

CAHIERS DU BURO

PHILIPPE FLAJOLET

**Analyse d'algorithmes de manipulation
d'arbres et de fichiers**

*Cahiers du Bureau universitaire de recherche opérationnelle.
Série Recherche*, tome 34-35 (1981), p. 5-208

http://www.numdam.org/item?id=BURO_1981__34-35__5_0

© Institut Henri Poincaré — Institut de statistique de l'université de Paris, 1981,
tous droits réservés.

L'accès aux archives de la revue « Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

INTRODUCTION

1. Notre travail a pour thème l'évaluation d'algorithmes reposant sur l'utilisation d'arbres.

Cette structure est l'une des structures de données les plus répandues en informatique. Elle apparaît soit comme description d'objets structurés : ce sont les arbres de syntaxe des programmes dans un langage comme PASCAL, les arbres d'expressions des compilateurs, les descriptions arborescentes des objets hiérarchisés... ; soit encore elle intervient comme structure superposée à un ensemble d'enregistrements d'un fichier, permettant à la fois la recherche rapide d'une clef et l'insertion efficace de nouveaux éléments dans le fichier.

L'analyse d'un algorithme consiste généralement en l'étude des variations du coût en fonction de la taille ou du nombre des données. Nous nous sommes ici intéressés à l'estimation en moyenne des coûts en temps ou en mémoires de certains des principaux algorithmes d'arbres, en nous attachant à obtenir des expressions permettant des évaluations précises.

Il apparaît que ces estimations en moyenne sont une mesure souvent fidèle du comportement des algorithmes dans les conditions d'utilisation pratique. En effet, les analyses révèlent fréquemment une distribution "pointue" des coûts, étroitement centrée autour de la moyenne, les configurations qui conduisent aux coûts extrêmes y étant rares. L'exemple du tri par partition-échange (tri-rapide ou "quicksort") est à cet égard significatif : ce tri devient inefficace dans des cas de dégénérescence hautement improbables, et la rareté de ces cas ainsi que l'excellent comportement en moyenne de l'algorithme, en font la méthode de tri la plus efficace pour une utilisation en mémoire centrale.

2. Cette démarche en algorithmique a été bien illustrée par les travaux de synthèse de Knuth. Il s'agit dans un premier temps d'un problème combinatoire de dénombrement des configurations associées à un coût donné ; l'étape suivante est

l'évaluation des coûts moyens sous forme de sommes pondérées de ces dénombrements et l'analyse asymptotique de ces quantités fournit en général lorsqu'on retient les deux ou trois premier termes des développements des expressions simples valides même pour de petites valeurs ($n = 10$) des tailles des données.

L'analyse d'algorithme emprunte ainsi à l'analyse combinatoire pour ce qui est des dénombrements, et à l'analyse asymptotique pour ce qui est des estimations approchées. Elle représente cependant une source de problèmes spécifiques :

- au plan des dénombrements : on est conduit à une recherche relativement exhaustive de dénombrements relatifs à certaines classes d'objets combinatoires : arbres, permutations, graphes. Ce sont des domaines où l'approche par séries génératrices joue un rôle important. L'étude des algorithmes révèle parfois certains paramètres non abordés par l'analyse combinatoire classique : le paramètre d'ordre d'un arbre binaire (ou fonction nombre de registres) en est un exemple. On peut même être amené à des objets combinatoires nouveaux comme les chemins valués (ou histoires) étudiés par Françon et Viennot.

- au plan des estimations asymptotiques : ici encore, il est besoin d'une approche systématique habituellement fournie, dans le cadre de l'analyse complexe, par l'étude des singularités des séries génératrices. On y rencontre également des problèmes nouveaux comme ceux où interviennent des quantités arithmétiques liées à divers systèmes de représentation d'entiers : ces systèmes -représentation binaire, code Gray...- jouent en effet un rôle important dans la conception et l'évaluation des algorithmes, comme le montre par exemple la structure de file binômiale introduite par J. Vuillemin.

De surcroît, l'une des nécessités en algorithmique nous paraît l'élaboration de méthodes et d'outils systématiques d'analyse de larges classes d'algorithmes pour lesquels on disposerait de procédés de passage quasi-automatiques de la formulation de l'algorithme à l'estimation de sa complexité. On est certes loin d'avoir atteint cet objectif. Mais les travaux récents de Rota, Schützenberger et Foata sur les correspondances entre objets et séries génératrices nous semblent une étape particulièrement utile à l'algorithmique. Ils nous ont en tout cas sensiblement influencé (notamment dans les résultats du chapitre V et la présentation du chapitre I). De même, les résultats de Delange sur la distribution des chiffres dans la représentation binaire des entiers sont de portée plus générale et permettent plusieurs analyses, comme nous le montrons aux chapitres II et III.

Enfin, comme nous l'avons dit, notre problématique informatique s'appuie largement sur l'approche de Knuth (chapitres II et III). L'évaluation des structures de données dynamiques repose sur une méthodologie très intéressante proposée par Françon et dont l'une des contributions de notre travail est d'avoir montré qu'elle se traite naturellement par l'algèbre des fractions continues.

3. Quant à son contenu, ce travail comprend cinq chapitres qui peuvent se résumer comme suit :

Le chapitre I constitue un chapitre de rappels et d'introduction. Le caractère récursif de la structure d'arbre conduit assez rapidement à des équations de séries génératrices pour les paramètres les plus simples fournis par l'étude des algorithmes. C'est ce que nous avons essayé de mettre en évidence dans une rédaction où de nombreux résultats connus sont rappelés et présentés de manière, espérons-nous, plus unifiée que de coutume. Le résultat nouveau principal y est l'analyse complète de la détection d'arbre ("matching"), laquelle illustre bien au plan des méthodes d'analyse l'intérêt qu'il y a, dans le domaine des arbres, à se ramener à des définitions inductives sur la structure. L'utilité de l'analyse est corroborée par des résultats de simulation qui montrent que les coûts d'exécution sont étroitement centrés autour de la valeur moyenne calculée.

Le chapitre II contient une analyse détaillée du problème d'allocation de registres en compilation. Le paramètre combinatoire sous-jacent est un paramètre d'arbre déjà apparu dans plusieurs travaux de statistiques en science de la nature ; il y est étudié en moyenne et en distribution. Les lois limites, où interviennent des fonctions θ de la théorie des fonctions elliptiques, y sont ici encore très pointues.

Le chapitre III complète le chapitre précédent ; il étend à d'autres systèmes de numérotation les résultats de Delange relatifs à la représentation binaire des nombres, ce qui permet de réaliser l'analyse des réseaux de tris par fusion paire-impair et d'expliquer simplement divers phénomènes de périodicité en analyse d'algorithmes, analogues à ceux du chapitre II.

Le chapitre IV concerne l'évaluation de structures de données dynamiques -pour l'essentiel des arbres- soumises de diverses manières à des suites d'adjonctions, interrogations et suppressions. Nous montrons que les analyses se traitent de manière synthétique par l'algèbre des fractions continues et la théorie des polynômes orthogonaux ; les structures de données usuelles de l'informatique conduisent aux familles classiques de polynômes de Laguerre, Hermite, Poisson-Charlier et Meixner (Mittag-Leffler).

Le chapitre V développe les implications combinatoires du théorème d'équivalence du chapitre précédent, et fournit les éléments d'une théorie combinatoire des fractions continues de type Jacobi-Stieltjes. On en déduit notamment une interprétation des coefficients de fonctions elliptiques de Jacobi et d'inverses de polynômes orthogonaux, ainsi que diverses preuves combinatoires d'identités classiques.

La plupart des résultats de ce travail ont été publiés, ou sont en cours de publication. Certaines sections utilisent des travaux effectués en collaboration avec J. Françon, J.C. Raoult et J. Vuillemin. Nous indiquons au début de chaque chapitre les références des résultats publiés.

4. Le mémoire présenté ici constitue la thèse de doctorat de l'auteur[†]. Certaines sections utilisent des travaux effectués en collaboration avec J. Françon, J.C. Raoult et J. Vuillemin, dont les références sont indiquées au cours de l'exposé. L'auteur tient à exprimer sa gratitude à l'égard des membres du jury de thèse H. Delange, E. Gelenbe, J. Giraud, M. Nivat, G. Poitou, M.P. Schützenberger et J. Vuillemin pour de très nombreux conseils et encouragements. Sont associés à ces remerciements J. Françon, H.M. Steyaert, J.C. Raoult, G. Viennot, G. Kahn ainsi que N. Parvan pour l'excellente préparation du manuscrit.

[†] Thèse de doctorat ès-sciences, Université Paris Sud (Sept. 1979).

CHAPITRE I

DENOMBREMENTS D'ARBRES ET ANALYSE D'ALGORITHMES.

Les arbres rencontrés habituellement en informatique ressortissent à des types divers que l'on peut classer selon trois traits principaux :

A - planarité de l'arbre.

B - présence d'étiquettes distinctes.

C - contraintes ordinales sur la position des étiquettes dans l'arbre.

De nombreuses classes sont ainsi définissables par exemple en variant les conditions de type C, ou en imposant des conditions sur la forme des arbres. Nous nous intéressons, pour l'essentiel, dans ce travail aux catégories suivantes :

- arbres planaires à étiquettes indépendantes ("arbres planaires"), soit la combinaison $A^+ B^- C^-$: ces structures sont appelées simplement arbres par Knuth.
- arbres planaires monotones soit $A^+ B^+ C^+$: il s'agit notamment des arbres tournois (monotonie le long des branches) et des arbres binaires de recherche (monotonie de gauche à droite).
- arbres non planaires étiquetés ("arbres étiquetés") soit $A^- B^+ C^-$: ces classes d'arbres ont été dénombrées par Cayley qui a montré que le nombre d'arbres de taille n vaut n^{n-1} .

Comme nous l'avons déjà indiqué dans l'introduction, ces arbres apparaissent dans des contextes très divers en informatique : les arbres planaires se présentent comme arbres de syntaxe, arbres de termes ... ; les arbres monotones apparaissent surtout comme structures de données dynamiques ; les arbres étiquetés servent à la statistique des "fonctions aléatoires" (en particulier génération de nombres aléatoires et évaluation de certaines méthodes dites de Monte-Carlo) et apparaissent dans les algorithmes d'équivalences ou de recherche d'arbre

de recouvrement de graphes.

Dans ce chapitre, nous présentons d'abord les méthodes de dénombrement des principales classes d'arbres (section 1). Nous montrons ensuite deux exemples d'application : l'un à l'analyse de la détection d'arbres ou "matching" (section 2) ; l'autre à l'évaluation de divers paramètres des forêts fonctionnelles binaires utiles à l'analyse heuristique d'un algorithme de factorisation (section 3). Les principes mis en oeuvre ici sont utilisés au chapitre II (statistique sur les arbres planaires), au chapitre IV (analyse de structures de données implantées comme arbres monotones) et au chapitre V (combinatoire des arbres tournois et des permutations).

1. DENOMBREMENTS DE DIVERSES CLASSES D'ARBRES.

La plupart des structures d'arbres courantes sont susceptibles de définitions récursives ; celles-ci traduisent la construction progressive des arbres à partir de leurs sous-arbres en commençant avec certains ensembles initiaux. Les constructions ont fréquemment un caractère "non contextuel" en ce sens qu'elles s'effectuent avec une information très limitée sur les sous-arbres. Cette situation favorable en analyse combinatoire, permet pour les paramètres essentiels de traduire de manière systématique les définitions inductives en équations fonctionnelles sur les séries génératrices de dénombrement.

Nous nous contentons ici d'indications intuitives sur la méthode, que nous illustrerons par divers exemples.

Supposons qu'on cherche à dénombrer un ensemble d'objets combinatoires $\mathcal{O}$ sur lequel est défini une notion naturelle de taille. Les objets -typiquement des arbres- sont construits à partir d'un ensemble de base au moyen de constructeurs $K_1, K_2, K_3 \dots$.

Il est tout d'abord commode de considérer les constructeurs comme opérant non sur les sous-ensembles de $\mathcal{O}$ mais sur les multiensembles[†] de $\mathcal{O}$. Soient $\mathcal{E}$ et $\mathcal{F}$ deux multiensembles d'éléments de $\mathcal{O}$

$$\mathcal{E} = \sum_{\omega \in \mathcal{O}} e_{\omega} \cdot \omega \quad \text{et} \quad \mathcal{F} = \sum_{\omega \in \mathcal{O}} f_{\omega} \cdot \omega$$

notés additivement ; une application, par exemple binaire, $K : \mathcal{O} \times \mathcal{O} \rightarrow \mathcal{O}$ s'étend aux multiensembles en posant

$$K(\mathcal{E}, \mathcal{F}) = \sum_{\omega, \omega'} e_{\omega} f_{\omega'} K(\omega, \omega').$$

Aux objets de $\mathcal{O}$ se trouve habituellement associée une notion naturelle de taille notée $|\cdot|$ qui est une application de $\mathcal{O}$ dans $\mathbb{N}$. Etant donné un multiensemble $\mathcal{E}$, on désigne par E_n le nombre d'objets de $\mathcal{E}$ de taille n , comptés chacun avec leur multiplicité et l'on s'intéresse à la série génératrice $E(z) = \sum_{n \geq 0} E_n z^n$.

[†] Un multiensemble est un ensemble d'éléments chacun doté d'un coefficient de multiplicité pris dans $\mathbb{Q}$. Nous renvoyons à [Eilenberg; 1974] pour des définitions classiques et une utilisation voisine de la nôtre. Nous notons $M(\mathcal{O}; K)$ la famille des multiensembles de $\mathcal{O}$ à coefficients dans K

Appelons α la correspondance qui associe à chaque multiensemble $\mathcal{E}$ la série $E(z)$:

$$\alpha : M(\mathcal{O}; K) \rightarrow \mathbb{C}[[z]].$$

Il est de fait que pour les objets de l'analyse combinatoire usuels (arbres, permutations, partitions...) chaque constructeur de base possède une image par α , c'est-à-dire que pour le constructeur K , par exemple binaire, existe une fonctionnelle ϕ telle que le diagramme suivant commute

$$\begin{array}{ccc} M(\mathcal{O}; \mathbb{C}) \times M(\mathcal{O}; \mathbb{C}) & \xrightarrow{K} & M(\mathcal{O}; \mathbb{C}) \\ \alpha \downarrow & & \alpha \downarrow \\ \mathbb{C}[[z]] \times \mathbb{C}[[z]] & \xrightarrow{\phi} & \mathbb{C}[[z]]. \end{array}$$

Les fonctionnelles images des constructeurs de base sont généralement simples et typiques de la classe d'objets (opérateurs algébriques pour les arbres planaires ; opérateurs intégraux pour les arbres tournois...). Les définitions récursives des objets se traduisent alors en équations fonctionnelles sur les séries génératrices de dénombrements.

Une bonne part de l'analyse combinatoire élémentaire peut être présentée dans ce cadre ; en particulier le cas où $\mathcal{O}$ est un monoïde libre a été largement étudié (cf. [Chomsky-Schützenberger ; 1963], [Eilenberg ; 1974], [Berstel ; 1978], [Salomaa, Soittola ; 1978])[†].

Cette approche présente pour nous un certain nombre d'avantages :

- de nombreuses sous-classes intéressantes s'obtiennent en limitant l'ensemble des constructions permises ou leur ordre d'applications (on pourrait appeler ces sous-classes "régulières" ou "rationnelles" par analogie avec la théorie des langages formels), et l'on obtient facilement les équations fonctionnelles correspondantes : c'est notamment le cas des classes d'arbres auxquelles on impose des limitations sur le degré des sommets.
- Les paramètres simples des structures sont habituellement liés au nombre d'applications de chaque constructeur dans la structure. On obtient ainsi des séries génératrices de dénombrement : soit en marquant par une variable l'appli-

[†] Pour d'autres types de structures combinatoires, cf. [Foata, Schützenberger ; 1970], [Foata ; 1971] et [Rota ; 1975].

cation de chaque constructeur ; soit en limitant le nombre de ces applications, ce qui équivaut à la résolution par approximations successives d'une équation au point fixe. Dans le cas des arbres, on traite de la sorte les dénombrements relatifs au nombre de feuilles, à la longueur de cheminement, à la hauteur... , ce d'une manière largement indépendante de la classe sur laquelle on opère.

Enfin, les séries génératrices obtenues sont aussi des fonctions. Ceci permet de leur appliquer des méthodes assez générales de l'analyse complexe :

- détermination des coefficients des séries entières par le théorème de Cauchy ; forme intégrale des coefficients des séries génératrices de Dirichlet.
- évaluations asymptotiques par localisation des singularités et méthode de Darboux pour les fonctions définies implicitement ; utilisation de la méthode de col pour les coefficients des séries entières.

