frac{1}{2} \zeta(2) \gamma + \frac{1}{3} \zeta(3) + \frac{1}{6} \gamma^3 + \\ &\quad \frac{1}{4} \frac{\log^2(N)}{N} + \frac{1}{2}(\gamma + 1) \frac{\log(N)}{N} + \frac{1}{4} (2\gamma + \gamma^2 - \zeta(2)) \frac{1}{N} - \frac{1}{24} \frac{\log^2(N)}{N^2} - \\ &\quad \left(\frac{1}{8} + \frac{\gamma}{12}\right) \frac{\log(N)}{N^2} + \mathcal{O}\left(\frac{1}{N^2}\right). \end{aligned}$$

Ces cas particuliers étant acquis, établissons maintenant le cas général.

Theorème 4.2.11 ([10]). Soit $g \in \mathcal{C}[\{P_w\}_{w \in Y^*}]$. Il existe des coefficients $a_j \in \mathbb{Z}$, $\alpha_j \in \mathbb{Z}$ et $\beta_j \in \mathbb{N}$ tels que

$$g(z) \sim \sum_{j=0}^{+\infty} a_j (1-z)^{\alpha_j} \log^{\beta_j}(1-z), \quad \text{lorsque } z \rightarrow 1.$$

Par conséquent, il existe des coefficients $b_i \in \mathbb{Q}[\gamma, \mathcal{Z}]$, $\eta_i \in \mathbb{Z}$ et $\kappa_i \in \mathbb{N}$ tels que

$$[z^n]g(z) \sim \sum_{i=0}^{+\infty} b_i n^{\eta_i} \log^{\kappa_i}(n), \quad \text{lorsque } n \rightarrow \infty.$$

Démonstration. En considérant le corollaire 4.1.13 page 44, on peut réduire la démonstration au cas où $g(z) = P_w(z)$. En effet, nous en déduisons les développements singuliers de $f(z)g(z)$, pour $f \in \mathcal{C}$ en remarquant que $z = 1 - (1 - z)$ et que $\frac{1}{z} = \sum_{n \geq 0} (1 - z)^n$.

Le développement singulier de $P_w(z)$ se déduit en fait de la proposition 3.10 page 37, qui relie le comportement de P_w en $z = 1$ au comportement d'une certaine combinaison polynômiale de fonctions $\{P_u\}_{u \in X^*}$ en $z = 0$. Ce point a déjà été illustré par l'exemple 3.3.6. De plus, d'après le théorème de Radford 2.2.3, nous pouvons supposer que les mots u entrant dans cette combinaison sont des mots de Lyndon et appartiennent donc à $x_0 X^* x_1 \cup \{x_0, x_1\}$. Mais dans ce cas, n'oublions pas que $P_u(z) = \sum_{n \geq 0} H_u(n) z^n$ et que $P_{x_0}(z) = (1 - z)^{-1} \log(z)$. Ainsi, le développement singulier s'ensuit.

De l'identité

$$(1 - z)^\alpha \log(1 - z)^\beta = (-1)^\beta \beta! (1 - z)^{\alpha+1} P_{y_1^\beta}(z), \quad (4.17)$$

nous déduisons alors le développement asymptotique de $H_w(n)$ en calculant les coefficients de Taylor $[z^n](1 - z)^\alpha \log^\beta(1 - z)$. Puisque nous avons déjà précisé au paragraphe 4.2.1.1 comment la multiplication par $(1 - z)^\alpha$ agit sur les coefficients de Taylor, le problème se ramène au calcul de $[z^n]P_{y_1^\beta} = H_{y_1^\beta}(n)$. Mais pour cela, nous utilisons la formule (4.8) qui achève cette démonstration. \square

Remarque 4.2.12. Malheureusement, même la connaissance complète du développement de $[z^n]g(z)$ ne permet en général que d'accéder au développement asymptotique de $g(z)$, lorsque $z \rightarrow 1$ jusqu'à l'ordre 0 (i.e. la partie singulière du développement). En effet, les polynômes $(1 - z)^k$, $k \geq 0$ sont invisibles en termes de coefficients de Taylor; par exemple,

$$\frac{1}{n} = [z^n] \text{Li}_1(z) = [z^n][\text{Li}_1(z) + (1 - z)^2], \quad \text{dès que } n > 2. \quad (4.18)$$

Remarque 4.2.13. Dans le cas d'une somme finie $\sum_{i \in I} b_i n^{\eta_i} H_1^{\kappa_i}(n)$, nous sommes capables de construire l'unique fonction $f \in \mathcal{C}[(P_w)_{w \in Y^*}]$ telle que,

$$\forall n \in \mathbb{N}, \quad [z^n]f(z) = \sum_{i \in I} b_i n^{\eta_i} H_1^{\kappa_i}(n), \quad (4.19)$$

comme illustré par les exemples 4.2.3 page 47 et 4.2.9 page 49.

Notons enfin que la preuve du théorème 4.2.11 page 50 fournit une construction *effective* du développement asymptotique des coefficients de Taylor. En particulier, appliquée à $g(z) = P_w(z)$ directement, celle-ci permet de trouver le développement de $H_w(N)$, comme illustré par l'exemple ci-dessous.

Exemple 4.2.14. *De l'exemple 3.3.6 nous obtenons, lorsque $z \rightarrow 1$*

$$\begin{aligned} P_{2,1}(z) &= \frac{\zeta(3)}{1-z} + \log(1-z) - 1 - \frac{\log^2(1-z)}{2} \\ &+ (1-z) \left(-\frac{\log^2(1-z)}{4} + \frac{\log(1-z)}{4} \right) + \mathcal{O}(|1-z|). \end{aligned}$$

Mais

$$\begin{aligned} [z^N] \zeta(3)(1-z)^{-1} &= \zeta(3), \\ [z^N] \log(1-z) &= -N^{-1}, \\ [z^N] \frac{\log^2(1-z)}{2} &= [z^N] \frac{2!(1-z)P_{y_1^2}(z)}{2} \\ &= [z^N](1-z)P_{y_1^2}(z) \\ &= H_{y_1^2}(N) - H_{y_1^2}(N-1), \\ &\vdots \end{aligned}$$

Finalement, à l'aide de l'exemple 4.2.10 :

$$[z^N]P_{2,1}(z) = H_{2,1}(N) = \zeta(3) - \frac{\log(N) + 1 + \gamma}{N} + \frac{1}{2} \frac{\log(N)}{N^2} + \mathcal{O}\left(\frac{1}{N^2}\right).$$

De plus, prolongeant l'exemple 4.1.14,

$$\begin{aligned} P_{1,2}(z) &= (1-z)P_1(z)P_2(z) - 2P_{2,1}(z) \\ &= \frac{-z \log(1-z)}{1-z} \left(-P_2(1-z) + \log(1-z)P_1(1-z) + \frac{\zeta(2)}{z} \right) - 2P_{2,1}(z). \end{aligned}$$

Ainsi,

$$[z^N]P_{1,2}(z) = H_{1,2}(N) = \zeta(2)\gamma - 2\zeta(3) + \zeta(2)\log(N) + \frac{\zeta(2) + 2}{2N} + \mathcal{O}\left(\frac{1}{N^2}\right).$$

4.2.2 Par la formule d'Euler-MacLaurin

Le principal défaut de l'algorithme précédemment évoqué est la construction préalable de la série $\sigma(P)Z$, σ étant le morphisme tel que $\sigma(x_i) = -x_{1-i}$ pour $i \in \{0, 1\}$. Cette construction est coûteuse (en termes de temps de calcul) si l'on cherche à obtenir la sé-

rie complète pour tous les mots d'un certain poids. Cette considération nous amène à l'algorithme suivant que nous allons d'abord présenter sous sa forme la plus simple.

4.2.2.1 Cas d'un mot convergent dont tous les suffixes sont convergents

En d'autres termes, intéressons-nous au cas d'un mot $w \in Y^*$ ne contenant pas la lettre y_1 . D'après la proposition 4.1.8 page 43, $H_w(n)$ peut s'exprimer récursivement en fonction des $H_{w'}(k)$, w' décrivant l'ensemble des suffixes de w . Dès lors, nous proposons l'algorithme récursif consistant à développer asymptotiquement le numérateur, puis à utiliser la formule sommatoire d'Euler-MacLaurin affirmant, pour tous entiers q, M, N avec $N > M$,

$$\sum_{n=M}^N f(n) = \int_M^N f(x)dx + \frac{f(M) + f(N)}{2} + \sum_{j=1}^m \frac{B_{2j}}{(2j)!} \left(f^{(2j-1)}(N) - f^{(2j-1)}(M) \right) + R_{2m}, \quad (4.20)$$

$$\text{où } R_m = \frac{1}{(2m+1)!} \int_M^N B_{2m+1}(x - [x]) f^{(2m+1)}(x) dx.$$

celle-ci permettant justement d'initialiser l'algorithme lorsque $|w| = 1$, cf équation (4.16).

Exemple 4.2.15. Soit $l = y_4 y_2$. D'après la proposition 4.1.8 page 43,

$$H_{4,2}(N) = \zeta(4, 2) - \sum_{i=N+1}^{\infty} \frac{H_2(i-1)}{i^4},$$

Mais $H_2(i-1) = \zeta(2) - \frac{1}{i} - \frac{1}{2} \frac{1}{i^2} + \mathcal{O}\left(\frac{1}{i^3}\right)$, d'où

$$H_{4,2}(N) = \zeta(4, 2) - \zeta(2) \sum_{i=N+1}^{\infty} \frac{1}{i^4} + \sum_{i=N+1}^{\infty} \frac{1}{i^5} + \frac{1}{2} \sum_{i=N+1}^{\infty} \frac{1}{i^6} + \sum_{i=N+1}^{\infty} \mathcal{O}\left(\frac{1}{i^7}\right)$$

En développant les sommes en N , nous trouvons finalement

$$H_{4,2}(N) = \zeta(4, 2) - \frac{1}{3} \frac{\zeta(2)}{N^3} + \frac{\frac{1}{2} \zeta(2) + \frac{1}{4}}{N^4} - \frac{\frac{1}{3} \zeta(2) + \frac{2}{5}}{N^5} + \mathcal{O}\left(\frac{1}{N^6}\right).$$

4.2.2.2 Et sinon ? Décomposition de Radford

Dans le cas général, nous cherchons un développement à l'ordre q , c'est-à-dire un polynôme $p \in \mathbb{R}[X, Y]$ vérifiant

$$H_w(N) = p\left(\log N, \frac{1}{N}\right) + \mathcal{O}\left(\frac{1}{N^q}\right). \quad (4.21)$$

Afin de proposer un algorithme le plus proche possible de sa version implémentée, nous distinguons la notation DA faisant référence à l'algorithme lui-même de la notation EM faisant référence à une utilisation directe de la formule d'Euler-MacLaurin. De plus, la notation $DA_q(H_w(n))$ désigne le développement de $H_w(n)$ jusqu'à l'ordre q .

L'algorithme est bâti en deux temps ([11]).

Dans un premier temps, nous "stockons" une table de développements, pour des mots $w \in \mathcal{Lyn}Y$. Pour cela nous procédons par récurrence sur la longueur de w .

- Si $w = y_s$, alors $DA_q(H_w(n)) = EM_q(H_w(n))$, un développement déjà donné explicitement en formule (4.16).
- Supposons connus les développements des $H_w(n)$ pour tous les mots de Lyndon de longueur inférieure ou égale à L , et considérons un mot de Lyndon de longueur $L + 1$, $w = y_s u$ (rappelons que d'après la définition des mots de Lyndon, y_s ne peut être égal à y_1 et donc le mot est convergent.) Nous posons alors

$$DA_q(H_w(N)) = \zeta(w) - EM_q \left(\sum_{i=N+1}^{\infty} \frac{DA_{q-s+1}(H_u(i-1))}{i^s} \right). \quad (4.22)$$

Il y a alors deux possibilités,

- ▷ Si $u \in \mathcal{Lyn}Y$ alors $DA_{q-s+1}(H_u(i-1))$ est supposé connu.
- ▷ Si $u \notin \mathcal{Lyn}Y$, la décomposition de Radford nous permet d'écrire u comme une somme finie de termes $t = c l_1 \sqcup \dots \sqcup l_r$, où $c \in \mathbb{Q}$, $l_i \in \mathcal{Lyn}Y$, et $|l_i| \leq |u|$, et donc

$$DA_{q-s+1}(H_t(i-1)) = c \prod_{p=1}^r DA_{q-s+1}(H_{l_p}(i-1)).$$

Dans un deuxième temps, si $w \notin \mathcal{Lyn}Y$, comme précédemment, nous utilisons la décomposition de Radford pour nous ramener au cas précédent.

4.2.2.3 Ou avec la décomposition de Taylor

Une solution alternative est la suivante : d'après le théorème de Radford 2.2.4, tout $w \in Y^*$ peut s'exprimer sous la forme,

$$w = \sum_{k=0}^{|w|} c_k(w) \sqcup y_1^{\sqcup k}, \quad (4.23)$$

avec $c_k(w) \in \mathbb{C}_2$, pour tout k . Cherchons à expliciter les polynômes convergents $c_k(w)$. Pour cela, posons $D : \mathbb{C}\langle Y \rangle \rightarrow \mathbb{C}\langle Y \rangle$ l'application linéaire définie, pour tout $p \in \mathbb{C}\langle Y \rangle$ et pour tout mot $w \in Y^*$ par la dualité

$$\langle Dp|w \rangle = \langle p|y_1 w \rangle. \quad (4.24)$$

En particulier, $Dw = 0$ lorsque w est convergent et $D(y_1 w) = w$ pour tout mot $w \in Y^*$. En fait, nous pouvons prouver que D est une dérivation pour le produit de mélange \sqcup .

Dans la suite, afin d'alléger les notations, les produits et les puissances doivent être effectués avec le produit \bowtie .

Proposition 4.2.16 ([11]). Soit $w \in Y^*$, un mot de longueur $|w|$. Alors les polynômes $c_k(w)$ sont données par

$$c_k(w) = \frac{1}{k!} \sum_{i=0}^{|w|-k} \frac{(-y_1)^i D^i}{i!} D^k(w).$$

Démonstration. Pour un polynôme $p \in \mathbb{C}[X]$ de degré l , le développement de Taylor est fini, et donné par

$$p(x) = p(y) + Dp(y)(x - y) + \cdots + \frac{D^l p(y)}{l!} (x - y)^l.$$

Donc, pour les valeurs $x = 0, y = y_1, p = D^k(w)$, nous obtenons

$$k!c_k(w) = D^k(w) - D D^k(w)y_1 + \cdots + \frac{D^l}{l!} D^k(w)(-y_1)^l.$$

□

Remarque 4.2.17. Comme $D^{k+i}(w) = 0$ dès que $k+i > |w|$, cette formule peut se résumer à

$$c_k(w) = \frac{1}{k!} e^{-y_1 D} D^k(w)$$

avec la convention $\exp(-y_1 D) = \sum_{i \geq 0} (-y_1)^i D^i / i!$, i.e. en faisant commuter D et y_1 .

Exemple 4.2.18. Soit $w = y_1 y_4 y_2$, pour lequel $D(w) = y_4 y_2$ et donc $D^k(w) = 0$, dès que $k \geq 2$. En appliquant la proposition 4.2.16,

$$\begin{aligned} c_0 &= \sum_{i=0}^3 \frac{(-y_1)^i D^i}{i!} (y_1 y_4 y_2) = y_1 y_4 y_2 - y_1 \bowtie y_4 y_2 \\ &= -y_4 y_1 y_2 - y_4 y_2 y_1 - y_4 y_3 - y_5 y_2 \\ c_1 &= \sum_{i=0}^2 \frac{(-y_1)^i D^i}{i!} (y_4 y_2) = y_4 y_2 \\ c_k &= 0 \text{ pour } k \geq 2 \end{aligned}$$

Ainsi, $y_1 y_4 y_2 = c_0 + c_1 \bowtie y_1 = -y_4 y_1 y_2 - y_4 y_2 y_1 - y_4 y_3 - y_5 y_2 + y_4 y_2 \bowtie y_1$. Remarquons que dans cet exemple, la décomposition de Taylor coïncide avec la décomposition de Radford.

Nous pouvons résumer la forme générale prise par les développements des $H_w(n)$ sous la forme suivante :

Proposition 4.2.19. *Pour tout $w \in Y^*$, et pour tout $M > 0$, il existe des coefficients $\alpha_i, \beta_{k,j}$ et C_w appartenant à la $\mathbb{Q}[\gamma]$ -algèbre engendrée par \mathcal{Z} tels que*

$$H_w(n) = \sum_{i=1}^{|w|} \alpha_i \ln^i(n) + C_w + \sum_{j=1}^M \frac{1}{n^j} \sum_{k=0}^{|w|-1} \beta_{k,j} \ln^k(n) + \mathcal{O}\left(\frac{1}{n^M}\right).$$

4.2.3 Constantes associées à des polyzêtas divergents

Définition 4.2.20. *Nous définissons, pour tout mot $w \in Y^*$, la “constante d’Euler généralisée”, notée $\zeta_{\sqcup}(w)$, comme la constante intervenant dans le développement asymptotique de la somme harmonique $H_w(n)$ dans l’échelle des $\{n^\alpha \log^\beta(n), \alpha, \beta \in \mathbb{Z}\}$.*

Proposition 4.2.21. *L’application $\zeta_{\sqcup} : (\mathbb{C}\langle Y \rangle, \sqcup) \rightarrow (\mathbb{C}, \cdot)$ est un morphisme d’algèbre (i.e. vérifie pour tous mots $u, v \in Y^*$, $\zeta_{\sqcup}(u \sqcup v) = \zeta_{\sqcup}(u)\zeta_{\sqcup}(v)$), tel que pour $w \in Y^* \setminus y_1 Y^*$, $\zeta_{\sqcup}(w) = \zeta(w)$ et tel que $\zeta_{\sqcup}(y_1) = \gamma$.*

Démonstration. Le fait que ζ_{\sqcup} vérifie la relation de mélange est une conséquence directe de la relation de mélange vérifiée par les sommes harmoniques (cf Proposition 4.1.7 page 43), et du fait que le développement de $H_w(n)$ ne fait intervenir que des puissances positives de $\log(n)$ et des puissances négatives de n et donc que le terme constant intervenant dans le développement de $H_u(n)H_v(n)$ ne peut être que le produit des termes constants $\zeta_{\sqcup}(u)\zeta_{\sqcup}(v)$.