1.1. Dénombrement d'arbres planaires.

Un arbre planaire est un arbre dessiné et planté au sens de la théorie des graphes : chaque sommet possède une suite de descendants ou fils (l'ordre de succession des fils étant figuré de gauche à droite dans le plan) ; enfin on distingue une racine dans l'arbre (représentée en haut de l'arbre). Notre terminologie suit celle de [Knuth ; 1968].

La figure 1 représente les arbres planaires comportant moins de 5 sommets.

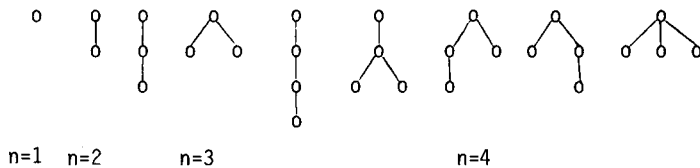
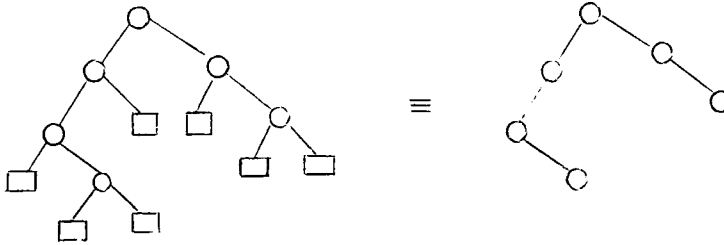


Figure 1 : Arbres formés de n sommets pour $n \leq 4$.

Les arbres binaires (planaires) sont tels que chaque sommet comporte soit 0 descendant (sommet externe) soit 2 descendants (sommet interne). Les sommets externes sont habituellement représentés par des carrés et l'on omet parfois de les faire figurer lorsqu'il n'y a pas ambiguïté. Voici un exemple de ces deux

représentations d'un arbre binaire :



La figure 2 donne le tableau des arbres binaires de taille inférieure à 4

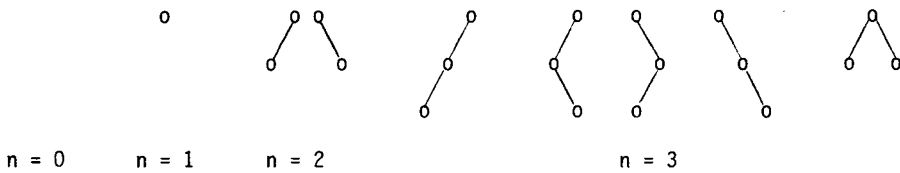


Figure 2 : Les arbres binaires à n sommets internes pour $n \leq 3$.

Nous commençons par le dénombrement des arbres binaires, suivi de celui des arbres (planaires) des plus généraux. Nous traitons avec un certain détail le cas des arbres binaires, afin d'illustrer la traduction des définitions récursives en équation sur les séries génératrices.

Arbres binaires :

La classe $\mathcal{B}$ des arbres binaires est définissable récursivement par l'équation

$$\mathcal{B} = \square + \begin{array}{c} \circ \\ \swarrow \quad \searrow \\ \mathcal{B} \quad \mathcal{B} \end{array}, \quad (*)$$

qui traduit qu'un arbre binaire est soit l'arbre formé d'un seul sommet externe (et de zéro sommet interne) encore appelé arbre vide, soit formé d'une racine et de deux arbres qui sont les sous-arbres gauche et droit de la racine. On peut écrire cette équation sous la forme

$$\mathcal{B} = K_0 + K_2(\mathcal{B}, \mathcal{B}) \quad (**)$$

où K_0 est le constructeur (d'arité zéro) dont la valeur est l'arbre vide, et où

$K_2(\mathcal{C}, \mathcal{F})$ consiste à construire de toutes les manières possibles des arbres dont les sous-arbres gauches sont dans $\mathcal{C}$ et les sous-arbres droits dans $\mathcal{F}$.

En d'autres termes, avec l'extension aux multi-ensembles :

$$K_2(\mathcal{C}, \mathcal{F}) = \sum_{\substack{s \in \mathcal{C} \\ t \in \mathcal{F}}} e_s f_t \begin{array}{c} \bigcirc \\ \swarrow \quad \searrow \\ s \quad t \end{array}.$$

La définition (*) est non ambiguë et l'équation (**) reste valide au sens des multi-ensembles.

Par ailleurs, pour les arbres binaires, on adopte généralement comme mesure de taille le nombre de sommets internes.

Soient $E(z)$ et $F(z)$ les séries génératrices associées[†] aux multi-ensembles $\mathcal{C}$ et $\mathcal{F}$; à $G = K_2(\mathcal{C}, \mathcal{F})$ se trouve associée la série $G(z)$ et l'on a l'égalité

$$G(z) = z E(z) F(z).$$

Cette propriété traduit simplement la relation évidente

$$G_n = \sum_{n_1+n_2+1=n} E_{n_1} F_{n_2}.$$

On a ainsi le diagramme commutatif :

$$\begin{array}{ccc} \mathcal{B} & \times & \mathcal{B} \\ \downarrow \alpha & & \downarrow \alpha \\ \mathbb{C}[[z]] & \times & \mathbb{C}[[z]] \end{array} \xrightarrow{K_2} \begin{array}{c} \mathcal{B} \\ \downarrow \alpha \\ \mathbb{C}[[z]] \end{array}$$

$$\mathbb{C}[[z]] \times \mathbb{C}[[z]] \xrightarrow{\phi} \mathbb{C}[[z]]$$

où $\phi(E(z), F(z)) = z E(z) F(z)$.

L'existence d'image pour ces constructeurs d'arbres conduit pour la série génératrice des arbres binaires $B(z) = \sum_{n \geq 0} B_n z^n$ où B_n est le nombre d'arbres binaires de taille n , à l'équation

$$B(z) = 1 + z B(z).B(z),$$

$$\text{soit } B(z) = \frac{1 - \sqrt{1-4z}}{2z} \quad \text{et} \quad B_n = \frac{1}{n+1} \binom{2n}{n}.$$

[†] Rappelons que E_n, F_n désignent le nombre d'objets de $\mathcal{C}$ resp. $\mathcal{F}$ de taille n et que $E(z) = \sum_{n \geq 0} E_n z^n$ $F(z) = \sum_{n \geq 0} F_n z^n$ sont les séries génératrices associées.

Les nombres B_n sont les nombre de Catalan. Ils sont apparus pour la première fois dans le décompte des triangulations de **polygones** par Euler et Segner.

Le paramètre hauteur (mesuré en nombre de sommets internes sur la branche la plus longue) correspond au nombre d'applications du constructeur K_2 . Soit $\mathcal{B}^{[h]}$ l'ensemble des arbres de hauteur $\leq h$. On peut écrire la suite d'égalités :

$$\begin{aligned}\mathcal{B}^{[0]} &= K_0 \\ \mathcal{B}^{[1]} &= K_2(\mathcal{B}^{[0]}, \mathcal{B}^{[0]}) \\ &\vdots \\ \mathcal{B}^{[h]} &= K_2(\mathcal{B}^{[h-1]}, \mathcal{B}^{[h-1]}),\end{aligned}$$

soit pour les séries énumératrices $B^{[h]}(z)$ qui sont ici des polynômes :

$$\begin{aligned}B^{[0]} &= 1 \\ B^{[1]} &= 1+z(B^{[0]})^2 \\ B^{[2]} &= 1+z(B^{[1]})^2 \\ &\vdots \\ B^{[h]} &= 1+z(B^{[h-1]})^2.\end{aligned}$$

On peut également marquer l'application de chaque constructeur au niveau des séries génératrices. Ainsi soit $c_{j_1, j_2, \dots, j_h}^{[h]}$ le nombre d'arbres de hauteur $\leq h$ ayant j_1 sommet à hauteur 1, j_2 sommets à hauteur 2 ... et soit $C^{[h]}(u_1, u_2, \dots, u_h)$ la série génératrice correspondante ; les $C^{[h]}$ vérifient les récurrences

$$\begin{aligned}C^{[0]}(z) &= 1 ; C^{[1]}(z ; u_1) = 1 + zu_1(C^{[0]}(z))^2 ; \\ C^{[2]}(z, u_1, u_2) &= 1 + zu_1(C^{[1]}(z ; u_2))^2 ; C^{[3]}(z, u_1, u_2, u_3) = 1 + zu_1(C^{[2]}(z ; u_2, u_3))^2.\end{aligned}$$

Ainsi la série génératrice des arbres binaires partagés selon le nombre de sommets à niveau 1, 2, 3, ... :

$$C(z ; u_1, u_2, \dots) = \lim_{h \rightarrow \infty} C^{[h]}(z ; u_1, u_2, \dots, u_h),$$

possède l'expression en "forme quadratique continue" :

$$C(z ; u_1, u_2, u_3, \dots) = 1 + zu_1(1 + zu_2(1 + zu_3(\dots)^2)^2)^2.$$

Soit enfin le paramètre longueur de cheminement. La longueur de cheminement interne d'un arbre (ici) est définie comme la somme des profondeurs des sommets internes.

Ce paramètre est définissable inductivement

$$lci(\square) = 0 \quad lci\left(\begin{array}{c} \circ \\ s \quad t \end{array}\right) = 1 + |s| + |t| + lci(s) + lci(t).$$

Introduisons le multi-ensemble

$$\mathcal{L} = \sum_{s \in \mathcal{B}} lci(s) \cdot s.$$

La définition inductive de lci se traduit par l'équation :

$$\mathcal{L} = \begin{array}{c} \circ \\ \mathcal{L} \quad \mathcal{B} \end{array} + \begin{array}{c} \circ \\ \mathcal{B} \quad \mathcal{L} \end{array} + \sum_{s \in \mathcal{B}} |s| \cdot s,$$

soit pour la série génératrice associée :

$$L(z) = 2z B(z) L(z) + \sum_{n \geq 0} n B_n z^n.$$

La résolution montre les égalités

$$L(z) = (1 - 2z B(z))^{-1} \cdot z \frac{d}{dz} B(z) = \frac{1}{1-4z} - \frac{1}{2z\sqrt{1-4z}} + \frac{1}{2z},$$

d'où pour la longueur de cheminement relative à l'ensemble des arbres de taille n , l'expression classique

$$L_n = 4^n - (2n+1)B_n.$$

Remarque : On obtient facilement par l'approximation de Stirling des factorielles les expressions asymptotiques

$$B_n = \frac{4^n}{\sqrt{\pi} n^{3/2}} (1 + O(\frac{1}{n})) \quad \text{et} \quad \frac{L_n}{B_n} = n\sqrt{\pi} + O(n).$$

Le paramètre $\frac{L_n}{n B_n}$ représente l'occupation moyenne de mémoire nécessaire à l'exploration d'un arbre de taille n .

Arbres planaires généraux :

La classe $\mathcal{G}$ des arbres sans contrainte sur le degré des sommets est définie par l'équation

$$\mathcal{G} = \circ + \begin{array}{c} \circ \\ \mathcal{G} \end{array} + \begin{array}{c} \circ \\ \mathcal{G} \quad \mathcal{G} \end{array} + \begin{array}{c} \circ \\ \mathcal{G} \quad \mathcal{G} \quad \mathcal{G} \end{array} + \dots$$

qu'on peut encore écrire

$$\mathcal{G} = K_0 + K_1(\mathcal{G}) + K_2(\mathcal{G}, \mathcal{G}) + K_3(\mathcal{G}, \mathcal{G}, \mathcal{G}) + \dots,$$

où le constructeur K_s est défini par

$$K_s(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_s) = \begin{array}{c} \circ \\ \mathcal{G}_1 \quad \mathcal{G}_2 \quad \dots \quad \mathcal{G}_s \end{array}.$$

La définition est non ambiguë et vaut pour les multi-ensembles. La taille d'un

arbre étant le nombre de sommets dans l'arbre, la série génératrice associée $G(z)$ vérifie l'équation

$$G(z) = z + z G(z) + z(G(z))^2 + z(G(z))^3 + \dots,$$

car, si $\mathcal{F} = K_s(\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_s)$, on a l'égalité :

$$F(z) = z E_1(z) E_2(z) \dots E_s(z).$$

On obtient la relation homographique

$$G(z) = \frac{z}{1-G(z)},$$

qui donne, en résolvant :

$$G(z) = \frac{1 - \sqrt{1-4z}}{2} \quad \text{et} \quad G_n = \frac{1}{n} \binom{2n-2}{n-1}.$$

On retrouve ainsi les nombres de Catalan : $G_n = B_{n-1}$.

On peut reprendre l'équation donnant $\mathcal{G}$ et marquer par une variable l'application du constructeur K_0 . Ceci permet de construire la série génératrice des arbres partitionnés selon le nombre de feuilles. Soit $G_{n,r}$ le nombre d'arbres de taille n comportant r feuilles et soit $G(z,u) = \sum_{n,r \geq 0} G_{n,r} u^r z^n$ la série génératrice correspondante. L'équation de définition

$$\mathcal{G} = K_0 + K_1(\mathcal{G}) + K_2(\mathcal{G}, \mathcal{G}) + \dots$$

se traduit, en marquant par u l'opérateur K_0 , en l'équation

$$G(z,u) = zu + zG(z,u) + z(G(z,u))^2 + \dots$$

soit en résolvant

$$G(z,u) = \frac{1 - z(1-u) - \sqrt{1 - z^2(1-u)^2 - 2z(1+u)}}{2}.$$

Naturellement $G(z,1) = G(z)$ et le développement en série de $G(z,u)$ montre (en utilisant par exemple le théorème d'inversion de Lagrange) l'expression

$$G_{n,r} = \frac{1}{n} \binom{n}{r} \binom{n-2}{r-1}.$$

Les nombres $G_{n,r}$ sont appelés nombres de Narayana par [Kreweras ; 1970] .

Soit $\mathcal{g}^{[h]}$ l'ensemble des arbres de hauteur $\leq h$ (la hauteur étant mesurée en nombre d'arêtes sur la branche la plus longue) ; les ensembles $\mathcal{g}^{[h]}$ vérifient les égalités :

$$\begin{aligned} g^{[0]} &= K_0 \\ g^{[1]} &= K_0 + K_1(g^{[0]}) + K_2(g^{[0]}, g^{[0]}) + \dots \\ &\dots \end{aligned}$$

$$g^{[h]} = K_0 + K_1(g^{[h+1]}) + K_2(g^{[h-1]}, g^{[h-1]}) + \dots,$$

ce qui se traduit par la récurrence

$$G^{[0]}(z) = z ; G^{[h]}(z) = \frac{z}{1 - G^{[h-1]}(z)} .$$

On résout cette récurrence homographique et l'on trouve

$$G^{[h]}(z) = z \frac{F_{h+1}(z)}{F_{h+2}(z)} ,$$

où les polynômes F_h sont donnés par

$$F_0(z) = 0 \quad F_1(z) = 1 \quad F_{h+2} = F_{h+1}(z) - zF_h(z) .$$

Nous retrouverons ces polynômes au chapitre II où ils apparaissent dans l'énumération des arbres selon le nombre de registres. L'identité des $G^{[h]}$ avec les réduites de la fraction continue

$$\frac{z}{1 - \frac{z}{1 - \frac{z}{1 - \frac{z}{\dots}}}}$$

a été observée par [Knuth, de Bruijn, Rice ; 1972] et sera généralisée et appliquée aux chapitres IV, V.

Remarque : 1. On traite de la même manière les dénombrements de classes d'arbres définies par des contraintes sur le degré des sommets. Par exemple l'équation de la série génératrice $P(z)$ des arbres dont tous les sommets ont degré (nombre de descendants) pair est

$$P(z) = z + z P^2(z) + z P^4(z) + \dots = \frac{z}{1 - P^2(z)} .$$

2. On traite également de cette façon les dénombrements d'arbres étiquetés tels que toutes les étiquettes soient indépendantes. Il suffit d'introduire autant d'opérateurs de construction qu'il y a d'étiquettes possibles. $\square$

Construction d'arbres par substitution aux feuilles :

La construction précédente des arbres correspond à une croissance "par le haut" : on construit des arbres par réunion de sous-arbres et adjonction d'une

racine. Il est possible également de construire les arbres "par le bas" ; ce procédé de bourgeonnement consiste à substituer aux feuilles des sous-arbres de hauteur ≤ 1 .