En fait, les conditions de la proposition sont des conditions suffisantes, et caractérisent ainsi complètement la constante $\zeta_{\sqcup}(w)$. En effet, comme nous venons de le voir dans la section 4.2.2.3, nous pouvons décomposer n’importe quel mot $w \in Y^*$ sous la

forme $w = \sum_{j=0}^{|w|} c_j(w) \sqcup y_1^{\sqcup j}$, avec $c_j(w)$ des polynômes convergents. En appliquant le

morphisme $u \mapsto H_u$, ceci est équivalent à

$$H_w = \sum_{j=0}^{|w|} H_{c_j(w)} H_1^j. \quad (4.25)$$

D’une part, sous les conditions de la proposition, la décomposition du mot w conduit

à $\zeta_{\sqcup}(w) = \sum_{j=0}^{|w|} \zeta(c_j(w)) \gamma^j$. D’autre part, la décomposition (4.25) amène précisément au

fait que le terme constant intervenant dans son développement asymptotique coïncide avec l’expression précédente. \square

Exemple 4.2.22. Soit $w = y_1^2 y_2$, alors

$$\begin{aligned} c_3(w) &= 0, \\ c_2(w) &= 2y_2, \\ c_1(w) &= -y_2 y_1 - y_3, \\ c_0(w) &= y_2 y_1^2 + y_3 y_1 + \frac{1}{2} y_4, \end{aligned}$$

et donc

$$\begin{aligned} H_w &= H_{c_0(w)} + H_{c_1(w)} H_1 + H_{c_2(w)} H_1^2, \\ \text{et } \zeta_{\boxplus}(w) &= \zeta(c_0(w)) + \zeta(c_1(w))\gamma + \zeta(c_2(w))\gamma^2 \\ &= \frac{7}{10}\zeta(2)^2 - 2\zeta(3)\gamma + \frac{\zeta(2)}{2}\gamma^2, \end{aligned}$$

après utilisation de la table de réduction des polyzêtas.

Proposition 4.2.23.

$$\zeta_{\boxplus}(y_1^k) = \sum_{\substack{s_1, \dots, s_k > 0 \\ s_1 + \dots + s_k = k}} \frac{(-1)^k}{s_1! \dots s_k!} (-\gamma)^{s_1} \left(-\frac{\zeta(2)}{2}\right)^{s_2} \dots \left(-\frac{\zeta(k)}{k}\right)^{s_k}.$$

Démonstration. Ceci est une conséquence directe de l'application du morphisme ζ_{\boxplus} sur l'égalité (4.8). \square

Par conséquent,

Theorème 4.2.24 ([54]). Pour $k > 0$, la constante $\zeta_{\boxplus}(y_1^k)$ est un polynôme de degré k en γ , à coefficients dans $\mathbb{Q}[\zeta(2), \zeta(2i+1)]_{0 < i \leq (k-1)/2}$. De plus, pour $l = 0, \dots, k$, le coefficient de γ^l est ou nul, ou de poids $k - l$.

Exemple 4.2.25.

$$\begin{aligned} \zeta_{\boxplus}(y_1^2) &= \frac{\gamma^2 - \zeta(2)}{2}, \\ \zeta_{\boxplus}(y_1^3) &= \frac{\gamma^3 - 3\zeta(2)\gamma + 2\zeta(3)}{6}, \\ \zeta_{\boxplus}(y_1^4) &= \frac{80\zeta(3)\gamma - 60\zeta(2)\gamma^2 + 6\zeta(2)^2 + 10\gamma^4}{240}, \\ \zeta_{\boxplus}(y_1^5) &= \frac{-20\zeta(2)\zeta(3) + 20\zeta(3)\gamma^2 - 10\zeta(2)\gamma^3 + \gamma^5 + 3\zeta^2(2)\gamma + 24\zeta(5)}{120}. \end{aligned}$$

Considérons les polynômes (exponentiels) de Bell $b_{n,k}(t_1, \dots, t_{n-k+1})$, en les variables $\{t_l\}_{l \geq 1}$, définis par les séries génératrices exponentielles :

$$\sum_{n=0}^{\infty} \sum_{k=0}^n b_{n,k}(t_1, \dots, t_{n-k+1}) \frac{v^n u^k}{n!} = \exp\left(u \sum_{l=1}^{\infty} t_l \frac{v^l}{l!}\right). \quad (4.26)$$

Exemple 4.2.26 (Polynômes $b_{n,k}$ pour $n \leq 5$).

$n \backslash k$	0	1	2	3	4	5
0	1					
1	0	t_1				
2	0	t_2	t_1^2			
3	0	t_3	$3t_1t_2$	t_1^3		
4	0	t_4	$3t_2^2 + 4t_1t_3$	$6t_1^2t_2$	t_1^4	
5	0	t_5	$10t_2t_3 + 5t_1t_4$	$10t_1^2t_3 + 15t_1t_2^2$	$10t_1^3t_2$	t_1^5

En particulier,

Lemme 4.2.27. Soit $t_m = (-1)^{m+1}(m-1)!\zeta_{\sqcup}(m)$, pour $m \geq 1$. Alors

$$\exp \left[- \sum_{k \geq 1} \zeta_{\sqcup}(k) \frac{(-y_1)^k}{k} \right] = 1 + \sum_{n \geq 1} \left[\sum_{k=1}^n b_{n,k}(\gamma, -\zeta(2), 2\zeta(3), \dots) \right] \frac{y_1^n}{n!}.$$

Construisons maintenant la série génératrice noncommutative des $\zeta_{\sqcup}(w)$ et considérons la partie constante dans chacun des deux membres de $H(N) \xrightarrow{N \rightarrow \infty} \text{Const}(N) \pi_Y Z$, nous obtenons

Théorème 4.2.28 ([55, 56]). Soit $Z_{\sqcup} = \sum_{w \in Y^*} \zeta_{\sqcup}(w) w$. Alors

$$Z_{\sqcup} = \left[1 + \sum_{n \geq 1} \left(\sum_{k=1}^n b_{n,k}(\gamma, -\zeta(2), 2\zeta(3), \dots) \right) \frac{y_1^n}{n!} \right] \pi_Y Z.$$

L'identification du coefficient de $y_1^k w$ dans chaque membre conduit au

Corollaire 4.2.29. Pour tout $w \in Y^* \setminus y_1 Y^*$ et $k \geq 0$, π_X désignant la projection canonique de Y sur X (réciproque de π_Y),

$$\zeta_{\sqcup}(y_1^k w) = \sum_{i=0}^k \frac{\zeta_{\sqcup}(\pi_X(y_1^{k-i} w))}{i!} \left[\sum_{j=1}^i b_{i,j}(\gamma, -\zeta(2), 2\zeta(3), \dots) \right].$$

Exemple 4.2.30.

$$\begin{aligned} \zeta_{\sqcup}(y_1^2 y_2) &= \zeta_{\sqcup}(x_1^2 x_0 x_1) + \zeta_{\sqcup}(x_1 x_0 x_1) b_{1,1}(\gamma) + \frac{\zeta(2)}{2!} (b_{2,1}(-\zeta(2)) + b_{2,2}(\gamma)) \\ &= 3\zeta(2, 1, 1) - 2\zeta(2, 1)\gamma + \frac{\zeta(2)}{2} (-\zeta(2) + \gamma^2), \end{aligned}$$

et après utilisation de la table de réduction, il vient

$$\zeta_{\sqcup}(y_1^2 y_2) = \frac{\zeta(2)}{2} \gamma^2 - 2\zeta(3)\gamma + \frac{7}{10}\zeta(2)^2,$$

résultat en accord avec l'exemple 4.2.22 page 56.

Remarque 4.2.31. Nous pouvons en fait déduire la proposition 4.2.23 page 57 du corollaire 4.2.29 page ci-contre, dans le cas particulier $w = \epsilon$, puisque toutes les valeurs $\zeta_{\sqcup}(x_1^{k-i})$ sont nulles, sauf si $i = k$, dans quel cas $\zeta_{\sqcup}(\epsilon) = 1$, d'où

$$\zeta_{\sqcup}(y_1^k) = \frac{1}{k!} \left[\sum_{j=1}^k b_{k,j}(\gamma, -\zeta(2), 2!\zeta(3), \dots, (-1)^{k-1}(k-1)!\zeta(k)) \right].$$

Par conséquent,

Theorème 4.2.32 ([54]). Pour $w \in Y^* \setminus y_1 Y^*$, $k \geq 0$, la constante $\zeta_{\sqcup}(y_1^k w)$ associée à $\zeta(y_1^k w)$ est un polynôme de degré k en γ et à coefficients dans \mathcal{Z} . De plus, pour $l = 0, \dots, k$, le coefficient de γ^l est ou nul ou de poids $|w| + k - l$.

En utilisant enfin l'expression explicite de $\zeta_{\sqcup}(x_1^k x_0 u)$, $u \in X^* x_1$ donnée dans [58]

$$\zeta_{\sqcup}(x_1^k x_0 u) = \zeta(x_0(x_1^k \sqcup u)),$$

nous en déduisons l'expression

$$\zeta_{\sqcup}(y_1^k w) = \sum_{i=0}^k \frac{\zeta(x_0[(-x_1)^{k-i} \sqcup u])}{i!} \left[\sum_{j=1}^i b_{i,j}(\gamma, -\zeta(2), 2!\zeta(3), \dots) \right], \quad (4.27)$$

avec $\pi_X w = x_0 u$.

Remarque 4.2.33. Dans cette section, nous avons effectué une régularisation de y_1 , pour le produit \sqcup à γ . Une régularisation à 0, comme proposée dans [58] ou dans [64] conduit alors à une expression de $\zeta_{\sqcup}(y_1^k w)$ comme un élément de l'algèbre \mathcal{Z} , expression que nous pouvons interpréter comme le terme constant intervenant dans le développement asymptotique de $H_{y_1^k w}(n)$, mais dans l'échelle des $\{n^\alpha H_1^\beta(n), \alpha, \beta \in \mathbb{Z}\}$. La démonstration de cette affirmation tient essentiellement à la décomposition de n'importe quel mot en un polynôme en y_1 , à coefficients dans C_2 , exactement comme dans la démonstration de la proposition 4.2.21.

4.2.4 Application pour l'amélioration des algorithmes

Le principal problème de l'approche de la section 4.2.2 est le coût (exponentiel) de la décomposition, soit sous sa forme "Radford", soit sous sa forme "Taylor". La récurrence (dans le cas convergent) s'effectuant sur la longueur du mot, le critère limitant est

précisément celui-ci, et absolument pas le poids. Par exemple, alors que le développement de $H_{52,123}(n)$ nécessite moins d'une seconde avec cette méthode, l'extraction du coefficient du mot $x_0^{51}x_1x_0^{122}x_1$ dans la série $\sigma(P)Z$ semble difficilement atteignable en un temps "raisonnable" (voire simplement atteignable).

Pour pallier cette difficulté en termes de coût de calcul, nous partons de la considération suivante. Nous souhaiterions utiliser la formule (4.22) en toute généralité sans se soucier du fait que le mot soit convergent ou divergent, sous la forme *purement heuristique* suivante

$$\text{DA}_q(H_w(N)) \approx \text{EM}_q \left(\sum_{i=1}^N \frac{\text{DA}_{q-s+1}(H_u(i-1))}{i^s} \right). \quad (4.28)$$

Il s'avère alors que les coefficients fournis par la formule sommatoire d'Euler-MacLaurin (4.20) sont *justes*, à l'exception notable du terme constant, la constante d'Euler généralisée $\zeta_{\boxminus}(w)$ que nous avons déjà explicitée.

Démonstration. Raisonnons par récurrence sur la longueur de w . L'initialisation pour un mot de longueur 1 tient aux formules (4.16). Supposons que

$$H_w(n-1) = \sum_{i=1}^{|w|} \alpha_i \ln^i(n) + \zeta_{\boxminus}(w) + \sum_{j=1}^M \frac{1}{n^j} \sum_{k=0}^{|w|-1} \beta_{k,j} \ln^k(n) + \mathcal{O}\left(\frac{1}{n^M}\right),$$

et considérons le mot $y_s w$.

$$\begin{aligned} H_{y_s w}(N) &= \sum_{n=1}^N \frac{H_w(n-1)}{n^s} \\ &= \sum_{n=1}^N \left(\sum_{i=1}^{|w|} \alpha_i \frac{\ln^i(n)}{n^s} + \frac{\zeta_{\boxminus}(w)}{n^s} + \sum_{j=1}^M \frac{1}{n^{j+s}} \sum_{k=0}^{|w|-1} \beta_{k,j} \ln^k(n) + \mathcal{O}\left(\frac{1}{n^{M+s}}\right) \right). \end{aligned}$$

Alors, l'utilisation de la formule d'Euler-MacLaurin entre 1 et N nous donne

$$H_{y_s w}(N) = \sum_{i=1}^{|w|+1} \alpha'_i \ln^i(N) + C_{y_s w} + \sum_{j=1}^{M+s-1} \frac{1}{N^j} \sum_{k=0}^{|w|} \beta'_{k,j} \ln^k(N) + \mathcal{O}\left(\frac{1}{N^{M+s-1}}\right),$$

où les constantes α'_i et $\beta'_{k,j}$ ne changent pas si on développe $H_w(n-1)$ à un ordre supérieur, ce qui n'est pas le cas pour la constante $C_{y_s w}$. Or cette constante d'Euler généralisée n'est autre que $\zeta_{\boxminus}(y_s w)$ et a déjà été explicitée au corollaire 4.2.29. \square

L'équation (4.28) peut donc être appliquée, en prenant soin de modifier le terme constant par la constante $\zeta_{\boxminus}(w)$.

Exemple 4.2.34. Pour $w = y_1 y_2$,

$$H_w(N) = \sum_{i=1}^N \frac{H_2(i-1)}{i},$$

or

$$H_2(i-1) = \zeta(2) - \frac{1}{i} - \frac{1}{2} \frac{1}{i^2} + \mathcal{O}\left(\frac{1}{i^3}\right).$$

L'application directe de la formule d'Euler (4.20) entre $M = 1$ et N conduit à l'expression

$$\zeta(2) \log(N) - \frac{1}{2} \zeta(3) - \zeta(2) + \zeta(2) \gamma + \frac{1}{N} \left(\frac{\zeta(2)}{2} + 1 \right) - \frac{1}{N^2} \left(\frac{\zeta(2)}{12} - \frac{1}{4} \right) + \mathcal{O}\left(\frac{1}{N^3}\right),$$

dans laquelle il convient de remplacer le terme $-\frac{1}{2} \zeta(3) - \zeta(2) + \zeta(2) \gamma$ par le terme $\zeta_{\mathbb{W}}(y_1 y_2) = \gamma \zeta(2) - 2 \zeta(3)$. Ainsi,

$$H_w(N) = \zeta(2) \log(N) + \gamma \zeta(2) - 2 \zeta(3) + \frac{1}{N} \left(\frac{\zeta(2)}{2} + 1 \right) - \frac{1}{N^2} \left(\frac{\zeta(2)}{12} - \frac{1}{4} \right) + \mathcal{O}\left(\frac{1}{N^3}\right).$$

4.2.5 Efficacités comparées

Voici quelques exemples de temps d'exécution, en secondes, pour calculer le développement de $H_w(N)$ jusqu'à $\mathcal{O}\left(\frac{1}{N^3}\right)$, dans lesquels la dénomination "Algo i ", $i \in \{1, 2, 3\}$ se réfère au i -ème algorithme présenté. Pour *Algo 1*, les séries L et $\sigma(L)Z$ sont supposées construites une fois pour toutes, et le temps d'exécution démarre avec l'extraction du coefficient de w dans les deux séries. La case vide sur la troisième ligne indique précisément que la construction de la série $\sigma(L)Z$ ne peut être calculée en un temps "raisonnable".

Ces exemples illustrent la nette efficacité du dernier algorithme proposé, et mettent en évidence le facteur limitant que constitue le poids ou la longueur pour chacun des deux autres algorithmes.

mot w	poids	longueur	Algo 1	Algo 2	Algo 3
$y_1 y_2 y_3$	6	3	2.4	0.7	0.1
$y_1 y_2 y_1 y_2$	6	4	2.5	3.6	0.4
$y_1^4 y_2$	6	5	2.5	11.4	1.3
$y_1 y_2 y_3 y_1$	7	4	3.5	2.2	0.3
$y_1^3 y_2^2$	7	5	2.5	28.5	0.8
$y_1 y_2 y_1^4$	7	6	2.6	20.1	2.3
$y_1 y_2 y_3 y_1^2$	8	5	3.5	5.1	0.6
$y_1^3 y_2 y_1^3$	8	7	2.7	154	3.2
$y_1 y_2^2 y_3 y_1^2$	10	6	4.7	13.4	0.9
$y_1^4 y_2 y_1^4$	10	8	3.3	>4000	22
$y_1^5 y_7$	12	6	-	45.9	5.1
$y_1^4 y_3^2 y_1^2$	12	8	-	>4000	20

Chapitre 5

Applications

Sommaire

5.1 Aux arbres hyperquaternaires	65
5.1.1 Autour de la racine	65
5.1.2 Retour sur les paramètres additifs	69
5.2 Aux maxima pour des données multidimensionnelles	74
5.2.1 Introduction	74
5.2.2 Équivalent asymptotique de $\text{Var}(K_{n,d})$	76
5.2.3 Détermination du développement complet	78

En guise d'introduction, considérons l'exemple suivant, tiré de l'article "Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes" [27]. Foata, Han et Lass considèrent le problème du collecteur de vignettes dans le cas où le collectionneur, supposé être l'aîné, possède une fratrie de r frères. Lorsqu'il trouve une vignette qu'il a déjà, il la donne à son frère le plus âgé, qui à son tour la conserve, ou bien la donne au plus âgé des $r - 1$ frères restants, et ainsi de suite. En définissant T le nombre (aléatoire) de barres chocolatées qu'il doit acheter pour compléter sa collection constituée de m vignettes différentes, et $M_T^{(j)}$ le nombre d'emplacements vides dans la collection du j -ème frère à cet instant, les auteurs prouvent que

$$\mathbb{E}[M_T^{(k)}] = \sum_{j=0}^k K_m^{(j)}, \quad (5.1)$$

où pour tout $m \geq 1$,

$$K_m^{(0)} = 1 \quad \text{et} \quad K_m^{(k)} = \sum_{j=2}^m \frac{K_j^{(k-1)}}{j}. \quad (5.2)$$

Les auteurs ont remarqué que ces nombres $K_m^{(k)}$, nommés *nombre hyperharmoniques* constituaient des variations des nombres harmoniques $H_r(m)$. En effet, pour $k \geq 1$, $K_m^{(k)}$ peut s'écrire

$$K_m^{(k)} = \underline{H}_{y_1^k}(m) - \underline{H}_{y_1^{k-1}}(m).$$

Démonstration. Par récurrence sur k : évident si $k = 1$ (rappelons que $\underline{H}_\epsilon \equiv 1$). Puis, supposons le résultat vrai pour $k > 1$, nous appliquons alors la proposition 4.1.8

$$\begin{aligned} K_m^{(k+1)} &= \sum_{j=2}^m \frac{K_j^{(k)}}{j} \\ &= \sum_{j=1}^m \frac{\underline{H}_{y_1^k}(j) - \underline{H}_{y_1^{k-1}}(j)}{j} \quad (\text{puisque } K_1^{(k)} = 0) \\ &= \underline{H}_{y_1^{k+1}}(m) - \underline{H}_{y_1^k}(m). \end{aligned}$$

□

Par conséquent, $\mathbb{E}[M_T^{(k)}]$ peut encore être simplifié en

$$\mathbb{E}[M_T^{(k)}] = \underline{H}_{y_1^k}(m).$$

Ainsi, lorsque m tend vers $+\infty$,

$$\begin{aligned} \mathbb{E}[M_T^{(2)}] &= \frac{1}{2} \ln^2(m) + \gamma \ln(m) + \frac{\gamma^2 + \zeta(2)}{2} + \frac{\ln(m) + \gamma - 1}{m} + \mathcal{O}\left(\frac{\ln(m)}{m^2}\right) \\ \mathbb{E}[M_T^{(3)}] &= \frac{1}{6} \ln^3(m) + \frac{1}{2} \ln^2(m) \gamma + \left(\frac{\gamma^2 + \zeta(2)}{2}\right) \ln(m) + \frac{\gamma \zeta(2)}{2} + \frac{\gamma^3}{6} + \frac{\zeta(3)}{3} + \\ &\quad \mathcal{O}\left(\frac{\ln^2(m)}{m}\right) \\ \mathbb{E}[M_T^{(4)}] &= \frac{1}{24} \ln^4(m) + \frac{1}{6} \ln^3(m) \gamma + \left(\frac{\gamma^2 + \zeta(2)}{4}\right) \ln^2(m) + \\ &\quad \left(\frac{\gamma \zeta(2)}{2} + \frac{\gamma^3}{6} + \frac{\zeta(3)}{3}\right) \ln(m) + \frac{1}{24} \gamma^4 + \frac{1}{3} \zeta(3) \gamma + \frac{9}{40} \zeta(2)^2 + \frac{\zeta(2) \gamma^2}{4} + \\ &\quad \mathcal{O}\left(\frac{\ln^3(m)}{m}\right). \end{aligned}$$

Au-delà de cet exemple introductif, l'objectif de ce chapitre est de présenter l'intérêt que peuvent revêtir les outils algébriques que nous avons introduits précédemment dans le contexte de données multidimensionnelles.

5.1 Aux arbres hyperquaternaires

5.1.1 Autour de la racine

5.1.1.1 Polynômes et base de Bernstein

Notation. Soit $S = \{w_1, \dots, w_s\}$ un ensemble de s mots de longueur d construits sur l'alphabet $\{0, 1\}$. Pour tous entiers $n_1, \dots, n_s \in \{0, \dots, n\}$ tels que

$$n_1 + \dots + n_s = n, \quad (5.3)$$

et pour tout entier $k \in \{1, \dots, d\}$, nous notons

$$N_k = n_1 i_k^1 + \dots + n_s i_k^s, \quad (5.4)$$

où i_k^j désigne la $k^{\text{ème}}$ lettre du mot binaire w_j de S . Il est immédiat que $0 \leq N_k \leq n$.

Nous rappelons que le problème posé par l'équation (1.9) réside dans le calcul de

$$J_n[S] = \int_0^1 \dots \int_0^1 f_S(t_1, \dots, t_d)^{n-1} dt_1 \dots dt_d, \quad \text{où } f_S(t_1, \dots, t_d) = \sum_{\epsilon \in S} \prod_{i=1}^d t_i^{\langle \epsilon_i \rangle},$$

avec les conventions $t_i^{\langle 0 \rangle} = t_i$ et $t_i^{\langle 1 \rangle} = 1 - t_i$. La fonction à intégrer peut donc s'écrire en toute généralité sous la forme :

$$f_S(t_1, \dots, t_d)^{n-1} = \sum_{i_1, \dots, i_d, j_1, \dots, j_d=0}^{n-1} A_{i_1, \dots, i_d}^{j_1, \dots, j_d} t_1^{i_1} (1 - t_1)^{j_1} \dots t_d^{i_d} (1 - t_d)^{j_d}.$$

Une base naturelle pour exprimer ces polynômes est la base de Bernstein.

Définition 5.1.1. Nous notons les polynômes de Bernstein de degré n par :

$$B_n^i(t) = \binom{n}{i} t^i (1 - t)^{n-i}.$$

L'ensemble des $\{B_n^i(t)\}_{0 \leq i \leq n}$ constitue la base de Bernstein de $\mathbb{R}_n[X]$, l'espace des polynômes de degré inférieur ou égal à n .

Par convention, pour $i < 0$ et $i > n$, $B_n^i \equiv 0$.

5.1.1.2 Calcul des intégrales $J_n[S]$, cas général

En introduisant les polynômes de Bernstein, f_S^n devient :

$$f_S^n(t_1, \dots, t_d) = \sum_{i_1, \dots, i_d=0}^n C_{i_1, \dots, i_d}^n B_n^{i_1}(t_1) B_n^{i_2}(t_2) \dots B_n^{i_d}(t_d), \quad (5.5)$$

où les coefficients C_{i_1, \dots, i_d}^n sont obtenus d'une part à partir des $A_{i_1, \dots, i_d}^{j_1, \dots, j_d}$ et d'autre part à partir des coefficients binômiaux pour l'écriture dans la base de Bernstein.

Pour simplifier le calcul, nous écrivons donc $f_S(t_1, \dots, t_d)$ dans la base de Bernstein de $\mathbb{R}_1[X]^{\otimes d}$:

$$f_S(t_1, \dots, t_d) = [B_1^{i_1^1}(t_1) \dots B_1^{i_d^1}(t_d)] + \dots + [B_1^{i_1^s}(t_1) \dots B_1^{i_d^s}(t_d)] \quad (5.6)$$

puis nous déduisons l'expression de $f_S^n(t_1, \dots, t_d)$ dans la base de Bernstein de $\mathbb{R}_n[X]^{\otimes d}$:

Lemme 5.1.2. Soit $S = \{i_1^1 \dots i_d^1, \dots, i_1^s \dots i_d^s\} \subset \{0, 1\}^d$. Nous avons :

$$f_S^n(t_1, \dots, t_d) = \sum_{\substack{n_1, \dots, n_s=0 \\ n_1 + \dots + n_s = n}}^n \frac{\binom{n}{n_1, \dots, n_s}}{\binom{n}{N_1} \dots \binom{n}{N_d}} B_n^{N_1}(t_1) \dots B_n^{N_d}(t_d).$$

Démonstration. En élevant à la puissance n le polynôme $f_S(t_1, \dots, t_d)$, nous obtenons successivement :

$$\begin{aligned} f_S^n(t_1, \dots, t_d) &= \sum_{\substack{n_1, \dots, n_s=0 \\ n_1 + \dots + n_s = n}}^n \binom{n}{n_1, \dots, n_s} [B_1^{i_1^1}(t_1) \dots B_1^{i_d^1}(t_d)]^{n_1} \dots [B_1^{i_1^s}(t_1) \dots B_1^{i_d^s}(t_d)]^{n_s} \\ &= \sum_{\substack{n_1, \dots, n_s=0 \\ n_1 + \dots + n_s = n}}^n \binom{n}{n_1, \dots, n_s} [B_{n_1}^{n_1 i_1^1}(t_1) \dots B_{n_1}^{n_1 i_d^1}(t_d)] \dots [B_{n_s}^{n_s i_1^s}(t_1) \dots B_{n_s}^{n_s i_d^s}(t_d)]. \end{aligned}$$

En utilisant le fait que pour tout $j_1 \in \{0, \dots, n_1\}, \dots, j_s \in \{0, \dots, n_s\}$,

$$B_{n_1}^{j_1}(t) \dots B_{n_s}^{j_s}(t) = \frac{\binom{n_1}{j_1} \dots \binom{n_s}{j_s}}{\binom{n_1 + \dots + n_s}{j_1 + \dots + j_s}} B_{n_1 + \dots + n_s}^{j_1 + \dots + j_s}(t), \quad (5.7)$$

nous déduisons alors l'expression désirée de $f_S^n(t_1, \dots, t_d)$ dans le cas particulier où chaque j_k est soit égal à 0 soit égal à n_k et où $n_1 + \dots + n_s = n$. \square

D'après le lemme 5.1.2, nous obtenons alors :

$$J_n[S] = \sum_{\substack{n_1, \dots, n_s=0 \\ n_1 + \dots + n_s = n-1}}^{n-1} \frac{\binom{n-1}{n_1, \dots, n_s}}{\binom{n-1}{N_1} \dots \binom{n-1}{N_d}} \int_0^1 \dots \int_0^1 B_{n-1}^{N_1}(t_1) \dots B_{n-1}^{N_d}(t_d) dt_1 \dots dt_d. \quad (5.8)$$

Proposition 5.1.3. [52] Soit $S = \{i_1^1 \dots i_d^1, \dots, i_1^s \dots i_d^s\} \subset \{0, 1\}^d$. Nous avons :

$$J_n[S] = \frac{1}{n^d} \sum_{\substack{n_1, \dots, n_s=0 \\ n_1 + \dots + n_s = n-1}}^{n-1} \frac{\binom{n-1}{n_1, \dots, n_s}}{\binom{n-1}{N_1} \dots \binom{n-1}{N_d}}.$$

Démonstration. En utilisant le fait que :

$$\int_0^1 B_{n-1}^i(t) dt = \frac{1}{n}, \quad i \in \{0, \dots, n-1\}, \quad (5.9)$$

pour tout $i_1, \dots, i_d \in \{1, \dots, n-1\}$, en appliquant le théorème de Fubini, nous déduisons le résultat suivant qui, d'après le lemme 5.1.2, donne le résultat désiré :

$$\int_0^1 \dots \int_0^1 B_{n-1}^{N_1}(t_1) \dots B_{n-1}^{N_d}(t_d) dt_1 \dots t_d = \frac{1}{n^d}. \quad (5.10)$$

□

5.1.1.3 Simplification du calcul des $J_n[S]$

Nous avons déjà remarqué que pour des ensembles S "grands", il était intéressant de s'intéresser au complémentaire \bar{S} , noté \bar{S} , puisque les quantités $J_n[S]$ et $J_n[\bar{S}]$ étaient reliées par la relation (1.10).

En dimension 3, dans le cas où $S = \{1, 2, 3, 4, 5, 6, 7\}$ (en identifiant le codage binaire et son équivalent décimal), alors $\bar{S} = \{0\}$, et

$$\begin{aligned} J_n[S] &= \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} \frac{1}{k^3} \\ &= -\frac{1}{n} \sum_{k=1}^n (-1)^k \binom{n}{k} \frac{1}{k^2}. \end{aligned}$$

Mais alors, nous sommes en face d'une tranformation d'Euler, et nous pouvons utiliser quelques propriétés reliant les opérateurs ∇ et Σ , définis par

$$\begin{aligned} \nabla : (a_n)_{n \geq 0} &\mapsto \left(\sum_{k=0}^n (-1)^k \binom{n}{k} a_k \right)_{n \geq 0}, \\ \Sigma : (a_n)_{n \geq 0} &\mapsto \left(\sum_{k=0}^n a_k \right)_{n \geq 0} \end{aligned}$$

En effet, Hoffman [33] a montré que

$$\Sigma \nabla \underline{H}_w = -\underline{H}_{w^\#}, \quad (5.11)$$

avec $w^\#$ associé bijectivement à w construit comme suit :

si $w = y_{s_1} \dots y_{s_r}$ a pour poids n , nous pouvons considérer les sommes partielles $\{s_1, s_1 + s_2, \dots, s_1 + \dots + s_{r-1}\}$ comme un (éventuellement vide) sous-ensemble de $\{1, \dots, n-1\}$, dont nous pouvons considérer l'ensemble complémentaire et construire à partir de là l'unique mot $w^\#$ de poids n , dont les "poids partiels" coïncident avec ce nouvel ensemble. Par exemple, $y_2^\# = y_1 y_1, y_1 y_3^\# = y_2 y_1^2$.

En revenant à $S = \{1, 2, 3, 4, 5, 6, 7\}$, et puisque $\Sigma \nabla = \nabla \Sigma^{-1}$, nous obtenons

$$J_n[S] = -\frac{1}{n} \nabla \Sigma^{-1} \underline{H}_{y_2}(n) = \frac{1}{n} \underline{H}_{y_2^\#}(n) = \frac{\underline{H}_{y_1^2}(n)}{n}.$$

Autre exemple, encore en dimension $d = 3$. Si $S = \{2, 3, 4, 5, 6, 7\}$, alors $\overline{S} = \{0, 1\}$, pour lequel $J_k[\overline{S}] = \frac{1}{k^2}$.

$$\begin{aligned} J_n[S] &= \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} \frac{1}{k^2} \\ &= -\frac{1}{n} \sum_{k=1}^n (-1)^k \binom{n}{k} \frac{1}{k} \\ &= -\frac{1}{n} \nabla \Sigma^{-1} \underline{H}_{y_1}(n) \\ &= \frac{1}{n} \underline{H}_{y_1^\#}(n) = \frac{\underline{H}_{y_1}(n)}{n}. \end{aligned}$$

Avec $d = 3$, et $S = \{3, 4, 5, 6, 7\}$, alors $\overline{S} = \{0, 1, 2\}$, pour lequel $J_k[\overline{S}] = \frac{\underline{H}_{y_1}(k)}{k^2}$.

$$\begin{aligned} J_n[S] &= \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} \frac{\underline{H}_{y_1}(k)}{k^2} \\ &= -\frac{1}{n} \sum_{k=1}^n (-1)^k \binom{n}{k} \frac{\underline{H}_{y_1}(k)}{k} \\ &= -\frac{1}{n} \nabla \Sigma^{-1} \underline{H}_{y_1^2}(n) \\ &= \frac{1}{n} \underline{H}_{y_1^2^\#}(n) = \frac{\underline{H}_{y_2}(n)}{n}. \end{aligned}$$

De fait, ces trois exemples se généralisent immédiatement à la dimension d . En effet,

en dimension d , on vérifie sans difficulté que

$$\begin{aligned} J_n[\{0\}] &= \frac{1}{n^d} \\ J_n[\{0, 1\}] &= \frac{1}{n^{d-1}} \\ J_n[\{0, 1, 2\}] &= \frac{H_{y_1}(n)}{n^{d-1}}. \end{aligned}$$

Par conséquent,

$$\begin{aligned} J_n[\{1, 2, \dots, 2^d - 1\}] &= \frac{1}{n} H_{y_{d-1}^\#}(n) = \frac{H_{y_1^{d-1}}(n)}{n} \\ J_n[\{2, \dots, 2^d - 1\}] &= \frac{1}{n} H_{y_{d-2}^\#}(n) = \frac{H_{y_1^{d-2}}(n)}{n} \\ J_n[\{3, \dots, 2^d - 1\}] &= \frac{1}{n} H_{y_{d-2} y_1^\#}(n) = \frac{H_{y_1^{d-3} y_2}(n)}{n}. \end{aligned}$$

En fait, comme $J_n[S] = \frac{1}{n^d}$ pour n'importe quel ensemble S réduit à un élément, la première de ces trois égalités reste vraie en remplaçant $\{1, 2, \dots, 2^d - 1\}$ par n'importe quel ensemble de cardinal $2^d - 1$.

5.1.2 Retour sur les paramètres additifs

5.1.2.1 Constantes universelles en dimension 2

Rappelons la récurrence fondamentale (1.8), propre aux arborescences hyperquaternaires en dimension 2, vérifiée par la moyenne des paramètres additifs

$$f_n = t_n + 4 \sum_{j=0}^{n-1} \pi_{n,j,2} f_j.$$

Si $t_n = \delta_{n,k}$, ce qui revient à interpréter f_n comme le nombre moyen de noeuds étant racines d'un sous-arbre de taille k , Labelle et Laforest [41] ont montré que f_n était asymptotiquement linéaire, avec une constante de linéarité explicite :

$$f_n \sim \lambda_k \cdot n \quad \text{avec} \quad \lambda_k = 3 - 18k + 6k(3k + 1)(\zeta(2) - 6H_2(k)).$$

De plus, cette constante admet une représentation intégrale

$$\lambda_k = \int_0^1 \Lambda(t) t^k dt \quad \text{avec} \quad \Lambda(t) = 6 \left(-\frac{5+t}{(1-t)^2} + \frac{2(1+2t)}{(1-t)^3} \log(t) \right).$$

Ces constantes λ_k sont nommées “constantes universelles”.

À partir de celles-ci, il est possible, par linéarité, de calculer, pour une propriété P donnée caractérisée par une suite $(t_n)_n$, la constante $\lambda(P)$, telle que $f_n \sim \lambda(P).n$, f_n étant interprété comme le nombre moyen de noeuds ayant la propriété P . En effet, [43]

$$\lambda(P) = \sum_{k=1}^{+\infty} t_k \lambda_k = \int_0^1 \Lambda(z) t(z) dz. \quad (5.12)$$

Exemple 5.1.4. – Pour le nombre de feuilles, $t_n = \delta_{n,1}$ d'où

$$\lambda(P) = \lambda_1 = 24\zeta(2) - 39.$$

– Pour le nombre de noeuds à 1 enfant, $t_n = \frac{4}{n^2}$, si $n \geq 2$ (cf Tableau 1.1.3), d'où

$$\lambda(P) = 24\zeta(3) - 156\zeta(2) + 228.$$

– Pour le nombre de noeuds à 2 enfants, $t_n = \frac{4}{n} - \frac{12}{n^2} + \frac{2}{n} \sum_{i=0}^{n-1} \frac{1}{2^i(n-i)}$, si $n \geq 3$.

Nous avons vu à l'exemple 1.2.1 que $\sum_{n \geq 1} \sum_{i=0}^{n-1} \frac{1}{2^i(n-i)} z^n = \frac{2}{2-z} \text{Li}_1(z)$, donc en

intégrant par rapport à $\frac{dz}{z}$, il vient

$$\sum_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{2^i(n-i)} z^n = \text{Li}_2(z) + \text{Li}_2(z-1) - \text{Li}_1(z) \log(2-z).$$

La série génératrice de cette suite particulière $(t_n)_n$, se trouve être

$$t(z) = 4 \text{Li}_1(z) - 10 \text{Li}_2(z) + 2 \text{Li}_2(z-1) - 2 \text{Li}_1(z) \log(2-z) + 6z + \zeta(2),$$

les deux derniers termes compensant simplement le fait que $t_0 = t_1 = t_2 = 0$.