Soit $S(\alpha, \beta)$ l'opérateur de substitution sur les arbres qui consiste à greffer les arbres de la famille β sur les feuilles des arbres de la famille α . Cet opérateur possède une image sur les séries formelles : soit $A(z, u)$ la série génératrice des arbres de la famille A partitionnés par nombre de feuilles :

$A(z, u) = \sum_{n, r \geq 0} A_{n, r} u^r z^n$ où $A_{n, r}$ est le nombre d'arbres de A de taille $n+r$ ayant r feuilles ; soit $B(z)$ la série génératrice des arbres de β et $C(z)$ la série génératrice associée à $\mathcal{C} = S(\alpha, \beta)$. La relation

$$\mathcal{C} = S(\alpha, \beta)$$

se traduit par l'égalité

$$C(z) = A(z, B(z)).$$

Ainsi, partant d'une définition par substitution des arbres généraux[†]:

$$\mathcal{G} = \square + S(\mathcal{G}, \mathcal{A}) \quad \text{où} \quad \mathcal{A} = \square + \begin{array}{c} \square \\ | \\ \square \end{array} + \begin{array}{c} \square \quad \square \\ | \quad | \\ \square \end{array} + \begin{array}{c} \square \quad \square \quad \square \\ | \quad | \quad | \\ \square \end{array} + \dots,$$

obtient-on pour la série $r(z, u)$ correspondante l'équation fonctionnelle :

$$r(z, u) = zu + r(z, \frac{z}{1-u}).$$

Ces définitions "par invariants" permettent parfois des calculs plus simples. Elles sont notamment utiles aux dénombrements d'arbres équilibrés.

Soit en effet $\mathcal{E}$ la famille des arbres 2-3 équilibrés, c'est-à-dire la famille des arbres dont chaque sommet a degré 2 ou 3 et dont toutes les feuilles sont à même profondeur. Ces arbres apparaissent comme support ([Aho, Hopcroft, Ullman ; 1974]) de plusieurs structures de données en informatique. La classe vérifie l'équation de substitution :

$$\mathcal{E} = \square + S(\mathcal{E}, \mathcal{A}) \quad \text{où} \quad \mathcal{A} = \begin{array}{c} \square \quad \square \\ | \quad | \\ \square \end{array} + \begin{array}{c} \square \quad \square \quad \square \\ | \quad | \quad | \\ \square \end{array}.$$

Si l'on marque par z les sommets internes et par u les sommets externes, on obtient l'équation donnant la série génératrice $E(z, u) = \sum E_{m, r} u^r z^m$ où $E_{m, r}$ est le nombre d'arbres 2-3 équilibrés à m sommets internes et r sommets externes :

$$E(z, u) = u + E(z, z(u^2 + u^3)).$$

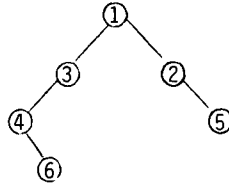
[†] L'opérateur S effectue la substitution aux feuilles marquées $\square$; ce détour est nécessaire à l'obtention d'une définition non ambiguë de $\mathcal{G}$.

Enfin, si l'on désigne par F_r le nombre d'arbres ayant r sommets externes $F_r = \sum_m E_{m,r}$, la série génératrice $F(u) = \sum F_r u^r$ vérifie l'équation fonctionnelle $F(u) = u + F(u^2 + u^3)$.

L'évaluation asymptotique des coefficients de $F(u)$ présente un problème analytique difficile résolu par [Odlyzko ; 1979].

1.2. Arbres monotones et permutations.

Un arbre tournoi est un arbre binaire dont les sommets internes sont étiquetés par des entiers consécutifs commençant à 1, de sorte que l'étiquetage soit croissant sur chaque branche. Par exemple :



La figure 3 donne le tableau des arbres tournois de taille inférieure à 4.

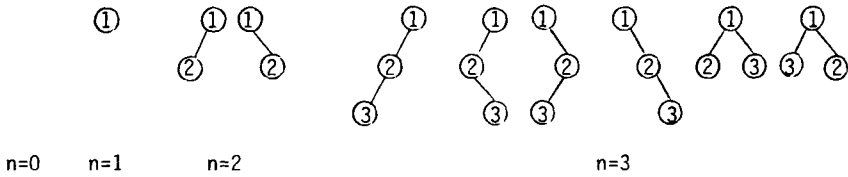
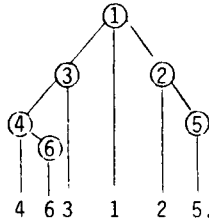
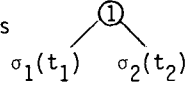


Figure 3 : Les arbres tournois de taille n pour $n \leq 3$.

Les arbres tournois sont une présentation commode des permutations : on passe d'un arbre tournoi à une permutation en lisant les étiquettes en ordre infixe, (symétrique), ou de manière équivalente en les projetant sur un axe horizontal :



Réciproquement on obtient l'arbre tournoi associé à une permutation π en factorisant la permutation, considérée comme un mot, en $\pi = \pi_g 1 \pi_d$, en plaçant 1 à la racine avec comme sous-arbre gauche le tournoi associé récursivement à π_g et comme sous-arbre droit le tournoi associé à π_d .

La construction des arbres tournois est simple : le constructeur L_0 construit l'arbre tournoi vide ; le constructeur L_2 associe à deux arbres tournois (t_1, t_2) le multi-ensemble des arbres tournois s obtenus en formant l'arbre $t =$  et en distribuant les étiquettes de $[2..|t_1|+|t_2|+1]$ dans les sous-arbres gauche et droit de toutes les manières qui préservent les relations d'ordre de t_1 et t_2 . En d'autres termes, soit $n_1 = |t_1|$, $n_2 = |t_2|$ et $n = n_1 + n_2 + 1$; $L_2(t_1, t_2)$ consiste en l'ensemble des arbres  où

- σ_1 est une application croissante de $[1..n_1]$ dans $[2..n]$
- σ_2 est une application croissante de $[1..n_2]$ dans $[2..n]$
- $\text{Im}(\sigma_1) \cup \text{Im}(\sigma_2) = [2..n]$,

les applications σ_1, σ_2 étant étendues aux arbres en opérant par renommage des étiquettes.

L'ensemble $\mathcal{C}$ des arbres tournois vérifie l'équation

$$\mathcal{C} = L_0 + L_2(\mathcal{C}, \mathcal{C})$$

encore valable en tant qu'équation sur les multi-ensembles. Cette équation se réécrit

$$\begin{aligned} \mathcal{C} &= L_0 + L_2(L_0 + L_2(\mathcal{C}, \mathcal{C}), L_0 + L_2(\mathcal{C}, \mathcal{C})) \\ &= L_0 + L_2(L_0, L_2(\mathcal{C}, \mathcal{C})) + L_2(L_2(\mathcal{C}, \mathcal{C}), L_0) + L_2(L_2(\mathcal{C}, \mathcal{C}), L_2(\mathcal{C}, \mathcal{C})), \end{aligned}$$

soit encore avec $\mathcal{U} = L_2(\mathcal{C}, \mathcal{C}) = \mathcal{C} - L_0$ l'ensemble des arbres tournois de taille ≥ 1 :

$$\mathcal{U} = L + L_g(\mathcal{U}) + L_d(\mathcal{U}) + L_2(\mathcal{U}, \mathcal{U})$$

où $L = L_2(L_0, L_0)$; $L_g(A) = L_2(A, L_0)$ et $L_d(A) = L_2(L_0, A)$.

Le constructeur L_2 correspond sur les séries génératrices exponentielles à l'intégrale d'un produit de Cauchy. En effet, soit $\mathcal{C} = L_2(\mathcal{A}, \mathcal{B})$; la taille d'un tournoi étant le nombre de ses sommets internes (étiquetés), avec les notations usuelles,

on a l'égalité

$$C_n = \sum_{n_1+n_2+1=n} \binom{n-1}{n_1, n_2} A_{n_1} B_{n_2},$$

dans laquelle le multinomial traduit le nombre de manière de répartir les étiquettes entre sous-arbres gauche et droit. De la sorte, entre les séries génératrices exponentielles $C(z) = \sum_{n \geq 0} C_n \frac{z^n}{n!}$, $A(z) = \sum_{n \geq 0} A_n \frac{z^n}{n!}$ et $B(z) = \sum_{n \geq 0} B_n \frac{z^n}{n!}$ on a la relation

$$C(z) = \int_0^z A(z) B(z) dz.$$

L'image de L_0 est simplement la constante 1.

On peut ainsi effectuer les dénombrements par des méthodes très voisines de celles du paragraphe précédent. La série génératrice des tournois vérifie l'équation

$$T(z) = 1 + \int_0^z T^2(z) dz$$

ce qui équivaut à :

$$\frac{dT(z)}{dz} = T^2(z) ; T(0) = 1$$

d'où le résultat attendu

$$T(z) = \frac{1}{1-z}.$$

Il est facile de voir que les points doubles, points simples à gauche, points simples à droite et feuilles de tournois binaires correspondent respectivement aux pics (P), doubles descentes (DD), doubles montées (DM) et creux (Q) de la permutation. Soit $V_{j,k,l,m}$ le nombre de permutations ayant j points P, k points DD, l points DM et m points Q, et soit

$$V(z) = V(z; u_0, u_1, \bar{u}_1, u_2) = \sum_{j,k,l,m \geq 0} V_{j,k,l,m} u_0^j u_1^k \bar{u}_1^l u_2^m \frac{z^{j+k+l+m}}{(j+k+l+m)!}$$

la série génératrice correspondante. L'équation donnant V s'obtient en marquant par $u_0, u_1, \bar{u}_1, u_2$ les applications des constructeurs L, L_g, L_d, L_2 dans la définition équationnelle de $\mathcal{U}$. Ainsi :

$$V(z) = u_0 z + (u_1 + \bar{u}_1) z \int_0^z V(z) dz + u_2 z \int_0^z V^2(z) dz.$$

On résout l'équation différentielle qui est à variables séparées :

$$\frac{dV(z)}{dz} = u_0 + (u_1 + \bar{u}_1) V(z) + u_2 V^2(z),$$

d'où la solution

$$V(z) = \delta \operatorname{tg}(z\delta) - \frac{u_1 + \bar{u}_1}{2} \quad \text{avec} \quad \delta = \frac{1}{2} \sqrt{4u_0u_2 - (u_1 + \bar{u}_1)^2}$$

En particulier le nombre de permutations alternantes (sans double montées, ni double descentes) sur $[2n+1]$, soit E_{2n+1} , a pour série génératrice

$$\sum_{n \geq 0} E_{2n+1} \frac{z^{2n+1}}{(2n+1)!} = \operatorname{tg} z,$$

comme on le voit en posant $u_0 = u_2 = 1$, $u_1 = \bar{u}_1 = 0$. On démontre de la même manière que le nombre de permutations alternantes sur $[2n]$ a pour série génératrice exponentielle $\sec z = \frac{1}{\cos z}$. Ceci constitue le théorème classique d'André.

Nous retrouverons ces nombres au chapitre IV (dénombrement de suites d'opérations) et au chapitre V où nous montrerons que les séries génératrices ordinaires correspondantes ont des développements en fraction continue de forme simple.

On traite de la même manière que précédemment les évaluations des paramètres hauteur et longueur de cheminement. Soit $\mathcal{C}^{[h]}$ la famille des tournois de hauteur $\leq h$ et soit $T^{[h]}$ la série exponentielle correspondante. Les polynômes $T^{[h]}$ vérifient la récurrence

$$T^{[0]}(z) = 1; \quad T^{[h+1]}(z) = 1 + \int_0^z (T^{[h]}(z))^2 dz,$$

relation qui en permet le calcul de proche en proche.

Enfin, soit $lci(t)$ la longueur de cheminement interne du tournoi t et $\mathcal{C}$ le multi-ensemble des tournois comptés avec, comme multiplicité, leur longueur de cheminement :

$$\mathcal{C} = \sum_{t \in \mathcal{C}} lci(t) \cdot t.$$

Comme dans le cas binaire non étiqueté, on a l'équation :

$$\mathcal{C} = L_2(\mathcal{C}, \mathcal{C}) + L_2(\mathcal{C}, \mathcal{L}) + D\mathcal{C}$$

$$\text{où } D(\mathcal{C}) = \sum_{t \in \mathcal{C}} |t| \cdot t.$$

Passant aux séries génératrices exponentielles, l'image $C(z)$ de $\mathcal{C}$ est donnée par l'équation

$$C(z) = 2 \int_0^z C(u) \frac{du}{1-u} + \frac{z}{(1-z)^2},$$

soit en dérivant

$$\frac{dC(z)}{dz} = 2 \frac{C(z)}{1-z} + \frac{1+z}{(1-z)^3},$$

et finalement

$$C(z) = \frac{1}{(1-z)^2} [2 \ln\left(\frac{1}{1-z}\right) - z] = z + 3z^2 + \frac{17}{3} z^3 + \dots,$$

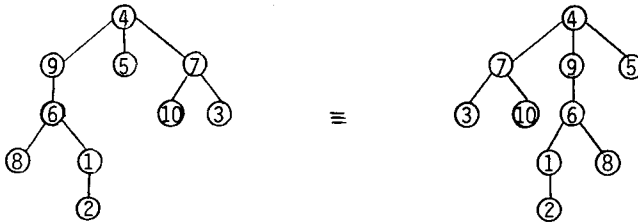
d'où le résultat classique

$$\frac{C_n}{n!} = 2(n+1)H_n - 3n.^\dagger$$

Ce résultat est utile à plusieurs analyses d'algorithmes : à celle des arbres binaires de recherche, ou à celles des implantations des files de priorité sous forme d'arbres tournois (cf. [Knuth ; 1973], [Fraçon, Viennot, Vuillemin ; 1978] [Fraçon ; 1979]).

1.3. Arbres étiquetés.

Nous considérons ici l'ensemble $\mathcal{E}$ des arbres non planaires enracinés et étiquetés par les entiers consécutifs 1,2,3,... et dont toutes les étiquettes sont distinctes. Ainsi, les deux représentations ci-dessous sont représentations équivalentes d'un même arbre (non planaire) étiqueté :



La figure 4 représente l'ensemble des arbres étiquetés de taille inférieure à 4.

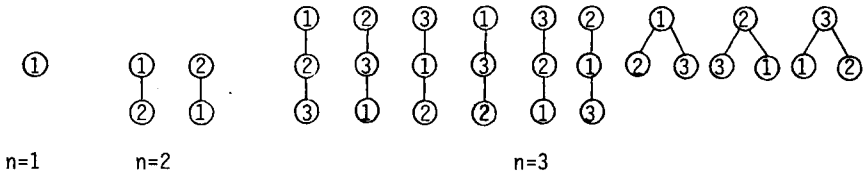


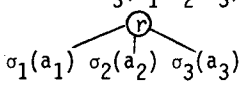
Figure 4 : Les arbres étiquetés de taille n , pour $n \leq 3$.

On introduit comme dans le cas des arbres tournois des constructeurs qui distribuent les étiquettes. Puisqu'il n'y a plus de contrainte de monotonie, ces

[†] Nous utilisons ici la notation classique des nombres harmoniques :

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}.$$

constructeurs distribuent également une étiquette à la racine de l'arbre. Par exemple, le constructeur ternaire des arbres étiquetés $M_3(a_1, a_2, a_3)$ est défini comme suit :

$M_3(a_1, a_2, a_3)$ est l'ensemble des arbres étiquetés de la forme

 où, avec $n_1 = |a_1|$, $n_2 = |a_2|$, $n_3 = |a_3|$, $n = 1 + n_1 + n_2 + n_3$:
 - $r \in [1..n]$
 - σ_j est une application croissante de $[1..n_j]$ dans $[1..n]$ pour $j = 1, 2, 3$.
 - $\text{Im}(\sigma_1) \oplus \text{Im}(\sigma_2) \oplus \text{Im}(\sigma_3) \oplus \{r\} = [1..n]$;

les applications σ sont étendues aux arbres par renommage des étiquettes.