Par des techniques de changement de variables et de développement en série entière, Labelle [40] a montré alors que

$$\lambda(P) = -336 + 201\zeta(2) + 24\zeta(3) + 96\zeta(2) \log(2) - 96 \sum_{k \geq 1} \frac{H_1(k)}{k^2 2^k} - 96 \sum_{k \geq 1} \frac{H_2(k)}{k^2 2^k}.$$

Mais alors, d'après les principes évoqués en section 4.2.1.2, nous pouvons affirmer que

$$\begin{aligned} \lambda(P) &= -336 + 201\zeta(2) + 24\zeta(3) + 96\zeta(2) \log(2) - 96 \left(\text{Li}_{2,1} \left(\frac{1}{2} \right) + \text{Li}_3 \left(\frac{1}{2} \right) \right) \\ &\quad - 96 \left(\text{Li}_{1,2} \left(\frac{1}{2} \right) + \text{Li}_3 \left(\frac{1}{2} \right) \right). \end{aligned}$$

Comme de plus, d'après l'équation (3.8),

$$\text{Li}_{2,1}(1-t) = -\text{Li}_3(t) + \log(t) \text{Li}_2(t) - \frac{1}{2} \log^2(t) \text{Li}_1(t) + \zeta(3),$$

et comme $\text{Li}_3\left(\frac{1}{2}\right) = \frac{7}{8}\zeta(3) - \frac{1}{2}\zeta(2) \log(2) + \frac{1}{6} \ln^3(2)$, nous en déduisons que

$$\text{Li}_{2,1}\left(\frac{1}{2}\right) = \frac{1}{8}\zeta(3) - \frac{1}{6} \log^3(2).$$

Enfin, comme $\text{Li}_{1,2} = \text{Li}_{x_1 x_0 x_1} = \text{Li}_{x_1} \text{Li}_{x_0 x_1} - 2 \text{Li}_{x_0 x_1^2}$, nous avons également

$$\text{Li}_{1,2}\left(\frac{1}{2}\right) = \frac{1}{2}\zeta(2) \log(2) - \frac{1}{6} \log^3(2) - \frac{1}{4}\zeta(3).$$

Finalement,

$$\lambda(P) = -336 + 201\zeta(2) + 144\zeta(2) \log(2) - 132\zeta(3).$$

5.1.2.2 Nombre de feuilles

La récurrence fondamentale devient, dans le cas où le paramètre étudié est le nombre de noeuds à k enfants :

$$f_{n,d} = p_{n,k,d} + 2^d \sum_{j=0}^{n-1} \pi_{n,j,d} f_{j,d},$$

$p_{n,k,d}$ désignant la probabilité que la racine d'un arbre hyperquaternaire de taille n soit d'arité k , probabilité déjà donnée par la proposition 1.1.5 page 11.

Pour $k = 0$, la racine n'étant une feuille que si $n = 1$, il est immédiat que $p_{n,0,d} = \delta_{n,1}$.

Par ailleurs, nous avons déjà vu en (1.15) l'équation différentielle vérifiée par la transformée d'Euler f^* . En introduisant l'opérateur K défini par

$$K(f)(z) = \int_0^z f(t) \frac{dt}{t},$$

et qui n'est autre que l'opérateur $J - I$, I et J étant les opérateurs introduits en section 1.2.2.

$$\begin{aligned} \left(z \frac{d}{dz}\right)^d \left((1-z)(f^*(z) - t^*(z)) \right) &= -2^d z f^*(z) \\ (1-z)(f^*(z) - t^*(z)) &= -(2K)^d (z f^*) \\ z(f^*(z) - t^*(z)) &= \left(-\frac{z}{1-z}\right) (2K)^d (z f^*) \end{aligned}$$

d'où $\left(Id + \frac{z}{1-z} (2K)^d \right) \left(z f^*(z) \right) = z t^*(z).$

Par conséquent,

$$zf^*(z) = \sum_{j \geq 0} \left(-\frac{z}{1-z} (2K)^d \right)^j (Id.t^*)(z).$$

Appliquons ceci à la fonction t^* , relative à l'étude du nombre de feuilles, i.e.

$$t^*(z) = -\frac{z}{(1-z)^2}.$$

Alors

$$K^2(Id.t^*)(z) = Li_2(z) - Li_1(z),$$

et de plus, pour tout mot w , $K(Li_w) = Li_{x_0 w}$.

Ainsi, pour $d = 3$,

$$\begin{aligned} \left(-\frac{z}{1-z} (2K)^3 \right) (Id.t^*)(z) &= -\frac{8z}{1-z} \left(Li_{x_0^2 x_1}(z) - Li_{x_0 x_1}(z) \right) \\ \left(-\frac{z}{1-z} (2K)^3 \right)^2 (Id.t^*)(z) &= \frac{8^2 z}{1-z} \left(Li_{x_0^2 x_1 x_0^2 x_1}(z) - Li_{x_0^2 x_1 x_0 x_1}(z) \right) \\ &\dots \end{aligned}$$

Par conséquent,

$$\begin{aligned} f^*(z) &= \sum_{k \geq 1} (-8)^k \left(P_{y_3^k}(z) - P_{y_3^{k-1} y_2}(z) \right) - \frac{z}{(1-z)^2} \\ f_n^* &= \sum_{k=1}^n (-8)^k \left(H_{y_3^k}(n) - H_{y_3^{k-1} y_2}(n) \right) \\ f_n &= \sum_{l, k=1}^n (-1)^l \binom{n}{l} (-8)^k \left(H_{y_3^k}(l) - H_{y_3^{k-1} y_2}(l) \right). \end{aligned}$$

Par un calcul parfaitement analogue, en dimension d quelconque,

$$f^*(z) = \sum_{k \geq 1} (-2^d)^k \left(P_{y_d^k}(z) - P_{y_d^{k-1} y_{d-1}}(z) \right) - \frac{z}{(1-z)^2} \quad (5.13)$$

$$f_n^* = \sum_{k=1}^n (-2^d)^k \left(H_{y_d^k}(n) - H_{y_d^{k-1} y_{d-1}}(n) \right) \quad (5.14)$$

$$f_n = \sum_{l, k=1}^n (-1)^l \binom{n}{l} (-2^d)^k \left(H_{y_d^k}(l) - H_{y_d^{k-1} y_{d-1}}(l) \right). \quad (5.15)$$

5.1.2.3 Noeuds à 1 fils

Si on s'intéresse maintenant au nombre de noeuds à 1 fils, la suite-test correspondante est $t_n = p_{n,1,d}$ i.e. la probabilité pour la racine d'être d'arité 1, probabilité donnée par la proposition 1.1.5 page 11,

$$p_{n,1,d} = \sum_{j=0}^1 (-1)^{1-j} \binom{2^d - j}{1 - j} \sum_{|S|=j} J_n[S] = \frac{2^d}{n^d}, \text{ si } n \geq 2.$$

La série génératrice correspondante est donc $t(z) = 2^d \text{Li}_d(z) - 2^d z$.

Par conséquent,

$$t^*(z) = \frac{2^d}{1-z} \text{Li}_d\left(\frac{z}{z-1}\right) + 2^d \frac{z}{(1-z)^2}.$$

Une approche possible pourrait consister à utiliser des relations fonctionnelles similaires à (3.8), et décrites dans [61], pour les transformations $z \mapsto \frac{1}{1-z}$ et $z \mapsto \frac{1}{z}$. En fait, il est ici plus rapide de travailler directement sur les coefficients, et d'utiliser la relation (5.11). En effet,

$$\begin{aligned} \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{1}{k^d} &= \nabla \Sigma^{-1} \underline{H}_{y_d}(n) \\ &= -\underline{H}_{y_d^*}(n) \\ &= -\underline{H}_{y_1^d}(n). \end{aligned}$$

L'expression $\underline{H}_{y_1^d}$ peut alors être décomposée en une somme de H_w , cf relation (4.10), et par suite, $t^*(z)$ s'exprime comme une combinaison linéaire de fonctions P_w .

Exemple 5.1.5. Pour $d = 3$, $-\underline{H}_{y_1^3}(n) = -\left(H_{y_1^3} + H_{y_1 y_2} + H_{y_2 y_1} + H_{y_3}\right)$, donc

$$t^*(z) = -8 \left(P_{y_1^3}(z) + P_{y_1 y_2}(z) + P_{y_2 y_1}(z) + P_{y_3}(z) \right) + \frac{8z}{(1-z)^2}.$$

Ainsi, en notant $w_3 = y_1^3 + y_1 y_2 + y_2 y_1 + y_3$, et en reprenant les résultats de la section précédente,

$$\begin{aligned} f^*(z) &= -8P_{w_3}(z) + \sum_{k \geq 1} (-8)^{k+1} P_{y_3^k w_3}(z) - \sum_{k \geq 1} (-8)^{k+1} \left(P_{y_3^k}(z) - P_{y_3^{k-1} y_2}(z) \right) + \frac{8z}{(1-z)^2} \\ f_n^* &= -8H_{w_3}(n) + \sum_{k=1}^n (-8)^{k+1} H_{y_3^k w_3}(n) - \sum_{k=1}^n (-8)^{k+1} \left(H_{y_3^k}(n) - H_{y_3^{k-1} y_2}(n) \right) \\ f_n &= -\frac{8}{n^3} + \sum_{l,k=1}^n (-1)^l \binom{n}{l} (-8)^{k+1} \left(H_{y_3^k w_3}(l) - H_{y_3^k}(l) + H_{y_3^{k-1} y_2}(l) \right). \end{aligned}$$

La généralisation de ce calcul en dimension d conduit à

$$f_n = -\frac{2^d}{n^d} + \sum_{l,k=1}^n (-1)^l \binom{n}{l} (-2^d)^{k+1} \left(H_{y_d^k w_d}(l) - H_{y_d^k}(l) + H_{y_d^{k-1} y_{d-1}}(l) \right),$$

où w_d est le polynôme tel que $\underline{H}_{y_d} = H_{w_d}$.

5.2 Aux maxima pour des données multidimensionnelles

Nous présentons dans cette deuxième section une approche combinatoire pour le calcul de la variance du nombre de maxima dans un hypercube. Celle-ci mène à une expression explicite, en terme de polyzêtas, du terme dominant dans le développement asymptotique de cette variance. De plus, nous obtenons un algorithme pour calculer ce développement, et prouvons que les coefficients intervenant dans ce développement appartiennent à la $\mathbb{Q}[\gamma]$ -algèbre engendrée par \mathcal{Z} .

5.2.1 Introduction

Soit $\mathcal{X} = \{x_1, \dots, x_n\}$ un ensemble de vecteurs aléatoires indépendants et identiquement distribués dans \mathbb{R}^d . Un point $x_i = (x_{i_1}, \dots, x_{i_d})$ est dit dominé par $x_j = (x_{j_1}, \dots, x_{j_d})$ si $x_{i_k} < x_{j_k}$ pour tout $k \in \{1, \dots, d\}$ et un point x_i est appelé un maximum de \mathcal{X} si aucun des autres points ne le domine. Le nombre de maxima de \mathcal{X} est noté $K_{n,d}$.

Bien que l'étude de cette variable aléatoire se distingue de l'étude des quadrees tant par la modélisation du problème que par ses applications, Labelle, Laforest et Provençal ont déjà noté dans [44] la similarité de résultats intervenant dans les deux domaines. Un lien existant entre ces deux structures, mais qui reste à exploiter, est celui de la racine de l'arbre, qui se trouve être un maximum si et seulement si son deuxième sous-arbre ("nord-est" si l'on travaille en dimension 2) est vide. À partir de là, en posant $G_i = \mathbb{1}_{\{x_i \text{ est un maximum}\}}$, le nombre de maxima peut être vue comme la variable

$$\sum_{i=1}^n G_i = \sum_{i=1}^n \mathbb{1}_{\{|T_2^{\sigma_i}|=0\}},$$

$T_2^{\sigma_i}$ désignant le deuxième sous-arbre issu de la racine, pour l'arbre hyperquaternaire obtenu en permutant les points x_1 et x_i . Les variables $\mathbb{1}_{\{|T_2^{\sigma_i}|=0\}}$ suivent alors des lois de Bernoulli de paramètre $\frac{1}{n} \underline{H}_{y_1^{d-1}}(n)$, mais sont bien entendus des variables dépendantes.

De nombreux articles ont déjà été consacrés à l'étude du nombre de maxima au sein d'un ensemble de points, étant donné que ce problème apparaît dans différents domaines, par exemple la détermination de l'ensemble de Pareto de \mathcal{X} , sous l'hypothèse

d'une distribution identique et continue des composantes des vecteurs [68]. De nombreuses applications précises sont fournies par Bai et al. dans [3], article dans lequel les auteurs prouvent la normalité asymptotique de $K_{n,2}$, lorsque les variables aléatoires $(x_i)_{1 \leq i \leq n}$ sont i.i.d. dans un polygone convexe. Plus récemment, dans [2], Bai et al. parviennent à obtenir une borne de Berry-Esseen pour le nombre de maxima dans un n -échantillon aléatoire dans $[0, 1]^d$, et proposent une méthode pour le calcul du développement asymptotique de la variance.

Cette étude de $\text{Var}(K_{n,d})$ dans le cas d'échantillons aléatoires dans $[0, 1]^d$ constitue précisément l'objectif de cette section. Pour cela, nous allons exploiter un résultat important, mis en premier en évidence par Ivanin [35] :

$$\mathbb{E}(K_{n,d}^2) = \mu_{n,d} + \sum_{1 \leq t \leq d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \sum_{(*)} \frac{1}{i_1 \dots i_{d-2} j_1 \dots j_{d-1}}, \quad (5.16)$$

où la somme $(*)$ est prise sur tous les indices vérifiant

$$1 \leq i_1 \dots \leq i_{t-1} \leq l, 1 \leq i_t \leq \dots \leq i_{d-2} \leq l$$

et

$$l+1 \leq j_1 \leq \dots \leq j_{d-1} \leq n.$$

Dans la formula (5.16), $\mu_{n,d}$ représente l'espérance de la variable $K_{n,d}$, calculée pour la première fois par Barndorff-Nielsen et Sobel [4] :

$$\mu_{n,d} = \mathbb{E}(K_{n,d}) = \sum_{1 \leq i_1 \leq \dots \leq i_{d-1} \leq n} \frac{1}{i_1 \dots i_{d-1}}. \quad (5.17)$$

Remarquons que cette espérance se déduit également immédiatement de notre remarque introductive concernant le fait que X_i était un maximum avec probabilité $\frac{1}{n} \underline{H}_{y_1}^{d-1}(n)$.

Cette formule (5.16) repose sur le calcul suivant,

$$\begin{aligned} \mathbb{E}(K_{n,d}^2) &= \mathbb{E} \left(\sum_{i=1}^n G_i \right)^2 \\ &= \mathbb{E}(K_{n,d}) + n(n-1) \mathbb{E}(G_1 G_2) \\ &= \mu_{n,d} + n(n-1) \sum_{t=1}^{d-1} \int_{(t)} \left(1 - \prod_{i=1}^d (1-x_i) - \prod_{i=1}^d (1-y_i) + \prod_{i=1}^t (1-x_i) \prod_{i=t+1}^d (1-y_i) \right)^{n-2} dx dy, \end{aligned}$$

avec $(t) = \{(\mathbf{x}, \mathbf{y}) / x_1 > y_1, \dots, x_t > y_t, x_{t+1} < y_{t+1}, \dots, x_d < y_d\}$.

Calculant cette intégrale par des manipulations algébriques, Bai et al. retrouvent la formule (5.16), et en déduisent, par des arguments essentiellement combinatoires, le résultat principal de [1], à savoir l'équivalent

$$\text{Var}(K_{n,d}) \sim \left(\frac{1}{(d-1)!} + \kappa_d \right) \ln^{d-1}(n),$$

avec

$$\kappa_d = \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \sum_{(**)} \frac{1}{i_1 \dots i_{t-1} j_1 \dots j_{d-2-t}}, \quad (5.18)$$

la somme $(**)$ étant calculée sur tous les indices vérifiant

$$1 \leq i_1 \leq \dots \leq i_{t-1} \leq l \quad \text{et} \quad 1 \leq j_1 \leq \dots \leq j_{d-2-t} \leq l.$$

5.2.2 Équivalent asymptotique de $\text{Var}(K_{n,d})$

Nous nous concentrons pour l'instant sur l'équivalent asymptotique de $\text{Var}(K_{n,d})$

$$\text{Var}(K_{n,d}) \sim \left(\frac{1}{(d-1)!} + \kappa_d \right) \ln^{d-1}(n),$$

κ_d donné par la formule (5.18). Celle-ci peut être ré-écrite (cf exemple 4.1.6)

$$\begin{aligned} \kappa_d &= \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \underline{\mathbb{H}}_{y_1^{t-1}}(l) \underline{\mathbb{H}}_{y_1^{d-2-t}}(l) \\ &= \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \underline{\mathbb{H}}_{y_1^{t-1} \sqcup y_1^{d-2-t}}(l) \\ &= \frac{1}{(d-1)!} \sum_{t=1}^{d-2} \binom{d-1}{t} \sum_{l \geq 1} \frac{1}{l^2} \underline{\mathbb{H}}_{y_1^{t-1} \sqcup y_1^{d-2-t}}(l) \end{aligned}$$

Par abus de notation, nous noterons $|w|_2$ plutôt que $|w|_{y_2}$ le nombre d'occurrences de la lettre y_2 dans le mot w . Nous introduisons de plus une dernière notation *ad hoc*, à savoir $\{y_1, y_2\}_{a+b}$ désignant l'ensemble des mots composés uniquement des lettres y_1 et y_2 , et de poids $a+b$.

Proposition 5.2.1.

$$y_1^a \sqcup y_1^b = \sum_{w \in \{y_1, y_2\}_{a+b}} (-1)^{|w|_2} \binom{a+b-2|w|_2}{a-|w|_2} w$$

Démonstration. Par récurrence sur $a+b$: le résultat est immédiat si $a+b = 1$. Supposons-le acquis pour $a+b = K$. Supposons également $b \geq 1$. Alors,

$$\begin{aligned}
y_1^{a+1} \sqcup y_1^b &= y_1(y_1^a \sqcup y_1^b) + y_1(y_1^{a+1} \sqcup y_1^{b-1}) - y_2(y_1^a \sqcup y_1^{b-1}) \\
&= \sum_{w \in \{y_1, y_2\}_{a+b}} (-1)^{|w|_2} \binom{a+b-2|w|_2}{a-|w|_2} y_1 w + \sum_{w \in \{y_1, y_2\}_{a+b}} (-1)^{|w|_2} \binom{a+b-2|w|_2}{a+1-|w|_2} y_1 w - \\
&\quad \sum_{w \in \{y_1, y_2\}_{a+b-1}} (-1)^{|w|_2} \binom{a+b-1-2|w|_2}{a-|w|_2} y_2 w \\
&= \sum_{w \in \{y_1, y_2\}_{a+b}} (-1)^{|w|_2} \binom{a+b+1-2|w|_2}{a+1-|w|_2} y_1 w - \sum_{w \in \{y_1, y_2\}_{a+b-1}} (-1)^{|w|_2} \binom{a+b-1-2|w|_2}{a-|w|_2} y_2 w \\
&= \sum_{w \in \{y_1, y_2\}_{a+b+1}} (-1)^{|w|_2} \binom{a+b+1-2|w|_2}{a+1-|w|_2} w.
\end{aligned}$$

□

En considérant le cas $a = t - 1$ et $b = d - 2 - t$ et en utilisant le lemme suivant,

Lemme 5.2.2.

$$\sum_{t=1}^{d-2} \binom{d-1}{t} \binom{d-3-2k}{t-1-k} = \binom{2(d-2-k)}{d-2-k}, \quad \text{si } k \leq \left\lfloor \frac{d-3}{2} \right\rfloor$$

nous pouvons maintenant simplifier la somme sur t comme suit

Theorème 5.2.3 ([12, 14]).

$$\kappa_d = \frac{1}{(d-1)!} \sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{|w|_2} \binom{2(d-2-|w|_2)}{d-2-|w|_2} \underline{\zeta}(y_2 w).$$

Démonstration.

$$\begin{aligned}
\kappa_d &= \frac{1}{(d-1)!} \sum_{t=1}^{d-2} \binom{d-1}{t} \sum_{l \geq 1} \frac{1}{l^2} \underline{\mathbb{H}}_{y_1^{t-1} \sqcup y_1^{d-2-t}}(l) \\
&= \frac{1}{(d-1)!} \sum_{t=1}^{d-2} \binom{d-1}{t} \sum_{l \geq 1} \frac{1}{l^2} \sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{|w|_2} \binom{d-3-2|w|_2}{t-1-|w|_2} \underline{\mathbb{H}}_w(l) \\
&= \frac{1}{(d-1)!} \sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{|w|_2} \sum_{t=1}^{d-2} \binom{d-1}{t} \binom{d-3-2|w|_2}{t-1-|w|_2} \sum_{l \geq 1} \frac{1}{l^2} \underline{\mathbb{H}}_w(l) \\
&= \frac{1}{(d-1)!} \sum_{w \in \{y_1, y_2\}_{d-3}} (-1)^{|w|_2} \binom{2(d-2-|w|_2)}{d-2-|w|_2} \underline{\zeta}(y_2 w).
\end{aligned}$$

□

Par exemple, pour $d = 7$, nous obtenons

$$6!\kappa_7 = \binom{10}{5} \underline{\zeta}(2, 1, 1, 1, 1) - \binom{8}{4} \left(\underline{\zeta}(2, 2, 1, 1) + \underline{\zeta}(2, 1, 2, 1) + \underline{\zeta}(2, 1, 1, 2) \right) + \binom{6}{3} \underline{\zeta}(2, 2, 2).$$

La dernière étape consiste à réduire en polyzêtas irréductibles, en appliquant la formule (4.10), par exemple

$$\underline{\zeta}(2, 1, 1, 2) = \zeta(2, 1, 1, 2) + \zeta(2, 1, 3) + \zeta(2, 2, 2) + \zeta(2, 4) + \zeta(3, 1, 2) + \zeta(3, 3) + \zeta(4, 2) + \zeta(6),$$

puis en utilisant la table de réduction.

Remarque 5.2.4. Les valeurs de κ_d sont données en appendice, pour d compris entre 2 et 13. Ces constantes ont déjà été données par Bai et al. [2] pour $d \leq 8$, et pour $d > 8$, nous faisons apparaître des polyzêtas irréductibles, i.e. ne pouvant s'exprimer à l'aide de la fonction zêta de Riemann, tout du moins si nous admettons la conjecture de Zagier.

5.2.3 Détermination du développement complet

Revenons à l'expression (5.16), que nous interprétons en terme de sommes harmoniques

$$\mathbb{E}(K_{n,d}^2) = \underline{H}_{y_1^{d-1}}(n) + \sum_{1 \leq t \leq d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \underline{H}_{y_1^{t-1}}(l) \underline{H}_{y_1^{d-t-1}}(l) \underline{H}_{y_1^{d-1}}(n; l+1),$$

Remarque 5.2.5. Pour traiter cette expression, nous nous permettons encore un abus de notation avec l'écriture $\underline{H}_w(n; l+1)$ qui correspond à la même définition que $\underline{H}_w(n)$, mais avec des indices tous supérieurs ou égaux à $l+1$.

Remarquons dès lors que la somme normalement calculée entre $l = 1$ et $l = n - 1$ peut être calculée entre $l = 1$ et $l = n$, puisque pour $l = n$ le terme ajouté est nul. Ainsi,

$$\mathbb{E}(K_{n,d}^2) = \underline{H}_{y_1^{d-1}}(n) + \sum_{1 \leq t \leq d-1} \binom{d}{t} \sum_{l=1}^n \frac{1}{l} \underline{H}_{y_1^{t-1}}(l) \underline{H}_{y_1^{d-t-1}}(l) \underline{H}_{y_1^{d-1}}(n; l+1).$$

– Dans un premier temps, nous transformons $\underline{H}_{y_1^{d-1}}(n; l+1)$ en un polynôme en $\underline{H}_r(n)$ et en $\underline{H}_r(l)$, pour $r = 1, \dots, d-1$. En effet,

Proposition 5.2.6. Pour tous entiers $n \geq l$,

$$\underline{H}_{y_1^d}(n; l) = \sum_{\substack{k_1 + \dots + k_d = d \\ k_1, \dots, k_d > 0}} \frac{\underline{H}_1^{k_1}(n; l) \dots \underline{H}_d^{k_d}(n; l)}{1^{k_1} k_1! \dots d^{k_d} k_d!}$$

Démonstration. C'est une conséquence directe de l'égalité (4.9). \square

Exemple 5.2.7.

$$\begin{aligned}\underline{H}_{y_1^2}(n; l+1) &= \frac{\underline{H}_1^2(n; l+1) + \underline{H}_2(n; l+1)}{2} \\ &= \frac{(\underline{H}_1(n) - \underline{H}_1(l))^2 + (\underline{H}_2(n) - \underline{H}_2(l))}{2}\end{aligned}$$

- Grâce à la proposition 4.1.7, nous sommes capables de transformer chaque polynôme (en des sommes harmoniques) en une combinaison linéaire de sommes harmoniques.

Exemple 5.2.8.

$$\begin{aligned}\mathbb{E}(K_{n,3}^2) &= \mu_{n,3} + \sum_{l=1}^n \frac{6}{l} \underline{H}_1(l) \left(\frac{(\underline{H}_1(n) - \underline{H}_1(l))^2 + (\underline{H}_2(n) - \underline{H}_2(l))}{2} \right) \\ &= \mu_{n,3} + 3 \left(\underline{H}_1^2(n) \sum_{l=1}^n \frac{\underline{H}_1(l)}{l} - 2\underline{H}_1(n) \sum_{l=1}^n \frac{\underline{H}_1^2(l)}{l} + \sum_{l=1}^n \frac{\underline{H}_1^3(l)}{l} + \right. \\ &\quad \left. \underline{H}_2(n) \sum_{l=1}^n \frac{\underline{H}_1(l)}{l} - \sum_{l=1}^n \frac{\underline{H}_1(l)\underline{H}_2(l)}{l} \right) \\ &= \mu_{n,3} + 3 \left(\underline{H}_1^2(n) \sum_{l=1}^n \frac{\underline{H}_1(l)}{l} - 2\underline{H}_1(n) \sum_{l=1}^n \frac{2\underline{H}_{y_1^2}(l) - \underline{H}_2(l)}{l} + \right. \\ &\quad \left. \sum_{l=1}^n \frac{6\underline{H}_{y_1^3}(l) - 3\underline{H}_{y_1 y_2}(l) - 3\underline{H}_{y_2 y_1}(l) + \underline{H}_3(l)}{l} + \underline{H}_2(n) \sum_{l=1}^n \frac{\underline{H}_1(l)}{l} - \right. \\ &\quad \left. \sum_{l=1}^n \frac{\underline{H}_{y_1 y_2}(l) + \underline{H}_{y_2 y_1}(l) - \underline{H}_3(l)}{l} \right)\end{aligned}$$

- Finalement, il ne reste plus que des sommes sur l du type $\sum_{l=1}^n \frac{\underline{H}_w(l)}{l}$, mais la proposition 4.1.8 nous permet de les réduire simplement à $\underline{H}_{y_1 w}(n)$.

Exemple 5.2.9.

$$\begin{aligned}\mathbb{E}(K_{n,3}^2) &= 6\underline{H}_{y_1^4}(n) - 6\underline{H}_{y_1 y_2 y_1}(n) + \underline{H}_{y_1^2}(n), \quad \text{d'où} \\ \text{Var}(K_{n,3}) &= \mathbb{E}(K_{n,3}^2) - \mu_{n,3}^2 \\ &= -\underline{H}_{y_2^2}(n) + 2\underline{H}_{y_1^2 y_2}(n) - 4\underline{H}_{y_1 y_2 y_1}(n) + 2\underline{H}_{y_2 y_1^2}(n) + \underline{H}_{y_1^2}(n).\end{aligned}$$

Remarquons que ce résultat précise celui donné par [7], sous une forme ne faisant pas intervenir de terme de reste.

Theorème 5.2.10 ([15]). *Pour tout $d \geq 2$, il existe des entiers $(\alpha_i)_{i \geq 1}$ nuls à partir d'un certain rang, et des mots $w_i \in Y^*$ tels que*

$$\mathbb{V}ar(K_{n,d}) = \sum_{i \geq 1} \alpha_i \underline{H}_{w_i}(n).$$

Conjecture. *Pour tout $i, w_i \in \{y_1, y_2\}_{2d-2}$.*

- En utilisant finalement les algorithmes décrits au chapitre 4, nous pouvons calculer le développement asymptotique de $\mathbb{V}ar(K_{n,d})$.

Theorème 5.2.11 ([12, 14]). *Il existe des coefficients effectivement calculables $\alpha_i, \beta_{j,k} \in \mathbb{Q}[\gamma, \mathcal{Z}]$ tels que, pour toute dimension d et tout ordre M ,*

$$\mathbb{V}ar(K_{n,d}) = \sum_{i=0}^{2d-2} \alpha_i \ln^i(n) + \sum_{j=1}^M \frac{1}{n^j} \sum_{k=0}^{2d-2} \beta_{j,k} \ln^k(n) + \mathcal{O}\left(\frac{1}{n^M}\right).$$

Des exemples de développement à l'ordre 0, pour $d \in \{3, 4, 5, 6\}$ sont donnés dans l'annexe C.

Conclusions et perspectives

Les travaux réalisés dans ce mémoire prolongent des méthodes combinatoires ayant déjà prouvé leur efficacité pour l'étude des polylogarithmes (indépendance linéaire [60], la détermination du groupe de Galois différentiel [53]), pour la résolution de systèmes dynamiques [57, 56], pour déterminer des relations entre polyzêtas [5], pour prolonger les fonctions polyzêtas de Hurwitz comme fonctions méromorphes à plusieurs variables [20]. Dans un autre domaine, l'utilisation de représentations symboliques, pour coder différents niveaux d'expressions de gènes, ont permis de proposer des solutions alternatives aux méthodes statistiques existantes pour les problèmes de biclustering en génomique [69].

Les techniques algébriques présentées au chapitre 2, notamment l'existence et la construction d'une base particulière, celle de Poincaré-Birkhoff-Witt-Lyndon, ont permis, au-delà de l'aspect théorique, une implantation des différents algorithmes proposés à la section 4.2, de façon pratique sous Maple 9.5, qu'il s'agisse du développement asymptotique d'une fonction appartenant à l'algèbre LI_C au voisinage de $z = 1$ ou du développement d'une somme harmonique $H_w(n)$ au voisinage de $+\infty$. Une fois ces procédures implémentées, les calculs relatifs aux applications du chapitre 5 ont pu être automatisés, comme les annexes permettent de le constater. En particulier, la formulation (4.27) des constantes d'Euler généralisées ont grandement amélioré nos algorithmes proposés dans [10] ou [11]. Cette formulation doit beaucoup à notre théorème "à l'Abel" 4.1.22, résultant lui-même de l'étude de la relation entre les séries génératrices non commutatives $L(z)$ et $L(1-z)$, relation projetée de l'alphabet X sur l'alphabet Y . Des techniques absolument similaires ont déjà permis par ailleurs d'obtenir la partie constante pour des séries génératrices (commutatives) de polyzêtas de Hurwitz [13].

Toutefois, les handicaps liés à Maple sont connus, ont déjà été soulevés dans la thèse de M. Bigotte [5], et encore maintenant il est par exemple difficile d'obtenir la série $\sigma(L)Z$ (cf équation (3.8)) au-delà du poids 10. Un codage spécifique à nos calculs en variables non commutatives, en langage C par exemple, devrait améliorer ces performances.

Concernant les constantes d'arbre de fouille, plusieurs questions restent en suspens. La réduction des intégrales $J_n[S]$ pour des ensembles S grands est automatisée par la transformation d'Euler et la relation (5.11). Mais il reste à déterminer, au départ, le cri-

tère de réduction d'une intégrale $J_n[S]$, a priori sous forme de combinaisons $\binom{n}{k}$, à une expression plus simple du type somme harmonique. Par ailleurs, la formule (5.15) nous présente f_n , l'espérance du nombre de feuilles dans un arbre à n noeuds, comme une combinaison de sommes harmoniques multiples. En fait, la conversion des sommes H_w en sommes de \underline{H}_w puis l'application de la transformation d'Euler (5.11), nous permet de réduire encore cette expression, mais difficilement sous forme close (en tout cas, pour une dimension d quelconque). La question du calcul explicite des constantes de linéarité asymptotique pour le nombre de feuilles dans un arbre hyperquaternaire aléatoire de points reste donc ouverte.

Par ailleurs, concernant les problèmes de maxima dans un ensemble de données multidimensionnelles, l'article de Barndorff-Nielsen et Sobel de 1966 [4] proposait des méthodes de calcul non simplement pour les moments d'ordre 1 et 2, mais pour la distribution complète de la variable $K_{n,d}$ selon deux hypothèses : d petit et n quelconque, et n petit, d quelconque. Récemment, les méthodes d'approximations de la variable K_n par une variable \overline{K}_n "proche" (pour la distance de Kolmogorov) appliquées dans [2] ont permis de trouver une borne de Berry-Esseen pour ce nombre de maxima (de l'ordre de la racine de l'écart-type). La forme des expressions pour les probabilités discrètes données dans [4] laisse à penser que ces quantités, et avec elles les moments de tout ordre de $K_{n,d}$, pourraient s'exprimer de manière symbolique comme combinaison de sommes harmoniques.

Annexe A

Valeurs des constantes κ_d

$$\begin{aligned}
 \kappa_2 &= 0 \\
 \kappa_3 &= \zeta(2) \\
 \kappa_4 &= 2\zeta(3) \\
 \kappa_5 &= \frac{33}{40}\zeta(2)^2 \\
 \kappa_6 &= \frac{5}{4}\zeta(5) + \frac{1}{6}\zeta(2)\zeta(3) \\
 \kappa_7 &= \frac{1451}{7560}\zeta(2)^3 + \frac{7}{72}\zeta(3)^2 \\
 \kappa_8 &= \frac{1729}{5760}\zeta(7) + \frac{181}{3600}\zeta(3)\zeta(2)^2 + \frac{13}{360}\zeta(2)\zeta(5) \\
 \kappa_9 &= -\frac{17}{1920}\zeta(6,2) + \frac{11}{160}\zeta(3)\zeta(5) + \frac{1}{320}\zeta(2)\zeta(3)^2 + \frac{1891}{89600}\zeta(2)^4 \\
 \kappa_{10} &= \frac{529}{75600}\zeta(2)^2\zeta(5) + \frac{33941}{6350400}\zeta(2)^3\zeta(3) + \frac{17}{3360}\zeta(2)\zeta(7) + \frac{199271}{4354560}\zeta(9) + \\
 &\quad \frac{11}{12960}\zeta(3)^3 \\
 \kappa_{11} &= \frac{209}{302400}\zeta(5)\zeta(2)\zeta(3) + \frac{2893}{6048000}\zeta(2)^2\zeta(3)^2 + \frac{3311}{460800}\zeta(3)\zeta(7) - \\
 &\quad \frac{517}{921600}\zeta(8,2) + \frac{39457}{9676800}\zeta(5)^2 + \frac{426341}{221760000}\zeta(2)^5 \\
 \kappa_{12} &= -\frac{13}{100800}\zeta(3)\zeta(6,2) + \frac{877}{302400}\zeta(2)\zeta(9) + \frac{299}{604800}\zeta(5)\zeta(3)^2 + \\
 &\quad \frac{13}{907200}\zeta(2)\zeta(3)^3 + \frac{7949}{6048000}\zeta(2)^2\zeta(7) + \frac{1081}{1411200}\zeta(5)\zeta(2)^3 + \\
 &\quad \frac{172157}{423360000}\zeta(2)^4\zeta(3) - \frac{586337}{232243200}\zeta(11) - \frac{13}{100800}\zeta(8,2,1)
 \end{aligned}$$

$$\begin{aligned}
\kappa_{13} = & \frac{169}{3456000} \zeta(5) \zeta(3) \zeta(2)^2 - \frac{1703}{43545600} \zeta(7) \zeta(2) \zeta(3) + \frac{9061}{8294400} \zeta(3) \zeta(9) + \\
& \frac{471809}{348364800} \zeta(5) \zeta(7) - \frac{13}{604800} \zeta(2)^2 \zeta(6, 2) - \frac{13}{215040} \zeta(2) \zeta(8, 2) + \\
& \frac{4667}{381024000} \zeta(2)^3 \zeta(3)^2 + \frac{11947}{174182400} \zeta(10, 2) - \frac{13}{483840} \zeta(8, 2, 1, 1) + \\
& \frac{13}{2903040} \zeta(3)^4 - \frac{2873}{87091200} \zeta(2) \zeta(5)^2 + \frac{11884374679}{152562009600000} \zeta(2)^6.
\end{aligned}$$