Pour chaque degré s , on définit ainsi un constructeur M_s ; les M_s sont étendus aux multi-ensembles par linéarité. Il importe de remarquer qu'à cause de la non-planarité, le multi-ensemble $M_s(\mathcal{C}, \mathcal{C}, \dots, \mathcal{C})$ représente la famille des arbres étiquetés dont le degré de la racine est s , chaque arbre étant compté avec la multiplicité $s!$ (chaque arbre est engendré autant de fois qu'il y a de manières de permuter les sous-arbres issus de la racine). Ainsi l'ensemble $\mathcal{C}$ vérifie-t-il l'équation

$$\mathcal{C} = M_0 + \frac{1}{1!} M_1(\mathcal{C}) + \frac{1}{2!} M_2(\mathcal{C}, \mathcal{C}) + \frac{1}{3!} M_3(\mathcal{C}, \mathcal{C}, \mathcal{C}) + \dots$$

Chaque constructeur M_s possède une image simple ; soit $C = M_s(A_1, A_2, \dots, A_s)$ et soient C_n et $A_{j,n}$ les nombres d'arbres de taille n (mesurée en nombre de sommets) des familles $\mathcal{B}$ et A_j pour $1 \leq j \leq s$. On a la récurrence

$$C_n = \sum_{n_1 + n_2 + \dots + n_s + 1 = n} \binom{n}{n_1, n_2, \dots, n_s} A_{1, n_1} A_{2, n_2} \dots A_{s, n_s}.$$

Le multinomial de la somme traduit le nombre de façons de distribuer les éléments en 1 élément à la racine, en n_1 éléments dans les arbres de A_1 de taille $n_1, \dots$.

Passant aux séries génératrices exponentielles :

$$B(z) = \sum_{n \geq 0} B_n \frac{z^n}{n!} \quad \text{et} \quad A_j(z) = \sum_{n \geq 0} A_{j,n} \frac{z^n}{n!},$$

on obtient la relation

$$B(z) = z A_1(z) A_2(z) \dots A_s(z).$$

La série génératrice exponentielle des arbres étiquetés satisfait donc l'équation

$$E(z) = \sum E_n \frac{z^n}{n!} = z + z E(z) + \frac{z}{2!} E^2(z) + \frac{z}{3!} E^3(z) + \dots = z \exp E(z).$$

Le théorème d'inversion de Lagrange permet de résoudre cette équation et donne comme valeur du coefficient $\frac{E_n}{n!}$

$$\frac{E_n}{n!} = \frac{1}{n} \text{coeff}(u^{n-1} ; (\exp u)^n) = \frac{1}{n} \frac{n^{n-1}}{(n-1)!}$$
$$E_n = n^{n-1}$$

ce qui constitue le résultat classique de Cayley.

Nous verrons à la section 3 l'application de ce type de dénombrement à l'analyse "heuristique" d'un algorithme de factorisation.

2. LA RECHERCHE DE MOTIFS DANS LES ARBRES.

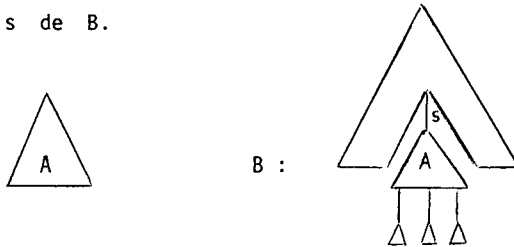
Nous présentons ici brièvement l'analyse de l'algorithme séquentiel[†] classique de recherche de motif dans les arbres planaires. Cette analyse illustre les techniques du paragraphe 1.1, en montrant l'intérêt qu'il y a, vis-à-vis des dénombrements d'arbres, à se ramener à des définitions inductives sur la structure. Il apparaît que le temps moyen de recherche est linéaire en la taille du texte et du motif, résultat qui contraste avec le comportement quadratique dans le cas le pire de l'algorithme.

2.1. L'algorithme séquentiel.

Nous considérons le cas des arbres (planaires) binaires non étiquetés.

DEFINITIONS. *Un arbre A est un préfixe d'un arbre B si B s'obtient à partir de A par substitution d'arbres binaires quelconques aux sommets externes de B. Un arbre A est arbre partiel d'un arbre B ssi A est préfixe d'un sous-arbre de B. La racine de B est alors appelée sommet d'occurrence de A dans B. □*

En d'autres termes, A occure en le sommet s de B ssi la représentation planaire de A se superpose à celle de B lorsqu'on fait coïncider la racine de A avec le sommet s de B.



L'algorithme séquentiel de détection d'un motif M dans un texte T consiste en une procédure d'exploration des sommets de B au cours de laquelle, en chaque sommet $s \in T$ est lancée une procédure de comparaison ; la procédure de comparaison explore simultanément en préordre le motif M et le sous-arbre T et s'interrompt

[†] Une première description de cette analyse est parue dans [Flajolet ; 1978].

dès qu'une non-coïncidence ("mismatch") est détectée. La version récursive est décrite à la figure 5.

```

procédure TRAVERSE(T);
  externe: M; arbre binaire: T, M;
  si T ≠ 0 : si COMPARE(M; T): écrire racine(T) fsi;
                                     TRAVERSE(gauche(T));
                                     TRAVERSE(droite(T))
                                fsi
  fsi
fproc
procédure booléen Z ← COMPARE(A, B);
  arbre binaire A, B;
  si A = ∅ : Z ← vrai
  sinssi B = ∅ : Z ← faux
  sinon      : Z ← COMPARE(gauche(A); gauche(B));
               si Z: Z ← COMPARE(droite(A); droite(B)) fsi
  fsi
fproc

```

Figure 5. L'algorithme de détection séquentielle.

Appliqué à un arbre texte T dans un contexte où le motif M est passé comme paramètre externe, cet algorithme construit la liste des occurrences de M dans T . Afin de décrire les paramètres de l'analyse, nous donnons deux définitions :

DEFINITION 1. Etant donné un arbre M de taille m , le segment de M d'ordre k , noté $\text{seg}_k(M)$ est l'arbre partiel de M induit par les k -premiers sommets de M dans l'ordre préfixe. $\square$

Par convention $\text{seg}_0(M) = \emptyset$ (l'arbre vide).

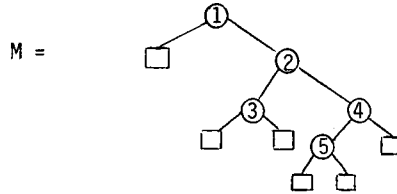
DEFINITION 2. Etant donné un arbre M de taille m , la fonction de feuille associée à M est la fonction $F_M : [0..m] \rightarrow [0..m+1]$ définie inductivement par :

- (i) si $M = \emptyset : F_M(0) = 1$
- (ii) si $M = \begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ M_1 \quad M_2 \end{array}$ avec $|M_1| = m_1$ et $|M_2| = m_2 :$

$$F_M(k) = \begin{cases} 0 & \text{pour } k = 0 \\ F_{M_1}(k-1) & \text{pour } 0 < k \leq m_1 \\ F_{M_2}(k-m_1-1)+m_1+1 & \text{pour } m_1 < k \leq m. \end{cases} \quad \square$$

En particulier $F_M(m) = m+1$; $F_M(k)$ compte le nombre de sommets externes rencontrés dans la visite en préordre des sommets de rang 1,2,... jusqu'à $k+1$.

Exemple : Pour l'arbre



dont les sommets sont étiquetés par leur rang en préordre, la fonction de feuille vaut

k	0	1	2	3	4	5
$F_M(k)$	0	1	1	3	3	6

□

Le décompte d'instructions lors de l'exécution de l'algorithme avec M comme motif et T comme texte montre :

LEMME : Le temps de la détection séquentielle du motif M dans le texte T vaut

$$T_S(M;T) = c_0 + c_1 \cdot t + c_2 \cdot \text{occ}(M;T) + c_3 \cdot \sum_{1 \leq k \leq m} \text{occ}(\text{seg}_k(M); T) + c_4 \cdot \sum_{1 \leq k \leq p} \text{occ}(\text{seg}_k(M); T) \cdot [F_M(k) - F_M(k-1)],$$

où $\text{occ}(A,B)$ désigne le nombre de sommets d'occurrence de A dans B ; $m = |M|$ et $t = |T|$. Les constantes numériques c_0, c_1, c_2, c_3, c_4 représentent des comptes d'instruction et sont caractéristiques de l'implantation. □

En particulier la somme où apparaît la fonction nombre de feuilles correspond à la somme des temps de la procédure de comparaison.

2.2. L'analyse en moyenne.

Le point de départ est le lemme de comptage suivant :

LEMME. Le nombre d'occurrences d'un motif M dans l'ensemble des arbres de taille t ne dépend que de la taille m de M et s'exprime par la forme binomiale

$$o_t^{(m)} = \binom{2t-m+1}{t+1}$$

Preuve : On traduit la définition inductive de la notion d'occurrence.

Soit $TETE(M;T)$ la quantité définie par :

$$TETE(M;T) = \begin{cases} 1 & \text{si } M \text{ occure à la racine de } T \\ 0 & \text{sinon.} \end{cases}$$

On définit les multi-ensembles (dépendant de M) :

$$\theta^{(M)} = \sum_{T \in \mathcal{B}} \text{occ}(M;T) \cdot T \quad \text{et} \quad \mathcal{E}^{(M)} = \sum_{T \in \mathcal{B}} TETE(M;T) \cdot T.$$

La définition de l'occurrence montre qu'on a l'équation

$$\theta^{(M)} = \begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ \theta^{(M)} \quad \mathcal{B} \end{array} + \begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ \mathcal{B} \quad \theta^{(M)} \end{array} + \mathcal{E}^{(M)}$$

qui traduit qu'une occurrence a lieu soit à droite, soit à gauche, soit à la racine de l'arbre texte. Par ailleurs $\mathcal{E}$ est donné inductivement sur M par

$$\mathcal{E}^{(\emptyset)} = \mathcal{B}; \quad \mathcal{E}^{(M)} = \begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ \mathcal{E}^{(M_1)} \quad \mathcal{E}^{(M_2)} \end{array} \quad \text{si } M = \begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ M_1 \quad M_2 \end{array}.$$

Ainsi pour les séries génératrices associées $o^{(M)}(z)$ et $T^{(M)}(z)$

$$\begin{cases} o^{(M)}(z) = 2z B(z) o(z) + T^{(M)}(z) \\ T^{(\emptyset)}(z) = B(z); \quad T^{(M)}(z) = z T^{(M_1)}(z) T^{(M_2)}(z). \end{cases}$$

La seconde équation montre que

$$T^{(M)}(z) = z^m (B(z))^{m+1},$$

d'où finalement en tenant compte de l'expression de $B(z)$:

$$\begin{aligned} o^{(M)}(z) &= \frac{1}{z\sqrt{1-4z}} (1 - \sqrt{1-4z})^{m+1}, \\ &= \frac{1}{(m+2)!} \frac{1}{z} \frac{d}{dz} [z^{m+2} B(z)^{m+2}]. \end{aligned}$$

En utilisant l'expression classique des coefficients des puissances de $B(z)$ déduite par exemple du théorème de Lagrange-Bürmann

$$(B(z))^k = \sum_{n \geq 0} [\binom{2n+k-1}{n} - \binom{2n+k-1}{n-1}] z^k,$$

on obtient, toutes réductions faites la valeur du coefficient $0_t^{(M)}$:

$$0_t^{(M)} = \binom{2t-m+1}{t+1},$$

quantité qui ne dépend que de la taille de M et qu'on note encore $0_t^{(m)}$. $\square$

Cette forme binomiale peut également s'obtenir par l'utilisation du théorème de Raney ([Raney ; 1960]).

COROLLAIRE. La probabilité d'occurrence d'un motif de taille m en un sommet d'un arbre texte de taille t vérifie les inégalités

$$\frac{1}{2^{m-1}} - \frac{(m-1)(m-2)}{t \cdot 2^{m+1}} < \frac{0_t^{(m)}}{t \cdot B_t} < \frac{1}{2^{m-1}} \quad t \geq m \geq 3. \quad \square$$

La décroissance exponentielle des probabilités d'occurrence d'un motif en fonction de la taille du motif est à l'origine du caractère linéaire en moyenne de la complexité de l'algorithme séquentiel. La forme simple de ce corollaire permet de surcroît de donner du temps de recherche une expression inductive sur la structure du motif.

PROPOSITION. Le temps moyen de détection séquentielle d'un motif M de taille t dans l'ensemble des textes de taille t

$$\bar{T}_S(M; t) = \frac{1}{B_t} \sum_{|T|=t} T_S(M; T)$$

est donné par la formule

$$\bar{T}_S(M, t) = t(\alpha + \beta \cdot 2^{-m} + \gamma \cdot v_S(P)) + O(1)$$

où v_S est une valuation définie inductivement sur la structure d'arbre par

$$v_S(\emptyset) = 1 ; \quad v_S\left(\begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ M_1 \quad M_2 \end{array}\right) = \frac{1}{2} [v_S(M_1) + \frac{1}{2^{m_1}} v_S(M_2)]$$

où $m_1 = |M_1|$. Les quantités α, β, γ sont des constantes dépendant de l'implantation.

Preuve : (indications). La preuve résulte de la combinaison du lemme donnant le temps de détection séquentielle avec la valeur asymptotique des probabilités d'occurrence. La valuation v_S traduit le temps de calcul des procédures de comparaison pondéré par ces probabilités d'occurrence ; la définition

de la valuation reflète la définition inductive de la fonction de feuille. $\square$

La figure ci-dessous donne l'exemple d'arbres de taille 6 avec la valeur de la valuation associée.

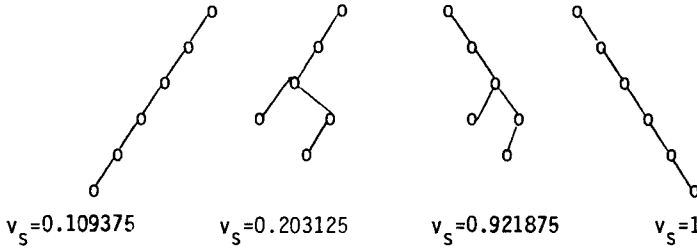


Figure 6 : Interdépendance de la valuation et de la forme de l'arbre.

Ainsi la valuation dépend-elle sensiblement de la forme de l'arbre motif.

L'expression inductive de v_s permet le calcul asymptotique de la valeur moyenne de v_s sur l'ensemble des arbres motifs de taille m , et ainsi conduit à l'évaluation de l'algorithme de détection séquentielle.

THEOREME 1. Le temps moyen de recherche séquentielle d'un motif de taille m dans un texte de taille t vaut

$$\bar{T}_s(m; t) = t(\alpha + \gamma \bar{v}_s)(1 + O(\frac{1}{m}))$$

où la constante $\bar{v}_s$ est égale à $6 - 4\sqrt{2}$; les quantités α et γ sont les constantes de temps liées à l'implantation de la proposition précédente.

Preuve : On considère le multi-ensemble

$$\mathcal{V} = \sum_{M \in \mathcal{B}} v_s(M) \cdot M.$$

La définition inductive de v_2 se traduit par l'équation

$$\mathcal{V} = \frac{1}{2} \left[\begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ \mathcal{B} \quad \mathcal{V} \end{array} + \begin{array}{c} \text{O} \\ \swarrow \quad \searrow \\ \mathcal{V} \quad \mathcal{K} \end{array} \right] + 1. \quad \square$$

où $\mathcal{K} = \sum_{M \in \mathcal{B}} \frac{1}{2^{|M|}} \cdot M.$

Soit $V(z)$ la série image de $\mathcal{V}$; V est donc la série génératrice des valuations totales V_m :

$$V_m = \sum_{|M|=m} v_s(M).$$

Le multi-ensemble $\mathcal{K}$ possède comme image la série

$$\begin{aligned} K(z) &= \sum \frac{1}{2^m} B_m z^m = B\left(\frac{z}{2}\right) \\ &= \frac{1 - \sqrt{1-2z}}{z}, \end{aligned}$$

d'où pour $V(z)$ l'équation

$$V(z) = \frac{1}{2} z B(z) V(z) + \frac{1}{2} z V(z) B\left(\frac{z}{2}\right) + 1,$$

soit en résolvant en V :

$$V(z) = \frac{4}{1 + \sqrt{1-4z} + 2\sqrt{1-2z}}.$$

La fonction V est ainsi une fonction algébrique dont la singularité la plus proche de l'origine est un point de branchement en $z = \frac{1}{4}$.

En multipliant par les quantités conjuguées, on met $V(z)$ sous la forme

$$\begin{aligned} V(z) &= -\frac{1}{2z} (1 + \sqrt{1-4z} - 2\sqrt{1-2z})(\sqrt{1-4z} + 1 - 2z). \\ &= B(z) \cdot \frac{1}{2} (2 - 2z - 2\sqrt{1-2z}) + H(z) \end{aligned}$$

où $H(z)$ a pour rayon de convergence $\frac{1}{2}$.

Le coefficient H_n de $H(z)$ est en $O(\rho^{-n})$ pour ρ quelconque inférieur à $\frac{1}{2}$, comme on le voit en appliquant une majoration triviale à la forme intégrale de H_n donnée par le théorème de Cauchy

$$H_n = \frac{1}{2i\pi} \int_{|z|=\rho} H(z) \frac{dz}{z^{n+1}} \quad 0 < \rho < \frac{1}{2}.$$

Ainsi par exemple $H_n = O(3^n)$ qui est exponentiellement petit devant B_n .