Annexe B

Expression exacte de $\text{Var}(K_{n,d})$

$$\text{Var}(K_{.,4}) = \underline{H}_{2,2,2} + 6\underline{H}_{1,1,1,2,1} - 14\underline{H}_{1,1,2,1,1} + 6\underline{H}_{1,2,1,1,1} + 6\underline{H}_{2,1,1,1,1} - 2\underline{H}_{2,2,1,1} + 4\underline{H}_{1,1,2,2} - 2\underline{H}_{1,2,1,2} - 2\underline{H}_{2,1,1,2} + \underline{H}_{1,1,1} - 2\underline{H}_{1,2,2,1} - 2\underline{H}_{2,1,2,1}$$

$$\begin{aligned} \text{Var}(K_{.,5}) = & -6\underline{H}_{2,1,2,1,1,1} - \underline{H}_{2,2,2,2} - 6\underline{H}_{1,1,1,1,2,2} + 14\underline{H}_{1,1,1,2,1,2} + 20\underline{H}_{1,1,1,1,2,1,1} - \\ & 6\underline{H}_{1,1,2,1,2,1} + 2\underline{H}_{1,2,2,2,1} - 6\underline{H}_{1,1,2,1,1,2} - 50\underline{H}_{1,1,1,2,1,1,1} + 2\underline{H}_{2,2,1,2,1} + 2\underline{H}_{2,2,1,1,2} - \\ & 6\underline{H}_{1,1,2,2,1,1} + 2\underline{H}_{2,2,2,1,1} - 6\underline{H}_{1,2,1,1,2,1} - 6\underline{H}_{1,2,1,1,1,2} + 20\underline{H}_{1,1,2,1,1,1,1} - \\ & 6\underline{H}_{2,1,1,1,2,1} + 2\underline{H}_{2,1,2,2,1} + \underline{H}_{1,1,1,1} - 6\underline{H}_{2,1,1,1,1,2} - 6\underline{H}_{1,2,1,2,1,1} + 2\underline{H}_{1,1,2,2,2} - \\ & 6\underline{H}_{2,1,1,2,1,1} + 14\underline{H}_{1,1,1,2,2,1} - 6\underline{H}_{2,2,1,1,1,1} + 20\underline{H}_{1,2,1,1,1,1,1} - 6\underline{H}_{1,2,2,1,1,1} + \\ & 2\underline{H}_{1,2,1,2,2} + 20\underline{H}_{2,1,1,1,1,1,1} + 2\underline{H}_{1,2,2,1,2} + 2\underline{H}_{2,1,2,1,2} + 2\underline{H}_{2,1,1,2,2} \end{aligned}$$

$$\begin{aligned} \text{Var}(K_{.,6}) = & 6\underline{H}_{1,2,1,1,1,2,2} + 6\underline{H}_{1,2,1,2,1,1,2} + \underline{H}_{1,1,1,1,1} - 20\underline{H}_{1,1,2,2,1,1,1,1} - 20\underline{H}_{1,2,1,1,2,1,1,1} - \\ & 182\underline{H}_{1,1,1,1,2,1,1,1,1} - 20\underline{H}_{1,1,2,1,1,2,1,1} - 2\underline{H}_{2,1,1,2,2,2} - 2\underline{H}_{2,1,2,2,2,1} - 2\underline{H}_{2,1,2,2,1,2} - \\ & 2\underline{H}_{1,2,1,2,2,2} - 2\underline{H}_{2,1,2,1,2,2} + \underline{H}_{2,2,2,2,2} - 20\underline{H}_{1,1,2,1,2,1,1,1} + 6\underline{H}_{1,1,2,1,1,2,2} - \\ & 2\underline{H}_{1,2,2,2,2,1} + 6\underline{H}_{1,1,2,1,2,2,1} + 6\underline{H}_{1,2,1,2,1,2,1} - 20\underline{H}_{1,2,1,1,1,1,1,2} - 20\underline{H}_{1,1,1,2,1,1,2,1} + \\ & 50\underline{H}_{1,1,1,1,2,1,2,1} + 6\underline{H}_{1,1,1,2,1,2,2} - 20\underline{H}_{1,1,1,2,1,1,1,2} + 6\underline{H}_{1,1,2,1,2,1,2} - \\ & 2\underline{H}_{2,2,1,2,2,1} - 20\underline{H}_{1,1,1,1,2,2,1} + 6\underline{H}_{1,2,1,1,2,1,2} - 2\underline{H}_{2,2,2,1,2,1} - 20\underline{H}_{1,1,2,1,1,1,2,1} - \\ & 20\underline{H}_{2,1,1,1,1,1,1,2} + 6\underline{H}_{2,1,1,2,1,2,1} + 50\underline{H}_{1,1,1,1,2,1,1,2} + 6\underline{H}_{2,1,1,1,2,1,2} + 6\underline{H}_{2,2,1,1,1,2,1} + \\ & 6\underline{H}_{1,1,1,2,2,1,2} + 50\underline{H}_{1,1,1,1,2,2,1,1} + 6\underline{H}_{2,1,2,1,2,1,1} + 6\underline{H}_{1,2,1,1,2,2,1} + 6\underline{H}_{1,2,2,1,1,1,2} - \\ & 20\underline{H}_{1,2,1,1,1,1,2,1} - 20\underline{H}_{1,1,1,1,2,1,2} - 20\underline{H}_{1,2,2,1,1,1,1,1} - 2\underline{H}_{1,1,2,2,2,2} + 6\underline{H}_{2,1,1,2,2,1,1} + \\ & 70\underline{H}_{1,1,2,1,1,1,1,1,1} - 20\underline{H}_{2,1,1,1,1,1,2,1} - 20\underline{H}_{1,1,1,2,2,1,1,1} + 6\underline{H}_{1,2,2,1,2,1,1} + \\ & 6\underline{H}_{1,1,2,2,1,2,1} - 20\underline{H}_{1,2,1,2,1,1,1,1} - 20\underline{H}_{2,1,1,1,1,2,1,1} - 20\underline{H}_{2,1,2,1,1,1,1,1} - 2\underline{H}_{2,2,1,1,2,2} - \\ & 2\underline{H}_{1,2,2,1,2,2} + 6\underline{H}_{2,1,2,2,1,1,1} - 20\underline{H}_{2,1,1,1,2,1,1,1} - 20\underline{H}_{1,2,1,1,1,2,1,1} - 20\underline{H}_{2,2,1,1,1,1,1,1} + \\ & 6\underline{H}_{1,1,2,2,1,1,2} - 2\underline{H}_{2,2,1,2,1,2} + 6\underline{H}_{1,2,2,2,1,1,1} + 70\underline{H}_{1,1,1,2,1,1,1,1,1} + 70\underline{H}_{2,1,1,1,1,1,1,1,1} + \\ & 6\underline{H}_{1,1,2,2,2,1,1} - 2\underline{H}_{2,2,2,1,1,2} + 6\underline{H}_{2,1,2,1,1,1,2} + 6\underline{H}_{2,2,1,1,1,1,2} - 14\underline{H}_{1,1,1,1,2,2,2} + \\ & 6\underline{H}_{2,1,1,1,2,2,1} + 6\underline{H}_{2,1,1,2,1,1,2} + 6\underline{H}_{1,2,1,2,2,1,1} - 20\underline{H}_{1,1,1,2,1,2,1,1} + 70\underline{H}_{1,1,1,1,1,2,1,1,1} - \\ & 20\underline{H}_{2,1,1,2,1,1,1,1} - 2\underline{H}_{2,2,2,2,1,1} + 6\underline{H}_{2,2,2,1,1,1,1} + 6\underline{H}_{2,1,2,1,1,2,1} - 2\underline{H}_{1,2,2,2,1,2} + \\ & 6\underline{H}_{1,1,1,2,2,2,1} + 6\underline{H}_{2,2,1,2,1,1,1} + 6\underline{H}_{2,2,1,1,2,1,1} + 70\underline{H}_{1,2,1,1,1,1,1,1,1} + 6\underline{H}_{2,1,1,1,1,2,2} - \\ & 20\underline{H}_{1,1,2,1,1,1,1,2} + 6\underline{H}_{1,2,2,1,1,2,1} \end{aligned}$$

$$\begin{aligned}
\text{Var}(K, \tau) = & 2 \underline{H}_{2,2,1,2,1,2,2} - 6 \underline{H}_{1,2,1,1,2,2,2,1} + 2 \underline{H}_{2,2,1,2,2,1,2} + 2 \underline{H}_{1,2,2,2,2,1,2} + 20 \underline{H}_{2,1,1,2,1,2,1,1,1} + \\
& 20 \underline{H}_{2,1,1,2,1,1,1,2,1} + 2 \underline{H}_{1,2,2,2,2,2,1} + 20 \underline{H}_{1,1,2,2,1,1,2,1,1} + 20 \underline{H}_{2,2,1,1,2,1,1,1,1} + 2 \underline{H}_{2,2,2,1,2,2,1} - \\
& 6 \underline{H}_{2,1,1,2,1,1,2,2} - 70 \underline{H}_{2,1,1,1,1,1,1,2,1} + 2 \underline{H}_{2,1,2,2,2,2,1} - 6 \underline{H}_{2,2,2,1,1,1,2,1} - 6 \underline{H}_{2,2,1,2,2,1,1,1} - \\
& 6 \underline{H}_{2,1,2,1,1,2,2,1} - 6 \underline{H}_{1,2,2,1,1,1,2,2} - 6 \underline{H}_{2,1,2,2,1,1,1,2} + 20 \underline{H}_{1,1,1,1,2,2,1,1,2} - 70 \underline{H}_{2,1,1,1,2,1,1,1,1,1} + \\
& 20 \underline{H}_{2,2,1,1,1,2,1,1,1} + 20 \underline{H}_{1,1,1,2,2,1,1,1,2} - 70 \underline{H}_{1,1,1,1,1,1,2,2,1,1} - 70 \underline{H}_{1,1,2,1,1,1,1,2,1,1,1} + 20 \underline{H}_{1,2,1,2,1,1,1,2,1,1} - \\
& 6 \underline{H}_{1,1,2,1,2,2,2,1} + 182 \underline{H}_{1,1,1,1,1,2,2,1,1,1} + 20 \underline{H}_{1,2,2,1,1,1,2,1,1} + 20 \underline{H}_{2,1,2,1,1,1,1,2,1} - 70 \underline{H}_{1,1,1,2,1,1,2,1,1,1} - \\
& 70 \underline{H}_{1,1,1,1,2,1,1,1,2,1} - 70 \underline{H}_{1,1,1,2,1,1,1,2,1,1} + 20 \underline{H}_{2,1,1,2,2,1,1,1,1} + 20 \underline{H}_{1,1,1,2,2,1,2,1,1} - 6 \underline{H}_{1,2,1,2,1,2,1,2} - \\
& 70 \underline{H}_{1,1,2,1,1,1,1,1,1,2} + 252 \underline{H}_{1,1,1,2,1,1,1,1,1,1,1} - 6 \underline{H}_{1,2,1,2,2,2,1,1} - 6 \underline{H}_{1,1,1,2,2,2,2,1} + \underline{H}_{1,1,1,1,1,1} - \\
& 70 \underline{H}_{1,2,1,1,1,1,1,2,1,1} + 20 \underline{H}_{2,1,1,1,1,2,1,1,1,2} - 70 \underline{H}_{2,1,2,1,1,1,1,1,1,1} - 70 \underline{H}_{1,1,1,1,2,1,2,1,1,1} + \\
& 20 \underline{H}_{2,1,1,1,1,2,1,2,1} - 6 \underline{H}_{2,1,1,1,2,2,1,2} - 70 \underline{H}_{1,2,2,1,1,1,1,1,1,1} + 20 \underline{H}_{1,1,1,2,1,2,1,2,1} - 6 \underline{H}_{1,2,1,2,1,1,2,2} - \\
& 6 \underline{H}_{1,2,1,2,1,2,2,1} - 6 \underline{H}_{2,2,2,1,1,1,1} - 6 \underline{H}_{2,2,1,1,1,1,2,2} + 2 \underline{H}_{2,2,2,1,1,2,2} + 2 \underline{H}_{2,1,2,2,1,2,2} + \\
& 20 \underline{H}_{1,2,1,1,1,2,1,2,1} - 6 \underline{H}_{1,2,1,2,2,1,2,1} - 50 \underline{H}_{1,1,1,1,2,1,2,2} - 6 \underline{H}_{2,1,1,2,2,1,1,2} + 20 \underline{H}_{2,1,2,1,2,1,1,1,1} + \\
& 20 \underline{H}_{1,2,1,1,2,1,1,1,2} + 20 \underline{H}_{2,1,2,1,1,1,2,1,1} + 2 \underline{H}_{2,2,2,1,2,1,2} + 182 \underline{H}_{1,1,1,1,2,1,2,1,1} + 20 \underline{H}_{2,2,1,1,1,1,1,2,1} - \\
& 70 \underline{H}_{2,2,1,1,1,1,1,1,1,1} + 20 \underline{H}_{1,2,1,2,1,2,1,1,1} - 6 \underline{H}_{2,1,1,2,1,2,2,1} + 20 \underline{H}_{1,1,2,2,2,1,1,1,1} - 6 \underline{H}_{2,2,1,1,1,2,2,1} + \\
& 20 \underline{H}_{1,1,2,1,1,2,1,1,2} + 2 \underline{H}_{2,2,2,2,1,1,2} + 20 \underline{H}_{2,2,1,2,1,1,1,1,1} + 20 \underline{H}_{1,1,1,2,1,1,1,2,2} + 2 \underline{H}_{2,2,2,2,1,2,1} - \\
& 70 \underline{H}_{2,1,1,1,1,1,2,1,1,1} - 70 \underline{H}_{1,1,2,2,1,1,1,1,1,1} + 2 \underline{H}_{2,1,1,2,2,2,2} + 20 \underline{H}_{1,1,2,1,1,2,2,1,1} - 6 \underline{H}_{2,1,2,1,2,1,2,1} - \\
& 6 \underline{H}_{1,1,2,2,1,2,2,1} + 2 \underline{H}_{1,1,2,2,2,2,2} - 6 \underline{H}_{2,2,1,2,1,1,1,2} - 6 \underline{H}_{1,2,2,1,1,2,1,2} - 6 \underline{H}_{2,2,2,1,1,1,1,2} + 2 \underline{H}_{2,1,2,2,2,1,2} - \\
& 6 \underline{H}_{2,2,1,1,2,1,1,2} + 20 \underline{H}_{1,1,1,2,2,2,1,1,1} - 6 \underline{H}_{1,2,2,2,1,1,2,1} - 6 \underline{H}_{1,1,1,1,2,2,2,2} + 20 \underline{H}_{1,2,1,2,1,1,1,1,2} + \\
& 20 \underline{H}_{1,1,2,2,1,1,1,2,1} + 20 \underline{H}_{2,1,1,2,1,1,1,1,2} - 70 \underline{H}_{1,1,2,1,1,1,1,2,1,1} - 6 \underline{H}_{2,1,2,2,1,1,2,1} + 20 \underline{H}_{1,2,1,1,1,1,1,2,1,2} - \\
& 20 \underline{H}_{2,1,1,1,1,2,2,1,1} + 20 \underline{H}_{2,1,1,1,2,1,1,1,2,1} - 6 \underline{H}_{2,1,1,2,2,1,2,1} - 70 \underline{H}_{2,1,1,1,1,1,1,1,1,2} + 20 \underline{H}_{1,1,1,2,1,1,2,1,2} - \\
& 70 \underline{H}_{1,1,2,1,1,2,1,1,1,1,1} + 20 \underline{H}_{1,2,2,1,2,1,1,1,1} - 70 \underline{H}_{1,1,2,1,1,1,1,1,2,1} + 20 \underline{H}_{1,1,1,1,2,2,1,1,1} + 20 \underline{H}_{1,1,1,2,1,2,1,1,2} - \\
& 6 \underline{H}_{2,2,1,2,1,2,1,1} - 70 \underline{H}_{1,2,1,1,1,1,1,1,1,2} - 70 \underline{H}_{2,1,1,1,1,2,1,1,1,1} + 20 \underline{H}_{1,1,1,2,1,2,2,1} - 70 \underline{H}_{1,2,1,1,1,1,2,1,1,1} - \\
& 6 \underline{H}_{1,2,2,1,2,2,1,1} + 20 \underline{H}_{1,2,1,1,1,2,1,1,1,2} - 70 \underline{H}_{1,1,1,2,1,1,1,1,1,2} - 6 \underline{H}_{2,2,1,2,1,1,2,1} + 20 \underline{H}_{2,1,1,1,1,1,2,1,2} - \\
& 70 \underline{H}_{1,1,1,1,2,1,1,2,1,1} + 20 \underline{H}_{1,2,2,1,1,1,1,1,2,1} + 20 \underline{H}_{2,2,1,1,1,1,1,1,1,2} - 6 \underline{H}_{1,1,2,2,1,2,1,2} + 20 \underline{H}_{1,1,1,2,1,2,2,1,1} - \\
& 6 \underline{H}_{2,1,2,1,2,2,1,1} - 6 \underline{H}_{2,1,1,1,1,2,2,2} - 6 \underline{H}_{2,1,2,2,1,2,1,1} - 6 \underline{H}_{2,2,1,1,1,2,1,2} + 252 \underline{H}_{1,2,1,1,1,1,1,1,1,1,1} - \\
& 70 \underline{H}_{1,2,1,1,1,2,1,1,1,1,1} + 20 \underline{H}_{2,1,2,1,1,1,1,1,1,2} - 6 \underline{H}_{1,1,2,1,1,1,1,1,1,2} + 252 \underline{H}_{1,1,1,1,1,1,2,1,1,1,1} + 20 \underline{H}_{1,1,2,2,1,1,1,1,1,2} + \\
& 182 \underline{H}_{1,1,1,1,1,2,1,1,1,2,1} - 6 \underline{H}_{1,2,1,1,1,2,2,2} - 6 \underline{H}_{1,2,2,2,1,2,1,1} + 20 \underline{H}_{2,1,1,1,2,1,1,1,1,2} + 20 \underline{H}_{1,2,2,1,1,2,1,1,1} - \\
& 70 \underline{H}_{1,1,1,1,2,2,1,1,1,1} + 20 \underline{H}_{1,1,2,1,1,1,2,2,1} - 6 \underline{H}_{2,1,2,1,1,1,2,2} - 6 \underline{H}_{1,1,2,2,2,2,1,1} - 6 \underline{H}_{1,2,2,2,2,1,1,1} - \\
& 70 \underline{H}_{2,1,1,1,1,1,2,1,1} + 20 \underline{H}_{1,2,1,2,2,1,1,1,1} + 20 \underline{H}_{1,1,1,1,2,1,1,2,2} - 6 \underline{H}_{1,1,1,2,2,1,2,2} + 20 \underline{H}_{1,2,1,1,2,2,1,1,1} - \\
& 6 \underline{H}_{1,1,1,2,2,2,1,2} + 2 \underline{H}_{2,2,1,1,2,2,2} + 20 \underline{H}_{1,2,1,1,2,1,1,2,1} + 252 \underline{H}_{1,1,1,1,2,1,1,1,1,1,1} + 20 \underline{H}_{1,2,1,2,1,1,1,2,1} + \\
& 2 \underline{H}_{2,2,2,2,2,1,1} - 6 \underline{H}_{2,1,1,2,2,2,1,1} - 6 \underline{H}_{1,2,1,1,2,1,2,2} + 20 \underline{H}_{2,2,2,1,1,1,1,1,1} + 20 \underline{H}_{1,1,1,2,1,1,2,2,1} + \\
& 20 \underline{H}_{1,1,1,1,2,2,1,2,1} - 6 \underline{H}_{2,1,2,1,2,1,1,2} - 6 \underline{H}_{1,1,2,2,1,1,2,2} - 6 \underline{H}_{2,1,2,2,2,1,1,1} - 50 \underline{H}_{1,1,1,1,1,2,2,1,2} + \\
& 20 \underline{H}_{1,1,2,1,2,1,1,2,1} - 70 \underline{H}_{1,1,2,1,2,1,1,1,1,1} + 20 \underline{H}_{1,1,1,2,2,1,1,2,1} - 6 \underline{H}_{2,1,1,1,2,1,2,2} + 20 \underline{H}_{1,2,1,1,2,1,2,1,1} - \\
& 6 \underline{H}_{2,1,1,2,1,2,1,2} - 70 \underline{H}_{1,1,1,1,2,1,1,1,1,2} + 20 \underline{H}_{2,1,1,1,2,1,2,1,1} - 672 \underline{H}_{1,1,1,1,1,2,1,1,1,1,1} - 6 \underline{H}_{1,2,2,1,2,1,1,2} - \\
& 6 \underline{H}_{1,2,2,1,2,1,2,1} - \underline{H}_{2,2,2,2,2,2} + 2 \underline{H}_{1,2,2,2,1,2,2} + 2 \underline{H}_{2,1,2,1,2,2,2} + 20 \underline{H}_{1,1,1,1,2,1,2,1,2} - 70 \underline{H}_{1,2,1,2,1,1,1,1,1,1} - \\
& 6 \underline{H}_{1,2,1,2,2,1,1,2} + 20 \underline{H}_{1,1,2,1,2,2,1,1,1} - 6 \underline{H}_{1,2,1,1,2,2,1,2} + 20 \underline{H}_{1,1,2,1,2,1,2,1,1} - 6 \underline{H}_{1,1,2,2,2,1,2,1} - \\
& 70 \underline{H}_{1,1,1,1,1,1,2,1,2,1} + 2 \underline{H}_{1,2,2,1,2,2,2} + 20 \underline{H}_{1,2,2,1,1,1,1,1,2} + 20 \underline{H}_{2,1,2,2,1,1,1,1,1} - 6 \underline{H}_{1,1,2,1,2,2,1,2} + \\
& 20 \underline{H}_{1,1,2,1,1,1,2,1,2} - 70 \underline{H}_{1,1,1,2,2,1,1,1,1,1} + 20 \underline{H}_{1,2,1,1,1,1,2,2,1} + 20 \underline{H}_{2,2,1,1,1,1,2,1,1} + 20 \underline{H}_{2,1,2,1,1,2,1,1,1} - \\
& 6 \underline{H}_{2,1,2,1,1,2,1,2} + 182 \underline{H}_{1,1,1,1,1,2,1,1,1,2} - 70 \underline{H}_{1,1,1,1,1,1,2,1,1,2} - 6 \underline{H}_{2,1,1,1,2,2,2,1} - 6 \underline{H}_{1,1,1,2,1,2,2,2} + \\
& 20 \underline{H}_{1,1,2,1,2,1,1,1,1,2} - 6 \underline{H}_{2,2,1,1,2,2,1,1} - 6 \underline{H}_{1,2,2,1,1,2,2,1} - 70 \underline{H}_{1,2,1,1,2,1,1,1,1,1} - 6 \underline{H}_{2,2,2,1,2,1,1,1} + \\
& 2 \underline{H}_{1,2,1,2,2,2,2} - 50 \underline{H}_{1,1,1,1,1,2,2,2,1} + 252 \underline{H}_{2,1,1,1,1,1,1,1,1,1,1} - 70 \underline{H}_{2,1,1,2,1,1,1,1,1,1,1} + 20 \underline{H}_{1,2,1,1,1,1,1,2,1,2} - \\
& 6 \underline{H}_{1,2,2,2,1,1,1,1,2} + 20 \underline{H}_{2,1,1,1,1,1,2,2,1} + 20 \underline{H}_{2,1,1,2,1,1,2,1,1} + 20 \underline{H}_{1,1,2,1,1,2,1,2,1} - 70 \underline{H}_{1,2,1,1,1,1,1,1,2,1} + \\
& 20 \underline{H}_{1,1,1,1,1,1,2,2,2} - 6 \underline{H}_{1,1,2,2,2,1,1,2} + 252 \underline{H}_{1,1,2,1,1,1,1,1,1,1,1} + 2 \underline{H}_{2,2,1,2,2,2,1} + 20 \underline{H}_{2,1,1,1,1,1,1,2,2} + \\
& 20 \underline{H}_{1,1,2,1,1,1,1,2,2} + 20 \underline{H}_{1,2,1,1,1,1,1,2,2} - 6 \underline{H}_{2,2,1,1,2,1,2,1} + 20 \underline{H}_{1,1,2,2,1,2,1,1,1} - 70 \underline{H}_{1,1,1,2,1,1,1,1,2,1} - \\
& 6 \underline{H}_{2,2,2,1,1,2,1,1} - 70 \underline{H}_{1,1,1,2,1,2,1,1,1,1} + 20 \underline{H}_{1,2,2,2,1,1,1,1,1} - 6 \underline{H}_{1,1,2,1,2,1,2,2} + 20 \underline{H}_{2,1,1,1,2,2,1,1,1}
\end{aligned}$$

Annexe C

Développement asymptotique de $\text{Var}(K_{n,d})$

$$\text{Var}(K_{n,3}) = \left(\frac{1}{2} + \kappa_3\right) \ln^2(n) + (-10\zeta(3) + 2\zeta(2)\gamma + \gamma) \ln(n) + \frac{1}{2}\gamma^2 - 10\zeta(3)\gamma + \frac{83}{10}\zeta(2)^2 + \zeta(2)\gamma^2 + \frac{1}{2}\zeta(2) + \mathfrak{o}(1)$$

$$\begin{aligned} \text{Var}(K_{n,4}) = & \left(\frac{1}{3!} + \kappa_4\right) \ln^3(n) + \left(-\frac{53}{5}\zeta(2)^2 + 6\zeta(3)\gamma + \frac{1}{2}\gamma\right) \ln^2(n) + \\ & \left(97\zeta(5) - \frac{106}{5}\zeta(2)^2\gamma + 16\zeta(2)\zeta(3) + 6\zeta(3)\gamma^2 + \frac{1}{2}\zeta(2) + \frac{1}{2}\gamma^2\right) \ln(n) + \\ & \frac{1}{3}\zeta(3) - \frac{53}{5}\zeta(2)^2\gamma^2 - \frac{3719}{70}\zeta(2)^3 + \frac{1}{6}\gamma^3 + \frac{1}{2}\zeta(2)\gamma + 16\zeta(2)\zeta(3)\gamma - \\ & 3\zeta(3)^2 + 2\zeta(3)\gamma^3 + 97\zeta(5)\gamma + \mathfrak{o}(1) \end{aligned}$$

$$\begin{aligned} \text{Var}(K_{n,5}) = & \left(\frac{1}{4!} + \kappa_5\right) \ln^4(n) + \left(\frac{1}{6}\gamma - \frac{98}{3}\zeta(5) + \frac{33}{10}\zeta(2)^2\gamma - \frac{13}{3}\zeta(2)\zeta(3)\right) \ln^3(n) + \\ & \left(\frac{10123}{140}\zeta(2)^3 + \frac{47}{2}\zeta(3)^2 + \frac{99}{20}\zeta(2)^2\gamma^2 + \frac{1}{4}\gamma^2 + \frac{1}{4}\zeta(2) - 13\zeta(2)\zeta(3)\gamma - \right. \\ & \left. 98\zeta(5)\gamma\right) \ln^2(n) + \left(\frac{1}{6}\gamma^3 + \frac{33}{10}\zeta(2)^2\gamma^3 + \frac{1}{2}\zeta(2)\gamma - 950\zeta(7) - \right. \\ & \left. 13\zeta(2)\zeta(3)\gamma^2 + 47\zeta(3)^2\gamma + \frac{1}{3}\zeta(3) - \frac{317}{5}\zeta(3)\zeta(2)^2 + \frac{10123}{70}\zeta(2)^3\gamma - \right. \\ & \left. 98\zeta(5)\gamma^2 - 222\zeta(2)\zeta(5)\right) \ln(n) - \frac{13}{3}\zeta(2)\zeta(3)\gamma^3 + \frac{47}{2}\zeta(3)^2\gamma^2 - \\ & \frac{317}{5}\zeta(3)\zeta(2)^2\gamma - \frac{98}{3}\zeta(5)\gamma^3 + \frac{33}{40}\zeta(2)^2\gamma^4 + \frac{32}{3}\zeta(3)\zeta(5) + \\ & \frac{10123}{140}\zeta(2)^3\gamma^2 - 222\zeta(2)\zeta(5)\gamma + \frac{1}{24}\gamma^4 - 950\zeta(7)\gamma + 50\zeta(6,2) + \\ & \frac{1}{4}\zeta(2)\gamma^2 + \frac{1}{3}\zeta(3)\gamma + \frac{9}{40}\zeta(2)^2 + \frac{95}{6}\zeta(2)\zeta(3)^2 + \frac{134739}{350}\zeta(2)^4 + \mathfrak{o}(1) \end{aligned}$$

$$\begin{aligned}
\text{Var}(K_{n,6}) = & \left(\frac{1}{5!} + \kappa_6 \right) \ln^5(n) + \left(\frac{1}{24} \gamma + \frac{25}{4} \zeta(5) \gamma + \frac{5}{6} \zeta(2) \zeta(3) \gamma - \frac{25}{6} \zeta(3)^2 - \right. \\
& \left. \frac{22711}{2520} \zeta(2)^3 \right) \ln^4(n) + \left(\frac{1}{12} \gamma^2 + \frac{1231}{30} \zeta(3) \zeta(2)^2 - \frac{50}{3} \zeta(3)^2 \gamma + \right. \\
& \left. \frac{8729}{24} \zeta(7) + \frac{127}{2} \zeta(2) \zeta(5) + \frac{1}{12} \zeta(2) + \frac{5}{3} \zeta(2) \zeta(3) \gamma^2 - \frac{22711}{630} \zeta(2)^3 \gamma + \right. \\
& \left. \frac{25}{2} \zeta(5) \gamma^2 \right) \ln^3(n) + \left(-55 \zeta(6, 2) - \frac{241}{6} \zeta(2) \zeta(3)^2 + \frac{1231}{10} \zeta(3) \zeta(2)^2 \gamma + \right. \\
& \left. \frac{1}{12} \gamma^3 + \frac{1}{4} \zeta(2) \gamma + \frac{8729}{8} \zeta(7) \gamma - \frac{2331589}{4200} \zeta(2)^4 + \frac{1}{6} \zeta(3) + \frac{25}{2} \zeta(5) \gamma^3 + \right. \\
& \left. \frac{5}{3} \zeta(2) \zeta(3) \gamma^3 - 342 \zeta(3) \zeta(5) - 25 \zeta(3)^2 \gamma^2 - \frac{22711}{420} \zeta(2)^3 \gamma^2 + \right. \\
& \left. \frac{381}{2} \zeta(2) \zeta(5) \gamma \right) \ln^2(n) + \left(\frac{8729}{8} \zeta(7) \gamma^2 - \frac{2331589}{2100} \zeta(2)^4 \gamma + \right. \\
& \left. \frac{381}{2} \zeta(2) \zeta(5) \gamma^2 - \frac{241}{3} \zeta(2) \zeta(3)^2 \gamma + \frac{1231}{10} \zeta(3) \zeta(2)^2 \gamma^2 - 19 \zeta(3)^3 + \right. \\
& \left. \frac{1}{4} \zeta(2) \gamma^2 + \frac{9}{40} \zeta(2)^2 - \frac{50}{3} \zeta(3)^2 \gamma^3 - 110 \zeta(6, 2) \gamma + \frac{25}{4} \zeta(5) \gamma^4 + \right. \\
& \left. \frac{1}{3} \zeta(3) \gamma + \frac{21919}{20} \zeta(2)^2 \zeta(5) - \frac{22711}{630} \zeta(2)^3 \gamma^3 + \frac{135593}{315} \zeta(2)^3 \zeta(3) + \right. \\
& \left. \frac{182179}{18} \zeta(9) + \frac{5}{6} \zeta(2) \zeta(3) \gamma^4 - 684 \zeta(3) \zeta(5) \gamma + \frac{19209}{8} \zeta(2) \zeta(7) + \right. \\
& \left. \frac{1}{24} \gamma^4 \right) \ln(n) + \frac{127}{2} \zeta(2) \zeta(5) \gamma^3 + \frac{1231}{30} \zeta(3) \zeta(2)^2 \gamma^3 - \frac{241}{6} \zeta(2) \zeta(3)^2 \gamma^2 - \\
& 342 \zeta(3) \zeta(5) \gamma^2 + \frac{1}{6} \zeta(2) \zeta(3) \gamma^5 + \frac{1}{6} \zeta(2) \zeta(3) + \frac{182179}{18} \zeta(9) \gamma + \\
& \frac{9}{40} \zeta(2)^2 \gamma - 19 \zeta(3)^3 \gamma + \frac{1}{6} \zeta(3) \gamma^2 - \frac{2331589}{4200} \zeta(2)^4 \gamma^2 + \frac{1}{5} \zeta(5) - \\
& 55 \zeta(6, 2) \gamma^2 - 325 \zeta(8, 2) + \frac{135593}{315} \zeta(2)^3 \zeta(3) \gamma - 55 \zeta(2) \zeta(6, 2) + \\
& \frac{1}{12} \zeta(2) \gamma^3 + \frac{8729}{24} \zeta(7) \gamma^3 - \frac{22711}{2520} \zeta(2)^3 \gamma^4 - \frac{25}{6} \zeta(3)^2 \gamma^4 + \frac{1}{120} \gamma^5 - \\
& 945 \zeta(5)^2 - \frac{6237237}{2200} \zeta(2)^5 + \frac{767}{30} \zeta(2)^2 \zeta(3)^2 - \frac{9031}{12} \zeta(3) \zeta(7) - \\
& 392 \zeta(5) \zeta(2) \zeta(3) + \frac{5}{4} \zeta(5) \gamma^5 + \frac{21919}{20} \zeta(2)^2 \zeta(5) \gamma + \frac{19209}{8} \zeta(2) \zeta(7) \gamma + \\
& \mathfrak{o}(1).
\end{aligned}$$

Bibliographie

- [1] Z.D. BAI, C.C. CHAO, H.K. HWANG, ET W.Q. LIANG. On the variance of the number of maxima in random vectors and its applications. *Annals of Applied Probability* 8 (1998), 886–895. MathReview : 99f :60019.
- [2] Z.D. BAI, L. DEVROYE, H.K. HWANG, ET T.H. TSAI. Maxima in hypercubes. *Random structures and Algorithmes* 27 (2005), 290–309.
- [3] Z.D. BAI, H.K. HWANG, W.Q. LIANG, ET T.H. TSAI. Limit theorems for the number of maxima in random samples from planar regions. *Electronic Journal of Probability* 6 (2001), Paper 3.
- [4] O. BARNDORFF-NIELSEN ET M. SOBEL. On the distribution of the number of admissible points in a vector random sample. *Theory of Probability and Applications* 11 (1966), 249–269.
- [5] M. BIGOTTE. *Étude symbolique et algorithmique des fonctions polylogarithmes et des nombres d'Euler-Zagier colorés*. Thèse, Université Lille 1, 2000.
- [6] F. BRABEK ET H. SAMET. <http://donar.umiacs.umd.edu/quadtrees/>.
- [7] A. CARLSUND. Notes on the variance of the number of maxima in three dimensions. *Random Structure Algorithms* 22 (2003), 440–447.
- [8] H.H. CHERN, M. FUCHS, ET H.K. HWANG. Phase changes in random point quad-trees. *ACM Transactions on Algorithms* 3 (2007).
- [9] H.H. CHERN, H.K. HWANG, ET T.H. TSAI. An asymptotic theory for cauchy-euler differential equations with applications to the analysis of algorithms. *Journal of Algorithms* 44 (2003), 177–225.
- [10] C. COSTERMANS, J.Y. ENJALBERT, ET HOANG NGOC MINH. Algorithmic and combinatoric aspects of multiple harmonic sums. *Discrete mathematics and theoretical computer science* (2005). <http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/article/view/dmAD0107/>.
- [11] C. COSTERMANS, J.Y. ENJALBERT, HOANG NGOC MINH, ET M. PETITOT. Structure and asymptotic expansion of multiple harmonic sums. In *International Symposium on Symbolic and Algebraic Computation* (2005), pp. 100–107. <http://doi.acm.org/10.1145/1073884.1073900>.