La fonction $G(z) = \frac{1}{2}(2 - 2z - \sqrt{1-2z})$ est régulière au voisinage de $z = \frac{1}{4}$;

on peut donc l'écrire sous la forme

$$G(z) = \sum_{k \geq 0} g_k \left(z - \frac{1}{4}\right)^k.$$

La multiplication par $B(z)$ montre que le produit $G(z) B(z)$ s'exprime par un développement

$$B(z) G(z) = \sum_{k \geq 0} h_k \left(z - \frac{1}{4}\right)^{k+\frac{1}{2}} + \sum_{k \geq 0} h'_k \left(z - \frac{1}{4}\right)^k,$$

où les h_k, h'_k sont des coefficients numériques.

Formellement, le n -ième coefficient de Taylor du terme $h_k \left(z - \frac{1}{4}\right)^{k+\frac{1}{2}}$ se comporte en $O(4^n n^{-3/2-k})$ et se trouve ainsi asymptotiquement négligeable devant B_n dès que $k \geq 1$. On montre qu'en effet il est analytiquement justifié pour obtenir

l'équivalent asymptotique du coefficient n -ième de $G(z)$ de négliger les termes autres que le premier terme : c'est-là le théorème de Darboux (cf. [Henrici ; 1978]).

L'on obtient de la sorte l'estimation :

$$V_n = B_n G(0) + O\left(\frac{1}{n} B_n\right) + O(\rho^n) \quad \text{pour } \rho < \frac{1}{2}.$$

Ainsi

$$\frac{V_m}{B_m} = (6 - 4\sqrt{2}) + O\left(\frac{1}{m}\right),$$

ce qui conduit immédiatement à l'énoncé du théorème. $\square$

La figure 7 présente le temps de calcul de de textes et motifs tirées entre 0 et 100, la taille du motif étant inférieure à celle du texte.

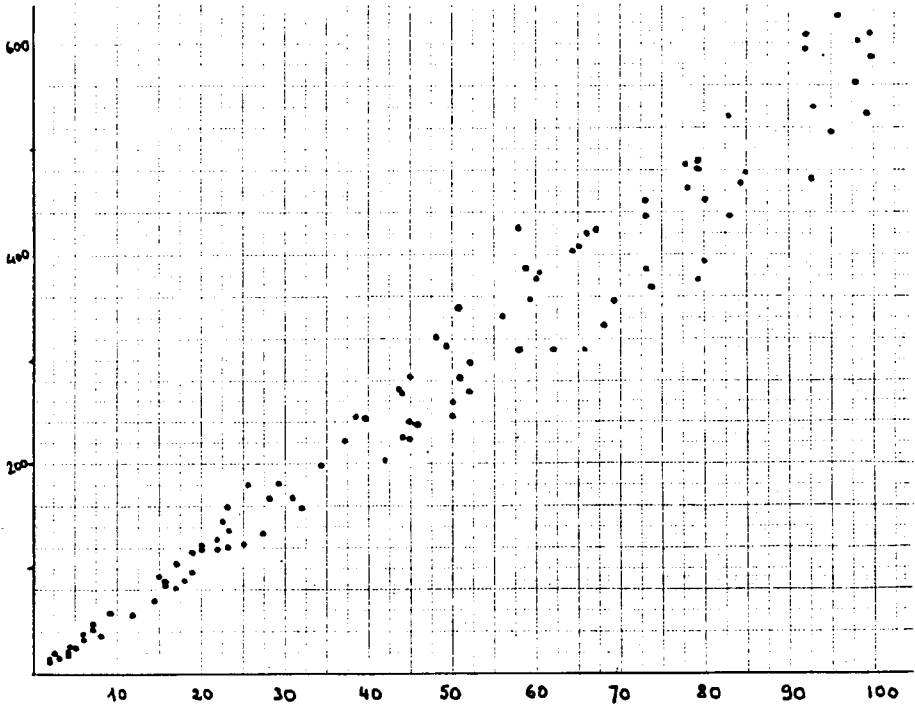


Figure 7 : Temps de calcul de l'algorithme séquentiel évalué en nombre d'appels des procédures EXPLORE, COMPARE et ECRIRE, en fonction de la taille de l'arbre texte.

Les arbres ont été tirés aléatoirement selon la statistique de Catalan au moyen d'un algorithme dû à J.L. Rémy. Le diagramme montre clairement la linéarité de l'algorithme et la faible dispersion des temps, justifiant ainsi d'un point de vue pratique l'analyse précédente.

On montre enfin ([Flajolet, Steyaert ; 1979]) que cette propriété de linéarité en moyenne est valable pour une large classe de familles d'arbres.

3. FORETS FONCTIONNELLES BINAIRES : L'ANALYSE HEURISTIQUE D'UN ALGORITHME DE FACTORISATION.

Nous montrons ici comment effectuer les dénombrements relatifs à des forêts greffées sur des cycles (graphes fonctionnels). Comme application, nous effectuons l'analyse, moyennant une hypothèse "heuristique" simple, d'un algorithme de factorisation par "méthode de Monte-Carlo" dû à [Pollard ; 1975].

3.1. Construction des forêts fonctionnelles binaires.

A toute fonction finie $f : E \rightarrow E$, on peut associer un graphe orienté dont les sommets sont l'ensemble E et dont les arêtes sont formées des couples $\langle x, f(x) \rangle$. Ce graphe est le graphe fonctionnel associé à la fonction f . La figure 7 donne le graphe fonctionnel de $f(x) = \lfloor 17 \sin^2 x \rfloor$ pour $x \in [0..16]$. Un graphe fonctionnel peut être considéré comme formé d'un ensemble de cycles sur lequel sont greffés des arbres non planaires étiquetés, les étiquettes étant distinctes des étiquettes des cycles. La bijection avec les fonctions de $[n]$ dans $[n]$ montre que le nombre de graphes fonctionnels construit sur $[n]$ vaut n^n .

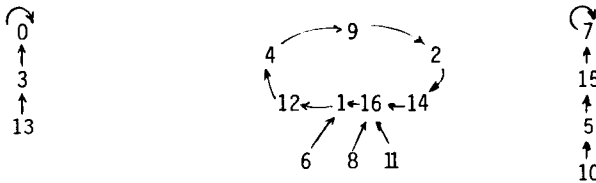


Figure : Graphe fonctionnel de $\lfloor 17 \sin^2 x \rfloor$ sur $[0..16]$.

Nous nous intéressons ici aux graphes fonctionnels binaires dans lesquels tout sommet possède 0 ou 2 antécédents. Ces graphes sont associés aux fonctions f telles que

$$\forall x \quad \text{card}\{f^{-1}(x)\} = 0 \quad \text{ou} \quad \text{card}\{f^{-1}(x)\} = 2.$$

Nous nous proposons d'évaluer différents paramètres des graphes fonctionnels binaires : longueur de cycle, distance des points au cycle. Pour cela, nous

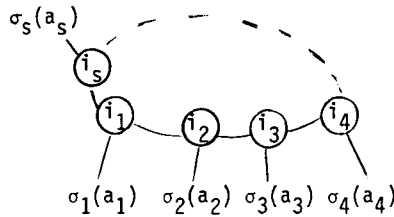
allons d'abord construire les graphes fonctionnels binaires.

Les arbres binaires étiquetés sont une famille $\mathcal{A}$ définie par

$$\mathcal{A} = \textcircled{1} + \frac{1}{2} M_2(\mathcal{A}, \mathcal{A})$$

où M_2 est le constructeur défini à la section 1.3 qui consiste à former un arbre à partir de deux sous-arbres en distribuant les étiquettes.

Les cycles greffés (binaires) sont des graphes fonctionnels (binaires) comportant un seul cycle. On les construit à partir de l'ensemble des arbres binaires étiquetés par greffe sur des cycles. Pour chaque entier $s \geq 1$, on définit un constructeur $G_s : G_s(a_1, a_2, \dots, a_s)$ est défini comme l'ensemble des cycles greffés de forme



où les i_k, σ_k vérifient, avec $n_k = |a_k|$ et $n = n_1 + n_2 + \dots + n_s + s$, les conditions

- pour tout $k, 1 \leq k \leq s, \sigma_k$ est une application croissante de $[1..n_k]$ dans $[1..n]$ étendue aux arbres par renommage des étiquettes ;
- $\text{Im}(\sigma_1) \circ \text{Im}(\sigma_2) \dots \circ \text{Im}(\sigma_s) \circ \{i_1, i_2, \dots, i_s\} = [1..n]$.

L'ensemble $\mathcal{C}$ des cycles greffés binaires vérifie l'équation :

$$\mathcal{C} = \frac{1}{1} G_1(\mathcal{A}) + \frac{1}{2} G_2(\mathcal{A}, \mathcal{A}) + \frac{1}{3} G_3(\mathcal{A}, \mathcal{A}, \mathcal{A}) + \dots,$$

dans laquelle le coefficient $\frac{1}{s}$ de $G_s(\mathcal{A}, \mathcal{A}, \dots, \mathcal{A})$ traduit le fait que chaque s -cycle est engendré s fois par $G_s(\mathcal{A}, \mathcal{A}, \dots, \mathcal{A})$ (autant de fois qu'il existe de permutations circulaires du cycle).

Enfin, les graphes fonctionnels binaires sont eux-mêmes construits à partir des cycles greffés par distribution des étiquettes, l'ordre des cycles n'étant pas pertinent. On définit pour chaque r un constructeur $P_r : P_r(c_1, c_2, \dots, c_r)$, où les c_k sont dans $\mathcal{C}$, représente l'ensemble des r -ensembles [†]

[†] ensemble de r éléments.

$\{\sigma_1(c_1), \sigma_2(c_2), \dots, \sigma_r(c_r)\}$ où les σ_k vérifient avec $n_k = |c_k|$ et $n = c_1 + c_2 + \dots + c_r$:

- pour tout k , $1 \leq k \leq r$, σ_k est une application croissante de $[1..n_k]$ dans $[1..n]$ étendue aux cycles par renommage des étiquettes ;
- $\text{Im}(\sigma_1) \circ \text{Im}(\sigma_2) \dots \circ \text{Im}(\sigma_r) = [1..n]$.

L'ensemble $\mathcal{F}$ des arbres fonctionnels binaires vérifie ainsi l'équation :

$$\mathcal{F} = P_0 + \frac{1}{1!} P_1(\mathcal{C}) + \frac{1}{2!} P_2(\mathcal{C}, \mathcal{C}) + \frac{1}{3!} P_3(\mathcal{C}, \mathcal{C}, \mathcal{C}) + \dots,$$

où le coefficient $\frac{1}{r!}$ de $P_r(\mathcal{C}, \mathcal{C}, \dots, \mathcal{C})$ traduit le fait que chaque r -ensemble de cycles est engendré $r!$ fois par $P_r(\mathcal{C}, \mathcal{C}, \dots, \mathcal{C})$ (autant de fois qu'il existe de permutation des composantes).

Chacun de ces constructeurs possède une image : soit $\mathcal{J}$ une famille d'arbres étiquetés et $\mathcal{J} = G_S(\mathcal{J}, \mathcal{J}, \dots, \mathcal{J})$ correspondant, on a la relation

$$J_n = s! \sum_{s+n_1+n_2+\dots+n_s=n} \binom{n}{n_1 n_2 \dots n_s} I_{n_1} I_{n_2} \dots I_{n_s},$$

où le multinomial traduit le nombre de manières de distribuer les étiquettes, et

où $s!$ représente le nombre de manières d'arranger les étiquettes dans le cycle.

Pour les séries génératrices exponentielles $I(z) = \sum_{n \geq 0} I_n \frac{z^n}{n!}$ et $J(z) = \sum_{n \geq 0} J_n \frac{z^n}{n!}$, ceci se traduit par l'égalité :

$$J(z) = z^S (I(z))^S.$$

Soit $\mathcal{K}$ une famille de cycles greffés et $\mathcal{L} = P_r(\mathcal{K}, \mathcal{K}, \dots, \mathcal{K})$; on a alors

$$L_n = \sum_{n_1+n_2+\dots+n_r=n} \binom{n}{n_1 n_2 \dots n_r} K_{n_1} K_{n_2} \dots K_{n_r},$$

somme dans laquelle le multinomial représente le nombre de manières de distribuer les étiquettes entre chacun des cycles greffés.

Ainsi pour les séries génératrices exponentielles $L(z) = \sum_{n \geq 0} L_n \frac{z^n}{n!}$ et

$K(z) = \sum_{n \geq 0} K_n \frac{z^n}{n!}$ obtient-on la relation

$$L(z) = (K(z))^r.$$

L'application au dénombrement de graphes fonctionnels binaires est simple : les équations

$$\begin{cases} \mathcal{A} = 1 + \frac{1}{2} M_2(\mathcal{A}, \mathcal{A}) \\ \mathcal{C} = \sum_{s \geq 1} \frac{1}{s} G_s(\mathcal{A}, \mathcal{A}, \dots, \mathcal{A}) \\ \mathcal{F} = \sum_{r \geq 0} \frac{1}{r!} P_r(\mathcal{C}, \mathcal{C}, \dots, \mathcal{C}), \end{cases}$$

se traduisent sur les séries génératrices exponentielles en les équations

$$\begin{cases} A(z) = z + \frac{z}{2} (A(z))^2 \\ C(z) = \sum_{s \geq 1} \frac{z^s (A(z))^s}{s} \\ F(z) = \sum_{r \geq 0} \frac{1}{r!} (C(z))^r, \end{cases}$$

soit en résolvant

$$\begin{cases} A(z) = \frac{1 - \sqrt{1-2z^2}}{z} \\ C(z) = -\ln \sqrt{1-2z^2} \\ F(z) = \frac{1}{\sqrt{1-2z^2}}, \end{cases}$$

d'où les résultats de dénombrement :

$$\begin{cases} A_{2n+1} = \frac{(2n+1)!}{2^n} \frac{1}{n+1} \binom{2n}{n} = 1.3 \dots (2n-1)(n+2)(n+3) \dots (2n+1) \\ C_{2n} = 2^n (2n-1)! \\ F_{2n} = \frac{(2n)!}{2^n} \binom{2n}{n} +. \end{cases}$$

On peut également marquer chacun des constructeurs. Soit $C(z, u)$ la série génératrice exponentielle en z des cycles partitionnés selon la longueur de cycle, et soit $F(z, u, t)$ la série génératrice exponentielle en z des graphes fonctionnels partitionnés selon la longueur totale de leurs cycles (marquée par u) et selon le nombre de leurs composantes connexes (marqué par t). Les équations de définition sont :

$$\begin{cases} A(z) = z + \frac{z}{2} (A(z))^2 \\ C(z, u) = \sum_{s \geq 1} \frac{z^s u^s}{s} (A(z))^s \\ F(z, u, t) = \sum_{r \geq 0} \frac{u^r}{r!} (C(z, u))^r, \end{cases}$$

⁺ Ce dernier résultat a été obtenu au moyen d'un raisonnement direct par P. Camion.

d'où :

$$\begin{cases} A(z) = \frac{1 - \sqrt{1-2z^2}}{2z} \\ C(z,u) = -\ln(1-u+u\sqrt{1-2z^2}) \\ F(z,u,t) = (1-u+u\sqrt{1-2z^2})^{-t}. \end{cases}$$

Ces équations permettent de calculer sans difficulté la longueur moyenne de cycle et le nombre moyen de composantes connexes d'un graphe fonctionnel binaire de taille n .

Nous nous intéressons maintenant aux deux paramètres suivant des graphes fonctionnels, lesquels sont utiles à l'analyse de factorisation de Pollard :

- distance moyenne d'un point du graphe à son cycle
- longueur moyenne du cycle vue d'un point du graphe.

On définit pour cela, les paramètres $\lambda(g)$ longueur totale de cycle rapportée aux points du graphe fonctionnel g et $\mu(g)$ longueur de cheminement au cycle du graphe fonctionnel g par :

$$\begin{cases} \lambda(a) = 0 \\ \lambda(G_S(a_1, a_2, \dots, a_s)) = s(|a_1| + |a_2| + \dots + |a_s| + s) \\ \lambda(P_r(c_1, c_2, \dots, c_r)) = \lambda(c_1) + \lambda(c_2) + \dots + \lambda(c_r) \\ \mu(1) = 1 ; \mu(M_2(a_1, a_2)) = \mu(a_1) + \mu(a_2) + |a_1| + |a_2| + 1 \\ \mu(G_S(a_1, a_2, \dots, a_s)) = \mu(a_1) + \mu(a_2) + \dots + \mu(a_s) \\ \mu(P_r(c_1, c_2, \dots, c_r)) = \mu(c_1) + \mu(c_2) + \dots + \mu(c_r), \end{cases}$$

pour $a, a_k \in \mathcal{A}$ et $c_k \in \mathcal{C}$.