- [12] C. COSTERMANS ET HOANG NGOC MINH. Explicit computation of the variance of the number of maxima in hypercubes. In *Fourth Colloquium on Mathematics and Computer Science* (2006). <http://mathinfo06.iecn.u-nancy.fr/papers/dmAG427-430.pdf>.
- [13] C. COSTERMANS ET HOANG NGOC MINH. Some results à l'abel obtained by use of techniques à la hopf. In *Global Integrability of Field Theories and Applications, Daresbury, UK* (2006), pp. 63–84. http://iaks-www.ira.uka.de/iaks-calmet/gift2006/GIFT2006_Hoang.pdf.
- [14] C. COSTERMANS ET HOANG NGOC MINH. Variance for the number of maxima in hypercubes and generalized euler constants. In *Formal Power Series and Algebraic Combinatorics* (2007). <http://www.fpsac.cn/PDF-Proceedings/Posters/63.pdf>.
- [15] C. COSTERMANS ET HOANG NGOC MINH. Noncommutative algebra, multiple harmonic sums and applications in discrete probability. *Journal of Symbolic Computation* (2008). À paraître.
- [16] M. DE BERG, M. VAN KREVELD, M. OVERMARS, ET O. SCHWARZKOPF. *Computational Geometry : Algorithms and Applications*, 2 ed. Springer-Verlag, 2000.
- [17] L. DEVROYE. Universal limit laws for depths in random trees. *SIAM Journal on Computing* 28 (1998), 409–432.
- [18] V. DRINFEL'D. Quasi-hopf algebras. *Len. Math. Journal* 1 (1990), 1419–1457.
- [19] J. ECALLE. Ari/gari, la dimorphie et l'arithmétique des multizêtas : un premier bilan. *Journal de théorie des nombres de Bordeaux* 15, 2 (2003), 411–478.
- [20] J.Y. ENJALBERT ET HOANG NGOC MINH. Analytic and combinatoric aspects of hurwitz polyzêtas. *Journal de Théorie des Nombres de Bordeaux* 19 (2007), 595–640.
- [21] R.A. FINKEL ET J.L. BENTLEY. Quadrees : A data structure for retrieval on composite keys. *Acta Informatica* 4 (1974), 1–9.
- [22] P. FLAJOLET. *Analytic Combinatorics*. <http://algo.inria.fr/flajolet/Publications/book.pdf>, 2007.
- [23] P. FLAJOLET, G. LABELLE, L. LAFOREST, ET B. SALVY. Hypergeometrics and the cost structure of quadrees. *Random Structures and Algorithms* 7 (1995), 117–144.
- [24] P. FLAJOLET ET A. ODLYZKO. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics* 3 (1990), 216–240.
- [25] P. FLAJOLET ET R. SEDGEWICK. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, 1995.
- [26] P. FLAJOLET ET B. VALLÉE. Continued fractions, comparison algorithms, and fine structure constants. In *Constructive, Experimental, and Nonlinear Analysis* (2000), pp. 53–82.

- [27] D. FOATA, G.N. HAN, ET B. LASS. Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes. *Journal électronique du Séminaire Lotharingien de Combinatoire* 47 (2001).
- [28] I. GESSEL. Multipartite p -partitions and inner products of skew schur functions. *Combinatorics and Algebra, Contemp. Math.* 34 (1984), 289–301.
- [29] G.H. HARDY. *Divergent Series*, 3e ed. Jacques Gabay, 2000.
- [30] M.E. HOFFMAN. The algebra of multiple harmonic series. *Journal of Algebra* 194 (1997), 477–495.
- [31] M.E. HOFFMAN. Quasi-shuffle products. *Journal of Algebraic Combinatorics* 11 (2000), 49–68. <http://arxiv.org/abs/math/9907173v1>.
- [32] M.E. HOFFMAN. The hopf algebra structure of multiple harmonic sums. *Nuclear Phys. B. Proc Suppl.* 135 (2004), 215–219.
- [33] M.E. HOFFMAN. *Algebraic Aspects of Multiple Zeta Values*, vol. 14 of *Dev. Math.* Springer, New York, 2005.
- [34] H.K. HWANG ET R. NEININGER. Phase change of limit laws in the quicksort recurrences under varying toll functions. *SIAM Journal on Computing* 31 (2002), 1687–1722.
- [35] V.M. IVANIN. Calculation of the dispersion of the number of elements of the pareto set for the choice of independent vectors with independent components. *Theory of Optimal Decisions, Kiev* 1 (1976), 90–100.
- [36] N. KAPUR. *Additive functionals on random search trees*. Thèse, Johns Hopkins University, 2003.
- [37] A. KLINGER. Patterns and search statistics. *Optimizing Methods in Statistics* (1971), 303–339.
- [38] D.E. KNUTH. Big omicron and big omega and big theta. *ACM SIGACT* 8, Issue 2 (1976), 18–24.
- [39] D.E. KNUTH. *The Art of Computer Programming, Vol. 3 : Sorting and Searching*, second ed. Addison-Wesley, 1998.
- [40] G. LABELLE. Some combinatorial results first found using computer algebra. *Journal of Symbolic Computation* 20, 5-6 (1995), 567–594.
- [41] G. LABELLE ET L. LAFOREST. Combinatorial variations on multidimensional quadtrees. *Journal of Combinatorial Theory* 69 (1995), 1–16.
- [42] G. LABELLE ET L. LAFOREST. Sur la distribution de l'arité de la racine d'une arborescence hyperquaternaire. *Discrete Mathematics* 139 (1995), 287–302.
- [43] G. LABELLE ET L. LAFOREST. Etude de constantes universelles pour les arborescences hyperquaternaires de recherche. *Discrete Mathematics* 153 (1996), 199–211.
- [44] G. LABELLE, L. LAFOREST, ET X. PROVENÇAL. Around the root of multidimensional quadtrees. *Discrete Mathematics and Theoretical Computer Science*

- (2006). <http://www.dmtcs.org/dmtcs-ojs/index.php/proceedings/article/view/dmAG0125/1702>.
- [45] L. LAFOREST. *Étude des arbres hyperquaternaires*. Thèse, Publications du LA-CIM/UQAM, 1990.
- [46] M. LASZLO ET S. MUKHERJEE. A genetic algorithm using hyper-quadtrees for low-dimensional k-means clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions* 28, Issue 4 (2006), 533–543.
- [47] M. LOTHAIRE. *Combinatorics on words*. Addison-Wesley, Reading, MA, 1983.
- [48] I. MACDONALD. *Symmetric functions and Hall polynomials*, 2e ed. The Clarendon Press, Oxford University Press, 1995.
- [49] C. MALVENUTO ET C. REUTENAUER. Duality between quasi-symmetric functions and the solomon descent algebra. *Jour. of Alg.* 177 (1995), 967–982.
- [50] D. MEHTA ET S. SAHNI, Eds. *Handbook of Data Structures and Applications, chapitre 16*. CRC Press, 2005. <http://www.cs.umd.edu/~hjs/pubs/SameCRC05.pdf>.
- [51] G. MELANÇON ET C. REUTENAUER. Lyndon words, free algebras and shuffles. *Canadian Journal of Mathematics* 41 (1989), 577–591.
- [52] HOANG NGOC MINH. A propos des intégrales $j_n[s]$. Communication privée à G. Labelle, 2001.
- [53] HOANG NGOC MINH. Differential galois groups and noncommutative generating series of polylogarithms. *7th World Multiconference on Systemics, Cybernetics and informatics, Orlando, Floride* (2003), 128–135.
- [54] HOANG NGOC MINH. Finite polyzêtas, poly-bernoulli numbers, identities of polyzêtas and noncommutative rational power series. In *4th International Conference on Words* (2003), pp. 232–250.
- [55] HOANG NGOC MINH. Des propriétés structurelles des polylogarithmes aux aspects algorithmiques des sommes harmoniques multiples. Séminaire "Polylogarithmes et nombres zêta multiples", Paris, 2005.
- [56] HOANG NGOC MINH. Algebraic combinatoric aspects of asymptotic analysis of nonlinear dynamical system with singular inputs. In *Computer Algebra and Differential Equations* (2007), vol. 67 of B, Acta Academiae Aboensis, pp. 117–126.
- [57] HOANG NGOC MINH ET G. JACOB. Symbolic integration of meromorphic differential systems via dirichlet functions. *Discrete Mathematics* 210 (2000), 87–116.
- [58] HOANG NGOC MINH, G. JACOB, N.E. OUSSOUS, ET M. PETITOT. De l’algèbre des ζ de riemann multivariées à l’algèbre des ζ de hurwitz multivariées. *Journal électronique du Séminaire Lotharingien de Combinatoire* 44 (2001).
- [59] HOANG NGOC MINH ET M. PETITOT. Lyndon words, polylogarithms and the riemann ζ function. *Discrete Mathematics* 217 (2000), 273–292.

- [60] HOANG NGOC MINH, M. PETITOT, ET J. VAN DER HOEVEN. Shuffle algebra and polylogarithms. In *Formal Power Series and Algebraic Combinatorics* (1998).
- [61] HOANG NGOC MINH, M. PETITOT, ET J. VAN DER HOEVEN. L'algèbre des polylogarithmes par les séries génératrices. In *Formal Power Series and Algebraic Combinatorics* (1999).
- [62] R. NEININGER ET L. RÜSCHENDORF. On the internal path length of d -dimensional quadrees. *Random Structures and Algorithms* (1999).
- [63] T. OSARAGI. Spatial clustering method for geographic data. *Working Paper Series 41, Centre for Advanced Spatial Analysis* (2002). <http://eprints.ucl.ac.uk/253/1/Paper41.pdf>.
- [64] G. RACINET. *Séries génératrices non commutatives de polyzêtas et associateurs de Drinfel'd*. Thèse, Université de Picardie, 2000.
- [65] D.E. RADFORD. A natural ring basis for shuffle algebra and an application to group schemes. *Journal of Algebra* 58 (1979), 432–454.
- [66] C. REUTENAUER. *Free Lie Algebras*, vol. 7 of *London Mathematical Society Monographs, New Series*. Oxford University Press, USA, 1993.
- [67] H. SAMET. The quadtree and related hierarchical data structures. *Computing Surveys* 16 (1984), 187–260.
- [68] L.A. SHOLOMOV. Survey of estimational results in choice problems. *Engrg. Cybernetics* 21 (1983), 51–75.
- [69] T. TRAN. *Approches combinatoires pour l'analyse des données de biopuces*. Thèse, Université Lille 2, 2006.
- [70] D. ZAGIER. Values of zeta functions and their applications. *First European congress of Mathematics* 2 (1994), 497–512.

Index

- $K_{n,d}$, 74
 X^* , 24
 X^+ , 24
 Z , 36
 Z_{\sqcup} , 58
 Z_{\sqcup} , 37
 $[z^n]f(z)$, 46
 $\mathbb{C}\langle X \rangle$, 24
 $\mathbb{C}\langle\langle X \rangle\rangle$, 24
 F_w , 40
 G_w , 40
 H , 45
 H_w , 42
 \underline{H}_w , 42
 L , 35
 LI_C , 34
 Λ , 45
 Li_w , 34
 Li_s , 33
 $\mathcal{L}yn X$, 26
 N_q , 47
 P , 37
 P_w , 37
 Σ , 67
 \mathcal{C} , 34
 \mathcal{Z} , 44
 \sqcup , 24
 $J_n[S]$, 10
 $J_n^*[S]$, 10
 $H_s(N)$, 9
 $H_s(N; k)$, 9
 $\underline{H}_s(N)$, 9
 $\underline{H}_s(N; k)$, 9
 $\pi_{n,k,d}$, 8
 κ_d , 42
 $\lambda_k^{(r)}$, 40
 $\langle S|w \rangle$, 24
 $\text{Comp}(r)$, 30
 Const , 45
 C_1 , 26
 C_2 , 26
 Mono , 45
 ∇ , 67
 ω_0 , 33
 ω_1 , 33
 π_Y , 26, 45
 $\psi_k^{(r)}$, 40
 \sqcup , 24
 \star , 25
 \sqcup , 24
 θ , 47
 $\underline{\zeta}(w)$, 43
 $\zeta(w)$, 34
 ζ_{\sqcup} , 56
 ζ_{\sqcup} , 36
 d -analogue, 18
 w^\sharp , 67
 $|w|$, 24
 $|w|_{x_i}$, 24
 algèbre
 de Hopf, 29
 de Lie, 27
 alphabet, 24
 antipode, 30
 arbre hyperquaternaire, 6
 base de PBW, 28
 bigèbre, 30
 changement de phase, 19

- collectionneur de vignettes, 63
- constante d'Euler, 56
- constante universelle, 21
- crochet
 - d'Hoffman, 25
 - de Lie, 27
- décomposition
 - en produit de mots de Lyndon, 27
 - de Radford, 26
 - de Taylor, 54
 - standard d'un mot de Lyndon, 27
- développement asymptotique
 - Algorithme 1, 50
 - Algorithme 2, 53
 - Algorithme 3, 59
- équation de Drinfel'd, 34
- facteur gauche, droit, 24
- fonction
 - quasi-symétrique, 40
 - symétrique, 40
- formule
 - d'Euler-MacLaurin, 53
 - de Waring, 40
- identité de Newton, 40
- longueur d'un mot, 24
- maxima, 74
- minus-stuffle, 24
- mot, 24
 - de Lyndon, 26
 - vide, 24
- nombre hyperharmonique, 64
- opérateur d'Euler, 47
- paramètre additif, 7
- poids d'un mot, 24
- polylogarithme, 33
- polynôme
 - de Bell, 57
 - de Bernstein, 65
- polyzêta, 34
 - large, 43
- produit
 - de concaténation, 24
 - de quasi-mélange, 25
- série
 - de Lie, 27
 - diagonale, 31
 - formelle, 24
 - group-like, 27
 - primitive, 27
 - régularisée, 36
- shuffle, 24
- somme harmonique multiple, 9
- stuffle, 24
- transformation d'Euler, 16

Remerciements

Je tiens pour une vérité absolue que celle d'un homme réside dans ses silences. L'exercice de style que constituent ces remerciements ne saurait donc constituer pour moi un épanchement en louanges, et nul voyageur égaré me connaissant peu ou prou ne devrait se sentir lésé s'il ne voyait son nom ici. Les qualités humaines ou professionnelles des uns et des autres ne se résument jamais à quelques considérations réductrices. Si je mentionnais ici ma femme et mon fils, de quoi donc pourrais-je aujourd'hui les remercier, si ce n'est d'exister ?

Les noms qui vont suivre occupent donc ou ont occupé une place particulière dans mon parcours. Pour citer mon père, "cela va de soi, mais ça va encore mieux en le disant."

Sobrement, je souhaiterais donc évoquer l'ensemble des professeurs de mathématiques croisés lors de ma (longue) scolarité, avec une attention particulière pour Mme Gournay, Mme Bermond et pour M. Lobry, et avec eux, leur passion du métier et de la matière.

Puis, d'autres enseignants passionnés m'ayant ouvert sur le monde de la recherche : M. Beckermann, Mme Philippe.

Préalablement au démarrage d'une thèse, il faut un directeur de thèse et un laboratoire d'accueil. Il m'a été donné de rencontrer les deux lors de mon mémoire de DEA, en les personnes de M. Hoang Ngoc Minh et de M. Beuscart, directeur du CERIM. Je sais tout ce que je leur dois.

Préalablement à la soutenance d'une thèse, il faut également un jury et des rapporteurs. Remerciements donc pour MM. Cresson, Duchamp, van der Hoeven, Mmes Viano, Tison et pour Mme Vallée, qui a accepté de présider mon jury.

Entre le début et la fin, d'autres rencontres. De celles qui font qu'aller "au labo" devient franchement différent d'aller "au boulot". L'ensemble du CERIM et d'EVALAB, donc, Loïc, Renaud, Sylvia et les autres. Un mot particulier pour Mme Nirel, et avec elle sa gentillesse et sa disponibilité.

Des rencontres de celles qui font (un peu) avancer la science, comme l'équipe Calcul Formel du LIFL, et son directeur, Michel Petitot, à qui je dois (et scientifiquement et matériellement) d'avoir visité la Chine.

Et puis, ceux qui ont partagé leur bureau avec moi, les collègues (et néanmoins amis), Tran, Arnaud (dont je salue ici la patience et le calme chaque fois que se tenait dans son bureau quelque relecture d'article animée...), Cristian, Costin.

Il me reste à évoquer deux noms, celui de Jean-Yves, dont je lui laisse le soin de déterminer à quel titre il se retrouve ici, et celui de Julien, pour lequel je n'évoquerai que deux de ses qualités : son extraordinaire curiosité et sa modestie qui n'a d'égale que l'étendue de ses compétences.

Enfin, un clin d'oeil, aussi discret que sincère, à la famille et aux amis.

Calcul symbolique non commutatif : analyse des constantes d'arbre de fouille

Résumé. L'étude de certaines variables aléatoires, comme les paramètres additifs sur les arbres hyperquaternaires de points, ou encore le nombre de maxima au sein d'un ensemble de n points indépendants, et uniformément distribués dans $[0, 1]^d$ font apparaître des suites particulières, les sommes harmoniques multiples (SHM), extensions des nombres harmoniques classiques à des multi-indices s .

Nos travaux visant à appliquer des méthodes symboliques pour l'étude de ces variables aléatoires, nous remplaçons l'utilisation de multi-indices par des codages sur des alphabets distincts, et nous appuyons alors sur des résultats importants en combinatoire des mots pour les appliquer à nos suites de SHM, et aux fonctions polylogarithmes, qui sont des variantes des génératrices ordinaires des SHM. Dans les cas convergents, les deux objets convergent (respectivement lorsque z tend vers 1 et lorsque N tend vers l'infini) vers la même limite, le polyzêta $\zeta(s)$. Pour les cas divergents, l'utilisation de séries génératrices non commutatives nous permet d'établir un théorème "à l'Abel", faisant apparaître une limite commune. Ce théorème permet de donner une forme explicite aux constantes d'Euler généralisées associées à des SHM divergentes et ainsi d'obtenir un algorithme très efficace pour calculer leur développement asymptotique.

Finalement, nous proposons des applications des sommes harmoniques dans le domaine des structures de données multidimensionnelles, pour lesquelles notre approche donne naissance à des calculs exacts, qui peuvent par la suite être aisément évalués asymptotiquement.

Mots clés : Calcul symbolique et formel, algèbre non commutative, polylogarithme, polyzêta, somme harmonique, analyse asymptotique, série génératrice, arbres hyperquaternaires, points maximaux

Noncommutative symbolic computation : analysis of search trees constants

Abstract. Working on some random variables, like additive parameters on multidimensional point quadrees, or the number of maxima among a set of n points, independent and uniformly distributed in $[0, 1]^d$ makes appear some particular sequences, called multiple harmonic sums (MHS), extensions of classical harmonic numbers to compositions s .

Aiming at applying symbolic methods to study these random variables, we replace the use of compositions by encodings over different alphabets, and then call upon important results in the theory of combinatorics on words to apply them to our special sequences of MHS and to polylogarithms, derivative of generating functions of MHS. In convergent cases (respectively as z tends to 1 and as N tends to infinity) both converge to the same limit, the polyzeta $\zeta(s)$. For divergent cases, the use of noncommutative generating series enables us to prove a theorem "à l'Abel", giving rise to a common limit. This theorem enables one to give an explicit form to generalized Euler constants associated to divergent MHS and so to get a very efficient algorithm to compute their asymptotic expansion.

Finally, we suggest some applications of harmonic sums in the field of multidimensional data structure, for which our approach gives rise to exact computations, which can be then easily asymptotically evaluated.

Keywords : Formal and symbolic computation, noncommutative algebra, polylogarithm, multiple zeta value, harmonic sum, asymptotic analysis, generating series, quadrees, maximal points