On introduit les multiensembles

$$\begin{aligned} \mathcal{L}\mathcal{A} &= \sum_{a \in \mathcal{A}} \lambda(a) \cdot a ; \mathcal{L}\mathcal{C} = \sum_{c \in \mathcal{C}} \lambda(c) \cdot c ; \mathcal{L}\mathcal{F} = \sum_{f \in \mathcal{F}} \lambda(f) \cdot f ; \\ \mathcal{M}\mathcal{A} &= \sum_{a \in \mathcal{A}} \mu(a) \cdot a ; \mathcal{M}\mathcal{C} = \sum_{c \in \mathcal{C}} \mu(c) \cdot c ; \mathcal{M}\mathcal{F} = \sum_{f \in \mathcal{F}} \mu(f) \cdot f ; \end{aligned}$$

et les définitions inductives de λ, μ se traduisent par les équations :

$$\begin{cases} \mathcal{L}\mathcal{A} = 0 \\ \mathcal{L}\mathcal{C} = \sum_{s \geq 1} G_s(\mathcal{A}, \mathcal{A}, \dots, \mathcal{A}) \cdot |G_s(\mathcal{A}, \mathcal{A}, \dots, \mathcal{A})| \\ \mathcal{L}\mathcal{F} = \sum_{r \geq 1} \frac{1}{r!} [P_r(\mathcal{L}\mathcal{C}, \mathcal{C}, \dots, \mathcal{C}) + \dots + P_r(\mathcal{C}, \mathcal{C}, \dots, \mathcal{L}\mathcal{C})] ; \end{cases}$$

soit avec $\mathcal{D}\mathcal{A} = \sum_{a \in \mathcal{A}} |a| \cdot a :$

$$\begin{cases} \mathcal{M}\mathcal{A} = 1 + \frac{1}{2} [M_2(\mathcal{M}\mathcal{A}, \mathcal{A}) + M_2(\mathcal{A}, \mathcal{M}\mathcal{A})] + \mathcal{D}\mathcal{A} \\ \mathcal{M}\mathcal{C} = \sum_{s \geq 1} \frac{1}{s!} [G_s(\mathcal{M}\mathcal{A}, \mathcal{A}, \dots, \mathcal{A}) + \dots + G_s(\mathcal{A}, \mathcal{A}, \dots, \mathcal{M}\mathcal{A})] \\ \mathcal{M}\mathcal{F} = \sum_{r \geq 1} \frac{1}{r!} [P_r(\mathcal{M}\mathcal{C}, \mathcal{C}, \dots, \mathcal{C}) + \dots + P_r(\mathcal{C}, \mathcal{C}, \dots, \mathcal{M}\mathcal{C})] . \end{cases}$$

Ainsi pour les séries génératrices exponentielles $LA(z)$, $LC(z)$ etc..., obtient-on les équations

$$\begin{cases} LA(z) = 0 \\ LC(z) = \frac{z \frac{d}{dz}(zA)}{(1-zA)^2} \\ LF(z) = LC(z) \exp C(z) \end{cases} \quad \begin{cases} MA(z) = z MA(z)A(z) + z \frac{d}{dz} A(z) \\ MC(z) = \frac{zMA(z)}{1-zA(z)} \\ MF(z) = MC(z) \exp C(z) . \end{cases}$$

On résout, et l'on trouve :

$$LF(z) = \frac{2z^2}{(1-2z^2)^2} \quad MF(z) = \frac{1}{(1-2z^2)^2} - \frac{1}{(1-2z^2)^{3/2}}$$

d'où les valeurs des coefficients

$$LF_{2n} = n \cdot 2^n 2n! \quad MF_{2n} = [(n+1)2^n - \frac{(n+1)}{2^{n+1}} \binom{2n+2}{n+1}] 2n! .$$

PROPOSITION. Les valeurs moyennes de longueur de cycle vues d'un point, et de distance au cycle dans l'ensemble des graphes fonctionnels binaires valent

$$\bar{\lambda}_{2n} = \frac{LF_{2n}}{2n \cdot F_{2n}} = \frac{4^n}{2 \binom{2n}{n}} \quad \bar{\mu}_{2n} = \frac{MF_{2n}}{2n \cdot F_{2n}} = \frac{n+1}{2n} \cdot \frac{4^n}{\binom{2n}{n}} - \frac{2n+1}{2n} ,$$

soit asymptotiquement

$$\bar{\lambda}_{2n} = \frac{1}{2} \sqrt{\pi n} + O\left(\frac{1}{\sqrt{n}}\right) \quad \text{et} \quad \bar{\mu}_{2n} = \frac{1}{2} \sqrt{\pi n} - 1 + O\left(\frac{1}{\sqrt{n}}\right) .$$

L'algorithme de factorisation de [Pollard ; 1975] qui factorise n au moyen d'une fonction $f : [n] \rightarrow [n]$ peut être succinctement décrit de la manière suivante :

1. Tirer x_0 dans $[1..n]$
2. Calculer itérativement $x_m = f(x_{m-1})$
 jusqu'à ce que $d = \text{pgcd}(n, x_m - x_j) \neq 1$
 pour un certain $j : 0 \leq j < m$
3. Ecrire d .

Le principe de l'algorithme est que si par exemple $n = pqr$ avec $p < q < r$ et p, q, r premiers, alors f présente un cycle modulo p plus rapidement que modulo n (en "moyenne" en $O(\sqrt{p})$ étapes) ; on a alors

$$x_m \equiv x_j \pmod{p},$$

et ainsi $\text{pgcd}(n, x_m - x_j)$ fournit un diviseur de n généralement non trivial. Cet algorithme s'implante d'une part en choisissant une classe de fonctions f particulières - typiquement $f(x) = x^2 + a$ pour a entier ; d'autre part en améliorant la procédure de recherche des cycles modulo un diviseur de n . L'algorithme devient ainsi

1. tirer a , init dans $[0..n-1]$; $r, \ell \leftarrow \text{init}$;
2. *iterer*

$$r \leftarrow (r^2 + a)^2 + a \pmod{n}$$

$$\ell \leftarrow \ell^2 + a \pmod{n}$$

$$d \leftarrow \text{pgcd}(r - \ell, n)$$

sur $d \neq 1$:

fiterer
3. *écrire* (d) .

Le paramètre de base de l'algorithme qui en détermine le temps d'exécution reste lié à la valeur de $\lambda + \mu$ associée au graphe fonctionnel de f modulo p pour p diviseur de n .

L'évaluation précédente s'applique sous l'hypothèse "heuristique" que les paramètres λ, μ sur la classe des graphes fonctionnels associés aux fonctions $\{x^2 + a\} \pmod{p}$ ont asymptotiquement (lorsque p est grand) même valeur que sur l'ensemble des graphes fonctionnels binaires. Moyennant cette hypothèse "heuristique

l'algorithme de Pollard possède en temps de calcul en $\frac{1}{2} \sqrt{\pi p}$ où p est le plus petit diviseur premier de n , soit $O(n^{1/4})$ (cf. [Camion, Flajolet, Monier ; 1979] pour une description plus détaillée accompagnée de résultats numériques).

CHAPITRE II

ALLOCATION DE REGISTRES ET EVALUATION D'EXPRESSIONS ARITHMETIQUES.

Nous étudions ici le problème de l'évaluation d'expressions arithmétiques. Ershov a montré, dès 1958, qu'il existe une stratégie d'évaluation d'expressions optimale vis-à-vis de l'occupation des registres de l'unité de calcul. Le paramètre nombre de registres de cette stratégie est une fonction de la forme de l'arbre sous-jacent à l'expression. Cette fonction de forme des arbres binaires est connue indépendamment en statistique dans les domaines de la géomorphologie et de la biologie sous le nom de nombre de Strahler ([Shreve ; 1966], [Mc Mahon ; 1975]).

La comparaison de la stratégie optimale à une stratégie d'évaluation plus simple - de gauche à droite, par exemple - nécessite le calcul de la distribution des arbres selon la fonction nombre de registres. Dans ce chapitre, nous analysons complètement la stratégie optimale : nous obtenons des résultats de distribution et de moyenne sous une forme simple qui permet un calcul exact. Nous effectuons ensuite l'évaluation asymptotique de ces résultats, ce qui conduit pour la distribution du nombre de registres sur l'ensemble des arbres de taille n à l'existence d'une infinité de lois limites dépendant de la partie fractionnaire de $\log_4 n$. Ce phénomène de périodicité se reflète au niveau du nombre moyen de registres nécessaire à l'évaluation des expressions de taille n pour lequel nous établissons la forme

$$R_n = \log_4 n + D(\log_4 n) + o(1)$$

où D est une fonction continue, périodique et de période 1. Ce résultat est à comparer au nombre moyen de registres, $\bar{R}_n^{GD}$ utilisé par l'évaluation (non optimale) procédant de gauche à droite :

$$\bar{R}_n^{GD} = \sqrt{\pi n} - \frac{1}{2} + o(1)$$

résultat qui se déduit de l'évaluation de la hauteur des arbres par [de Bruijn, Knuth, Rice ; 1972].

Avant l'exposé proprement dit, signalons que l'obtention des résultats de distribution s'effectue par les méthodes de l'analyse combinatoire spécifiques aux arbres, lesquelles ont été présentées au chapitre 1. L'analyse asymptotique utilise un résultat de Delange concernant la distribution des chiffres dans les représentations binaires d'entiers. Nous montrerons au chapitre suivant qu'un problème en apparence non relié d'analyse de réseaux de tri est justiciable de méthodes analogues.

Ce chapitre prolonge un travail effectué en collaboration avec J.C. Raoult et J. Vuillemin ([Flajolet, Raoult, Vuillemin ; 1977 et 1979]). R. Kemp [1977] a obtenu des résultats analogues par des méthodes différentes que nous discuterons au chapitre suivant. Les résultats de distribution, l'analyse de variance (poursuivie au chapitre suivant) et la statistique des ordres de sommets sont nouveaux.

1. PROBLEMATIQUE INFORMATIQUE.

Un programme écrit dans un langage de programmation de type FORTRAN, ALGOL, PASCAL etc... est traduit par un compilateur en une suite d'instructions plus élémentaires directement exécutables par une machine. Les modèles usuels de machines ('de Von Neumann') comportent une unité de calcul opérant sur un petit nombre de registres (par ex. 16) ; les opérandes généralement issus de calculs antérieurs sont pris dans une mémoire de capacité plus grande, et chargés dans les registres sur lesquels l'unité de calcul opère. Un modèle simplifié d'un évaluateur d'expressions arithmétiques est décrit par la figure 1.

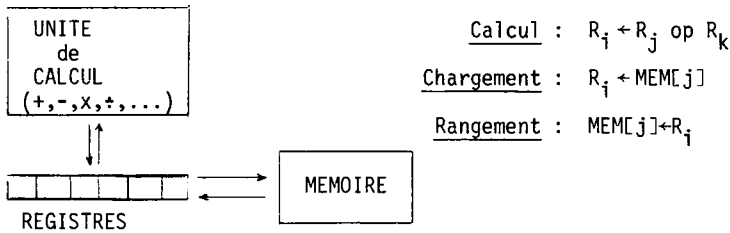


Figure 1 : *Modèle d'évaluation d'expressions arithmétiques.*

Il est clair que le nombre de registres de calcul d'une expression dépend de l'ordre d'évaluation des sous-expressions. Ainsi pour l'expression $a \times (b+c)$, les évaluations gauche-droite et droite-gauche conduisent aux suites d'instructions :

gauche-droite

$R_0 \leftarrow a$
 $R_1 \leftarrow b$
 $R_2 \leftarrow c$
 $R_1 \leftarrow R_1 + R_2$
 $R_1 \leftarrow R_0 \times R_1$

droite-gauche

$R_0 \leftarrow b$
 $R_1 \leftarrow c$
 $R_0 \leftarrow R_0 + R_1$
 $R_1 \leftarrow a$
 $R_0 \leftarrow R_1 \times R_0$

La première stratégie d'évaluation utilise les registres (R_0, R_1, R_2) , la seconde les registres (R_0, R_1) . Le nombre de registres utilisé par l'évaluation gauche-droite est une fonction Reg^{GD} définie inductivement sur la structure d'arbre par les règles :

$$\text{Reg}^{\text{GD}}(\square) = 0 ; \quad \text{Reg}^{\text{GD}}(T_1 \circ T_2) = \max(\text{Reg}^{\text{GD}}(T_1) ; 1 + \text{Reg}^{\text{GD}}(T_2)).$$

(Par convention, les registres sont numérotés à partir de zéro, et le nombre de registres est le maximum des indices de registres utilisés). On dispose d'une formule analogue pour l'évaluation droite-gauche et globalement (en moyenne et en distribution) les deux stratégies ont des comportements équivalents.

La définition inductive de Reg^{GD} montre que pour un arbre T , $\text{Reg}^{\text{GD}}(T)$ est égal à la hauteur droite de l'arbre (binaire) T , et donc égal à la hauteur de l'arbre général $\rho(T)$ associé à T par la correspondance de rotation (cf. chapitre I). L'analyse de la stratégie d'évaluation gauche-droite se ramène donc au problème de la hauteur des arbres résolu par de Bruijn et al. dont la valeur moyenne déjà indiquée dans l'introduction, vaut

$$\sqrt{n} - \frac{1}{2} + O(1).$$

Parmi les stratégies ne dépendant que de la forme de l'arbre sous-jacent, on démontre ([Sethi, Ullman ; 1970]) l'optimalité de la stratégie due à [Ershov, 1958]. Celle-ci consiste à calculer de l'intérieur, en priorité, les expressions les plus "lourdes" : la fonction nombre de registres de l'évaluation optimale, Reg , est donnée par :

$$\begin{cases} \text{Reg}(\square) = 0 \\ \text{Reg}(T_1 \circ T_2) = \begin{cases} \text{si } \text{Reg}(T_1) = \text{Reg}(T_2) \text{ alors } \text{Reg}(T_1) \\ \text{sinon } \max(\text{Reg}(T_1); \text{Reg}(T_2)) \end{cases} \end{cases}$$

l'évaluation de $T_1 \circ T_2$ se fait dans l'ordre T_1-T_2 si $\text{Reg}(T_1) \geq \text{Reg}(T_2)$ et dans l'ordre T_2-T_1 si $\text{Reg}(T_1) < \text{Reg}(T_2)$.

Ainsi pour l'expression $E = a + (b - c) \times (d - e) * (f + g)$, l'ordre d'évaluation optimal est-il

$$(f+g) ; (d-e) ; (d-e) * (f+g) ; (b-c) \times (d-e) * (f+g) ; E ;$$

le nombre de registres $\text{Reg}(E)$ vaut 2, et les instructions machines correspondantes sont :

$R_0 \leftarrow f$	$R_1 \leftarrow R_1 - R_2$	$R_0 \leftarrow R_1 \times R_0$
$R_1 \leftarrow g$	$R_0 \leftarrow R_1 * R_0$	$R_1 \leftarrow a$
$R_0 \leftarrow R_0 + R_1$	$R_1 \leftarrow b$	$R_0 \leftarrow R_1 + R_0$
$R_1 \leftarrow d$	$R_2 \leftarrow c$	
$R_2 \leftarrow e$	$R_1 \leftarrow R_1 - R_2$	

Par contraste, l'évaluation gauche-droite utilise 4 registres ; l'ordre d'évaluation des sous-expressions :

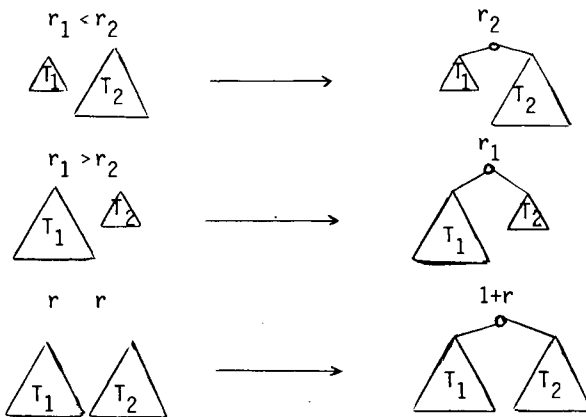
$(b-c)$; $(d-e)$; $(f+g)$; $(d-e) * (f+g)$; $(b-c) \times (d-e) * (f+g)$; E ,

correspond à la suite d'instructions

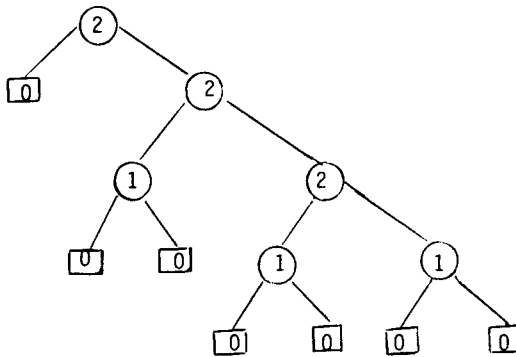
$R_0 \leftarrow a$	$R_3 \leftarrow e$	$R_2 \leftarrow R_2 * R_3$
$R_1 \leftarrow b$	$R_2 \leftarrow R_2 - R_3$	$R_1 \leftarrow R_1 \times R_2$
$R_2 \leftarrow c$	$R_3 \leftarrow f$	$R_0 \leftarrow R_0 + R_1$
$R_1 \leftarrow R_1 - R_2$	$R_4 \leftarrow g$	
$R_2 \leftarrow d$	$R_3 \leftarrow f+g$	

Remarque : Certains modèles de machine autorisent l'utilisation d'un opérande en mémoire ; cette commodité équivaut à disposer d'un registre supplémentaire pour les opérandes droits qui sont des variables. Le nombre de registres de la stratégie optimale pour ce modèle mixte diffère donc d'au plus 1 de la fonction Reg . Enfin les résultats de ce chapitre s'interprètent comme des bornes supérieures pour les algorithmes qui utiliseraient les répétitions éventuelles de sous-expressions. $\square$

Le nombre de registres Reg associé à un arbre d'expression se détermine en affectant 0 à chacun des sommets externes et en construisant de bas en haut une valuation sur les sommets de l'arbre selon les règles données pour Reg :



Sur l'arbre de l'expression E, on obtient la valuation :

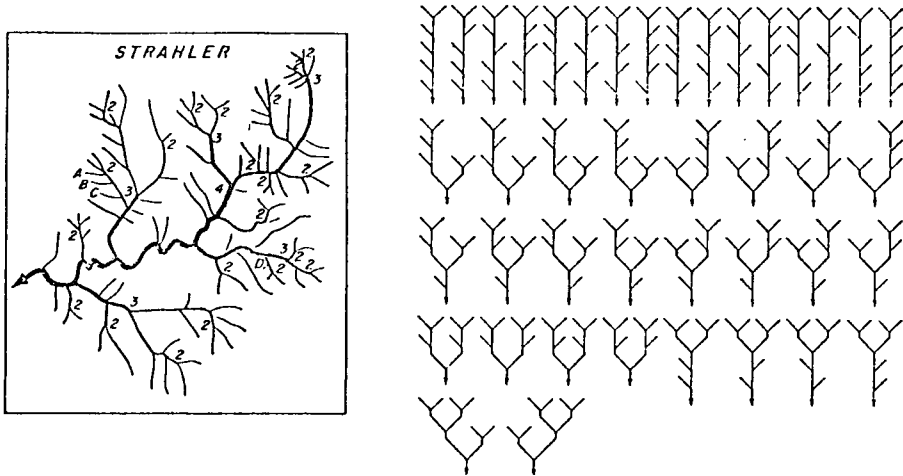


Sous cette forme, on reconnaît que la quantité $\text{Reg}(T)$ est égale à la hauteur du plus grand arbre parfait plongeable (en tant qu'arbre partiel) dans T .

C'est sous cette forme dite nombre de Strahler que la fonction Reg apparaît en géomorphologie. Reprenant les termes de Mc Mahon :

" Starting at the point of ultimate ramification, Strahler assigns to these lowest-order units the numeral 1. In the description of drainage patterns, these units are the initial feeder streams ; in the description of a tree, they are the outmost twigs. Whenever two such first order links meet they combine to form another link that continues onward. The continuing link is of the next high order in the system ; it is assigned the numeral 2. When as frequently happens in an asymetrical pattern of ramification, a first order link meets a second order link, the continuation is assigned the numeral of the higher order ... ".

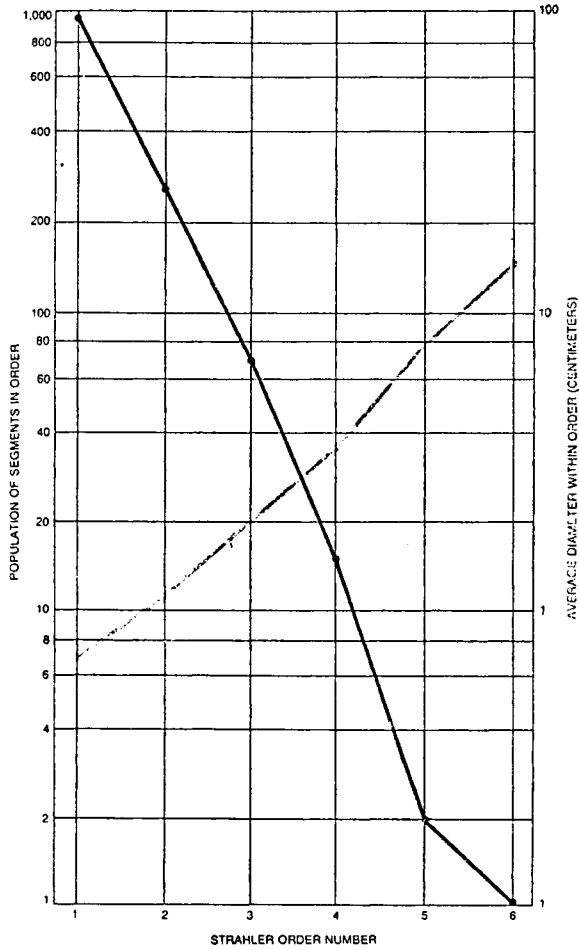
La figure 2 montre la numérotation de Strahler appliquée à la "Upper Hiwassee River (Géorgie du Nord) (reproduit d'après [Shreve, 1966]) et la classification des arbres de taille 5 selon la valeur du nombre de Strahler



-Schematic diagrams of the $N(6)=42$ topologically distinct channel networks with $l=11$ links and $n=6$ first-order Strahler streams or six Horton streams. In a topologically random population these networks would all be equally likely. Top row shows the possible second-order networks, for which $\Omega = 2$, $B=6.00$, $N(6;2)=16$, $p(6;2)=16/42=0.381$, $N[6,1]=16$, and $p[6,1]=16/16=1.000$. Bottom four rows show the possible third-order networks, for which $\Omega = 3$, $B=2.45$, $N(6;3)=26$, $p(6;3)=26/42=0.619$, $N[6,2,1]=24$, $p[6,2,1]=24/26=0.923$ (first three of the four rows), $N[6,3,1]=2$, and $p[6,3,1]=2/26=0.077$ (last row). Arrowhead indicates outlet in each diagram.

Figure 2 : Numérotation de Strahler : la Upper Hiwassee River et classification des arbres de taille 5 selon la valeur du nombre de Strahler.

Certaines observations statistiques relient dans de nombreuses familles d'arbres naturelles (peupliers, pins...) le nombre de Strahler à l'épaisseur des branches. Ces familles naturelles semblent également caractérisées par un rapport constant entre le nombre de sommets d'ordre r et le nombre de sommets d'ordre $(r+1)$ (cf. Figure 3 reproduite d'après [Mc Mahon, 1975])



SMALL POPLAR TREE, analyzed according to the Strahler system, has 960 first-order segments, 256 second-order segments, 68 third-order segments and 15 fourth-order segments that arise from a main stem bearing only two branches. As with the asymmetrical model shown on the opposite page, when the populations of segments are plotted on a semilogarithmic graph, the descending slope is approximately level. As the size of each population decreases moving from twig to trunk the average-diameter of members of each order increases. Ascending slope (color) that traces this increase in diameter is also approximately level.

Figure 3 : *Diagramme de Strahler : effectifs des populations et diamètre des segments.*

Nous présentons à la section 4 les statistiques concernant l'ordre des sommets dans les arbres.

2. DISTRIBUTION ET MOYENNE DE LA FONCTION NOMBRE DE REGISTRES.

Nous donnons ici des résultats de dénombrements explicites concernant la fonction nombre de registres. Les dénombrements s'expriment sous forme binomiale ou, de manière équivalente, trigonométrique. Comme conséquence, nous obtenons deux résultats d'équivalence de distribution entre hauteur et nombre de registres.

2.1. Réurrences de base.

Soit $\mathcal{B}$ la famille des arbres binaires, $\mathcal{R}_p$ la famille des arbres nécessitant exactement p registres, $\mathcal{S}_p$ la famille de ceux qui nécessitent au moins p registres. On désigne par $\mathcal{B}_n, \mathcal{R}_{p,n}, \mathcal{S}_{p,n}$ les sous-familles de $\mathcal{B}, \mathcal{R}_p, \mathcal{S}_p$ formées des arbres de taille n (comprenant n sommets internes); les dénombrements correspondants sont notés $B_n, R_{p,n}, S_{p,n}$ et les séries génératrices ordinaires sont notées

$$B(z) = \sum_{n \geq 0} B_n z^n; \quad R_p(z) = \sum_{n \geq 0} R_{p,n} z^n; \quad S_p(z) = \sum_{n \geq 0} S_{p,n} z^n.$$

De l'origine combinatoire de Reg résultent les égalités :

$$\begin{aligned} \text{(i)} \quad R_{0,0} &= 1 & R_{0,n} &= 0 & n \geq 1 \\ \text{(ii)} \quad R_{1,0} &= 0 & R_{1,n} &= 2^{n-1} & n \geq 1 \\ \text{(iii)} \quad R_{p,2^p-1} &= 1 & R_{p,n} &= 0 & n \geq 2^p \end{aligned}$$

où (ii) signifie que tout arbre calculable avec un registre est filiforme, donc déterminé par une suite de choix gauche/droite et (iii) traduit le fait que l'arbre parfait est l'arbre unique de taille minimum évaluable avec p registres.

On obtient ainsi les relations :

$$R_0(z) = 1 \quad R_1(z) = \frac{z}{1-2z} \quad \text{et} \quad \text{val}(R_p(z)) = 2^p - 1.$$

D'après la définition, on a $\mathcal{S}_p = \sum_{j \geq p} \mathcal{R}_j$ et $S_p(z) = \sum_{j \geq p} R_j(z)$, somme qui converge au sens des séries formelles. La définition inductive de Reg se traduit par l'équation :

$$\begin{cases} \mathcal{R}_0 = \square \\ \mathcal{R}_p = \mathcal{R}_{p-1} \begin{array}{c} \diagup \diagdown \\ \circ \end{array} \mathcal{R}_{p-1} + \mathcal{R}_p \begin{array}{c} \diagup \diagdown \\ \circ \end{array} \sum_{j < p} \mathcal{R}_j + \sum_{j < p} \mathcal{R}_j \begin{array}{c} \diagup \diagdown \\ \circ \end{array} \mathcal{R}_p, \end{cases}$$

soit pour les séries génératrices associées la récurrence :

$$\begin{cases} R_0 = 1 \\ R_p = zR_{p-1} + 2zR_p \sum_{j < p} R_j. \end{cases}$$

Cette récurrence montre que R_p se calcule rationnellement en fonction de $R_0, R_1, \dots, R_{p-1}$, et l'on obtient les premières valeurs

$$R_0 = 1 ; R_1 = \frac{z}{1-2z} ; R_2 = \frac{z^3}{1-6z+10z^2-4z^3} ;$$

$$R_4 = \frac{z^7}{1-14z+78z^2-220z^3+330z^4-252z^5+84z^6-8z^7}.$$

Cette même récurrence permet de calculer de proche en proche les premières valeurs des $R_{p,n}$ données par la figure 4.

$\begin{smallmatrix} n \\ p \end{smallmatrix}$	1	2	3	4	5	6	7	8	9	10
1	1	2	4	8	16	32	64	128	256	512
2			1	6	26	100	364	1288	4488	15504
3							1	14	118	780

Figure 4 : Tables des premières valeurs de $R_{p,n}$.

2.2. Expression des séries génératrices.

Nous résolvons ici la récurrence de base sur les R_p en la réduisant à une récurrence quadratique de forme particulière. Nous en déduisons des expressions explicites de R_p et S_p où interviennent les polynômes de Fibonacci (Tchebycheff).

PROPOSITION 1^A. Les quantités R_p et S_p sont données par les relations :

$$\frac{1}{2i \sin \phi} R_p = \frac{t^{2p}}{1-t^{2p+1}}, \quad \frac{1}{2i \sin \phi} S_p = \frac{t^{2p}}{1-t^{2p}},$$

où $\cos \frac{\phi}{2} = \frac{1}{2\sqrt{z}}$ et $t = e^{-i\phi} = zB^2(z)$.

Preuve : Partant de la récurrence de base sur les R_p :

$$R_0 = 1 ; R_p = zR_{p-1}^2 + 2zR_p \sum_{j < p} R_j,$$

on divise les deux membres par zR_p afin d'isoler le radical ; ainsi :

$$\frac{1}{z} = \frac{R_{p-1}^2}{R_p} + 2 \sum_{j < p} R_j ;$$

la relation homologue dans laquelle p est remplacé par $p+1$ s'écrit :

$$\frac{1}{z} = \frac{R_p^2}{R_{p+1}} + 2 \sum_{j < p+1} R_j.$$

Par soustraction de ces deux égalités membre à membre, on obtient

$$0 = \frac{R_p^2}{R_{p+1}} - \frac{R_{p-1}^2}{R_p^2} + 2R_p ,$$

et l'on a ainsi remplacé une récurrence sur la suite des valeurs par une récurrence du second ordre (non linéaire). La relation obtenue est homogène et peut se mettre en divisant par R_p sous la forme :

$$\frac{R_p}{R_{p+1}} = \left(\frac{R_{p-1}}{R_p} \right)^2 - 2 .$$

En posant $U_p = \frac{R_{p-1}}{R_p}$ pour tout $p \geq 1$, nous avons ainsi :

$$U_{p+1} = U_p^2 - 2.$$

Cette récurrence quadratique apparaît dans certains problèmes de théorie des nombres (tests de primalité des nombres de Mersenne). Pour la résoudre ici, nous utilisons la similitude entre cette récurrence et les formules trigonométriques de doublement d'arc :

$$\cos 2\theta = 2 \cos^2 \theta - 1.$$

On pose $U_p = 2 \cos \phi_p$; la récurrence devient

$$2 \cos \phi_{p+1} = 4 \cos^2 \phi_p - 2 = 2 \cos 2\phi_p.$$

Ainsi en définissant ϕ par $U_1 = 2 \cos \phi$, obtient-on la solution

$$U_p = 2 \cos 2^{p-1} \phi .$$

Le passage de z à ϕ est donné par la condition initiale

$$2 \cos \phi = U_1 = \frac{R_0}{R_1} = \frac{1}{z} - 1.$$

Cette relation peut s'exprimer de plusieurs manières différentes :

$$\begin{cases} 2 \cos \frac{\phi}{2} = \frac{1}{2\sqrt{z}} ; & i \sin \phi = \frac{\sqrt{1-4z}}{2z} ; \\ e^{-i\phi} = B(z)-1 = zB^2(z) ; & \text{ou encore } z = \frac{1}{4 \cos^2 \frac{\phi}{2}} . \end{cases}$$

On obtient R_p par le produit des quotients U_p :

$$\begin{aligned} \frac{R_0}{R_p} &= U_1 \cdot U_2 \dots U_p \\ &= (2 \cos \phi)(2 \cos 2\phi)(2 \cos 4\phi) \dots (2 \cos 2^{p-1} \phi). \end{aligned}$$

Cette expression se simplifie lorsqu'on la multiplie par $\sin \phi$:

$$\begin{aligned} \frac{\sin \phi}{R_p} &= \sin \phi (2 \cos \phi) (2 \cos 3\phi) \dots (2 \cos 2^{p-1}\phi) \\ &= \sin 2\phi (2 \cos 2\phi) (2 \cos 4\phi) \dots (2 \cos 2^{p-1}\phi) \\ &= \sin 4\phi (2 \cos 4\phi) \dots (2 \cos 2^{p-1}\phi) \\ &\dots = \sin 2^p \phi. \end{aligned}$$

Ainsi $R_p = \frac{\sin \phi}{\sin 2^p \phi}$, ce qui en posant $t = e^{-i\phi}$ s'écrit encore :

$$R_p = 2i \sin \phi \frac{t^{2^p}}{1-t^{2^{p+1}}}.$$

On obtient $S_p = \sum_{j \geq p} R_j$ par sommation. Pour cela on observe l'identité

$$\frac{u}{1-u^2} + \frac{u^2}{1-u^4} + \frac{u^4}{1-u^8} + \frac{u^8}{1-u^{16}} + \dots = \frac{u}{1-u},$$

qui traduit la propriété de tout entier d'être factorisable en une puissance de 2 et un nombre impair. En effet :

$$\sum_{j \geq 0} \frac{u^{2^j}}{1-u^{2^{j+1}}} = \sum_{j, k \geq 0} u^{2^j} u^{k \cdot 2^{j+1}} = \sum_{j, k \geq 0} u^{2^j (2k+1)} = \sum_{s \geq 0} u^s.$$

Le remplacement de u par t^{2^p} montre que

$$\frac{1}{2i \sin \phi} \sum_{j \geq p} R_j = \frac{t^{2^p}}{1-t^{2^p}},$$

d'où l'expression de S_p

$$S_p = 2i \sin \phi \frac{t^{2^p}}{1-t^{2^p}}. \square$$

PROPOSITION 1^B. Les fractions rationnelles $R_p(z)$ et $B(z) - S_p(z)$ ont

les expressions explicites suivantes :

$$R_p(z) = \frac{z^{2^p-1}}{F_{2^p+1}(z)} \quad B(z) - S_p(z) = \frac{F_{2^p-1}(z)}{F_{2^p}(z)}$$

où les polynômes $F_m(z)$ sont les polynômes de Fibonacci

$$F_m(z) = \sum_j \binom{m-1-j}{j} (-z)^j.$$

Preuve : Pour obtenir l'expression explicite de R_p , on part de la relation

$$R_p = \frac{\sin \phi}{\sin 2^p \phi} = \frac{2 \sin \frac{\phi}{2} \cos \frac{\phi}{2}}{\sin 2^p \phi} \quad \text{avec} \quad \cos \frac{\phi}{2} = \frac{1}{2\sqrt{z}}$$

L'expression de R_p suggère d'introduire les polynômes de Tchebycheff. Les polynômes T_m et U_m de première et seconde espèce sont définis par :

$$\cos m\theta = T_m(\cos \theta) ; \quad \frac{\sin m\theta}{\sin \theta} = U_m(\cos \theta),$$

et l'on a $\deg(T_m) = \deg(U_{m+1}) = m$. Les premières valeurs de $T_m(x)$ et $U_m(x)$ sont données par :

$$\begin{aligned} T_0 &= 1 ; T_1 = x ; T_2 = 2x^2 - 1 ; T_3 = 4x^3 - 3x ; T_4 = 8x^4 - 8x^2 + 1 \\ U_0 &= 0 ; U_1 = 1 ; U_2 = 2x ; U_3 = 4x^2 - 1 ; U_4 = 8x^3 - 4x \end{aligned}$$

Les identités trigonométriques

$$\begin{aligned} \cos(m+2)\theta + \cos m\theta &= 2 \cos \theta \cos(m+1)\theta \\ \frac{\sin(m+2)\theta}{\sin \theta} + \frac{\sin m\theta}{\sin \theta} &= 2 \cos \theta \frac{\sin(m+1)\theta}{\sin \theta}, \end{aligned}$$

fournissent pour T_m et U_m les récurrences

$$\begin{aligned} T_{m+2}(x) + T_m(x) &= 2xT_{m+1}(x) \\ U_{m+2}(x) + U_m(x) &= 2xU_{m+1}(x). \end{aligned}$$

Le retour à l'expression de R_p s'effectue en exprimant $\sin 2^p \phi$ en fonction de $\cos \frac{\phi}{2}$ au moyen du polynôme de Tchebycheff U_{2^p+1} :

$$R_p = 2 \cos \frac{\phi}{2} \frac{\sin \frac{\phi}{2}}{\sin 2^p \frac{\phi}{2}} = 2 \cos \frac{\phi}{2} \frac{1}{U_{2^p+1}(\cos \frac{\phi}{2})}, \text{ avec } \cos \frac{\phi}{2} = \frac{1}{2\sqrt{z}}.$$

Il est alors naturel d'introduire les variantes élémentaires des polynômes de Tchebycheff :

$$E_m(z) = z^{m/2} T_m\left(\frac{1}{2\sqrt{z}}\right) ; \quad F_m(z) = z^{(m-1)/2} U_m\left(\frac{1}{2\sqrt{z}}\right),$$

lesquels satisfont les récurrences

$$E_{m+2}(z) = E_{m+1}(z) - zE_m(z) ; \quad F_{m+2}(z) = F_{m+1}(z) - zF_m(z)$$

avec les valeurs initiales

$$E_0 = 1 ; E_1 = \frac{1}{2} ; F_0 = 0 ; F_1 = 1.$$

En reportant dans l'expression de R_p , on obtient ainsi :

$$R_p(z) = \frac{z^{2^p-1}}{U_{2^p+1}(z)}.$$

Partant des relations $B=1+t$ et $S_p=2i \sin \phi \frac{t^{2^p}}{1-t^{2^p}}$, avec $t=e^{-i\phi}$, on déduit l'expression

$$B-S_p = (1+t) \frac{1-t^{2^p-1}}{1-t^{2^p}},$$

soit en fonction de $u = e^{i\phi/2}$:

$$B-S_p = (u + \frac{1}{u}) \frac{u^{2^p-1} - u^{-2^p+1}}{u^{2^p} - u^{-2^p}},$$

ce qui équivaut à la forme trigonométrique

$$B-S_p = 2 \cos \frac{\phi}{2} \frac{\sin(2^p-1)\frac{\phi}{2}}{\sin 2^p \frac{\phi}{2}}.$$

Comme $2 \cos \frac{\phi}{2} = \frac{1}{\sqrt{2}}$, utilisant les variantes des polynômes de Tchebycheff, on obtient l'expression

$$B(z) - S_p(z) = \frac{F_{2^p-1}(z)}{F_{2^p}(z)}.$$

Les polynômes $F_m(z)$ sont appelés polynômes de Fibonacci par [Kreweras, 1970] dont nous suivons ici les notations. La relation de récurrence

$$F_{m+2}(z) = F_{m+1}(z) + zF_m(z); \quad F_0(z) = 0; \quad F_1(z) = 1,$$

conduit pour la série génératrice de ces polynômes

$$F(u; z) = \sum F_m(z) u^m,$$

à l'expression

$$F(u; z) = \frac{u}{1 - u + zu^2},$$

qui se développe en série, donnant

$$F(u; z) = u \sum_{n \geq 0} u^n (1-zu)^n = \sum_{\substack{m \geq 0 \\ k \geq 0}} \binom{n}{k} (-z)^k u^{n+1+k}$$

$$F(u; z) = \sum \binom{m-1-k}{k} (-z)^k u^m,$$

Il s'ensuit l'expression explicite des polynômes $F_m(z)$. $\square$

Remarquons que les dénominateurs des $R_p(z)$ sont construits sur les diagonales du triangle de Pascal prises aux puissances de 2, comme l'indique la figure 5.

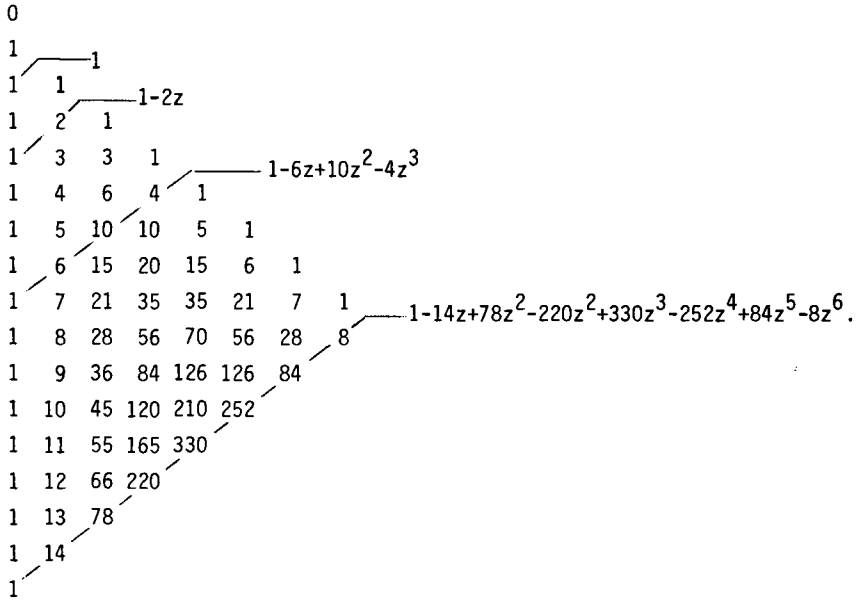


Figure 5 : Construction de $R_p(z)$.

Les polynômes de Fibonacci sont ainsi appelés car pour $z=-1$, $F_m(z)$ vaut le nombre de Fibonacci F_m . De l'identité entre séries formelles

$$B(z) = \sum_{p \geq 0} \frac{z^{2^p-1}}{F_{2^{p+1}}(z)},$$

qui traduit la partition des arbres selon la valeur du nombre de registres, il s'ensuit en faisant $z=-1$, l'égalité numérique

$$\frac{\sqrt{5}-1}{2} = 1 - \frac{1}{F_4} - \frac{1}{F_8} - \frac{1}{F_{16}} \dots$$

Ce développement du nombre conjugué du nombre d'or dont la convergence double exponentielle est extrêmement rapide était connu de [Lucas, 1891].

Une autre application de nature combinatoire concerne l'identité de la distribution du nombre de registres sur les arbres binaires et de la distribution de certaines fonctions de la hauteur sur les arbres généraux.

Corollaire : (i) Le nombre $B_{n-S_{p,n}}$ d'arbres binaires de taille n

évaluables avec au moins p registres est égal au nombre d'arbre généraux de taille $(n+1)$ dont la hauteur[†] est strictement inférieure à 2^p-1 .

(ii) Le nombre $R_{p,n}$ d'arbres binaires de taille n évaluables avec exactement p registres est égal au nombre d'arbres généraux de taille $n+2^p$ de hauteur égale à 2^p-2 et dont la branche la plus à gauche est de longueur 2^p-2 .

Preuve : La série génératrice des arbres de hauteur $\leq h$ vaut

$G^{[h]}(z) = z \frac{F_{h+1}(z)}{F_{h+2}(z)}$; la série génératrice des arbres de hauteur égale à h , et dont la branche gauche a longueur h vaut $D^{[h]}(z) = \frac{z^{h+1}}{F_{h+2}(z)}$ (cf. chapitre I).

Il suffit de comparer ces séries aux expressions

$$R_p(z) = \frac{z^{2^p-1}}{F_{2^p+1}(z)} \quad \text{et} \quad B(z) - S_p(z) = \frac{F_{2^p-1}(z)}{F_{2^p}(z)}. \quad \square$$

Ces relations suggèrent l'existence d'une correspondance entre arbres généraux et arbres binaires qui transformerait par pliage une branche de longueur h maximale d'un arbre général en un arbre parfait plongé dans un arbre binaire et de hauteur $\lceil \log_2 h \rceil$. Une telle correspondance n'est pas connue mais [Françon, 1977] a donné de la partie (i) de ce résultat une preuve purement géométrique fondée sur les correspondances arbres-chemins.

Pour déterminer le nombre moyen de registres correspondant à l'ensemble des arbres de taille n , on utilise la proposition suivante :

PROPOSITION 1^C. Soit T_n le nombre total de registres de l'ensemble des arbres de taille n . La série génératrice $T(z) = \sum T_n z^n$ s'exprime par :

$$T(z) = 2i \sin \phi \sum_{k \geq 1} v_2(k) t^k, \quad \text{où } t = e^{-i\phi} = zB^2(z) \quad \text{et} \quad \cos \phi = \frac{1}{2z} - 1.$$

La quantité $v_2(k)$ est la valuation dyadique^{††} de k .

Preuve : Par définition de T_n , on a

[†] La hauteur est mesurée en nombre d'arêtes sur la branche la plus longue.

^{††} La valuation dyadique de k , $v_2(k)$, est l'unique entier tel que $k = 2^{v_2(k)}(2q+1)$ avec $q \in \mathbb{N}$.

$$T_n = \sum_{p \geq 0} p R_{p,n} \text{ et par conséquent } T(z) = \sum_{p \geq 0} p R_p(z).$$

L'expression de R_p : $R_p = 2i \sin \phi \frac{t^{2p}}{1-t^{2p+1}}$, montre :

$$\frac{1}{2i \sin \phi} T(z) = \sum_{p \geq 0} p \frac{t^{2p}}{1-t^{2p+1}}.$$

La somme de droite se réécrit grâce à l'égalité :

$$1 \frac{t^2}{1-t^4} + 2 \frac{t^4}{1-t^8} + 3 \frac{t^8}{1-t^{16}} + \dots = \sum_{k \geq 1} v_2(k) t^k,$$

laquelle s'obtient simplement en développant :

$$\sum_{p \geq 0} p \frac{t^{2p}}{1-t^{2p+1}} = \sum_{p, s \geq 0} p t^{2p} t^{s2p+1} = \sum_{p, k \geq 0} p t^{2p(2s+1)}.$$

Ainsi chaque terme $t^{2p(2s+1)}$ est pris avec le coefficient p , ou encore chaque terme t^k est pris avec le coefficient $v_2(k)$. $\square$

2.3. Dénombrements et résultats de distribution.

Nous donnons maintenant des expressions explicites des coefficients des séries génératrices $R_p(z)$, $S_p(z)$ et $T(z)$ précédemment calculées. D'une part, ces séries s'expriment en fonction de $B(z) = \frac{1-\sqrt{1-4z}}{2z}$ dont les développements des puissances sont connus ; d'autre part R_p et $B-S_p$ sont des fractions rationnelles dont les dénominateurs, liés aux polynômes de Tchebycheff, ont des zéros exprimables explicitement en fonction de racines de l'unité. Ainsi chaque dénombrement présente-t-il une forme binomiale et une forme trigonométrique.

THEOREME 2^A. Les nombres $R_{p,n}$, $S_{p,n}$ et T_n ont les expressions explicites :

$$\begin{aligned} R_{p,n} &= \sum_{q \in I} \left[\binom{2n}{n+1-q} - \binom{2n}{n-q} \right] \quad \text{où} \quad I = \{2^p(2k+1) \mid k \in \mathbb{Z}\} \\ S_{p,n} &= -B_n + \sum_{q \in J} \left[\binom{2n}{n+1-q} - \binom{2n}{n-q} \right] \quad \text{où} \quad J = \{2^p \cdot k \mid k \in \mathbb{Z}\} \\ T_n &= \sum_{q \in K} v_2(q) \left[\binom{2n}{n+1-q} - \binom{2n}{n-q} \right] \quad \text{où} \quad K = \mathbb{Z} - \{0\}. \end{aligned}$$

Preuve : La preuve plus simple diffère de celle que nous avons donné précédemment dans [Flajolet, Raoult, Vuillemin ; 1978]. Contrairement aux méthodes utilisées par de Bruijn et al., elle n'utilise pas l'analyse complexe.

Les polynômes de Fibonacci satisfont à la récurrence linéaire

$$F_{m+2}(z) = F_{m+1}(z) - F_m(z),$$

dont l'équation caractéristique

$$y^2 = y - z$$

admet comme solution les séries

$$G(z) = \frac{1 - \sqrt{1-4z}}{2} \quad \text{et} \quad \bar{G}(z) = \frac{1 + \sqrt{1-4z}}{2}.$$

Le polynôme F_m s'exprime sous la forme

$$F_m(z) = \frac{\bar{G}^m - G^m}{G - \bar{G}} = \frac{\bar{G}^m}{\sqrt{1-4z}} \left(1 - \left(\frac{G}{\bar{G}}\right)^m\right).$$

Avec $B(z) = \frac{1 - \sqrt{1-4z}}{2z} = \frac{1}{z} G(z)$, le quotient $\frac{G(z)}{\bar{G}(z)}$ se réécrit $zB^2(z)$ et $\frac{1}{\bar{G}(z)}$ vaut $B(z)$. On en déduit pour $\frac{1}{F_m(z)}$ l'expression :

$$\frac{1}{F_m(z)} = \sqrt{1-4z} \frac{B^m(z)}{1 - z^m B^{2m}(z)}.$$

On peut remplacer $\sqrt{1-4z}$ par $1 - 2zB(z)$, et développer la fraction d'où :

$$\frac{1}{F_m(z)} = (1 - 2zB(z)) \sum_{k \geq 0} z^{km} B^{2k+1}(z)^m.$$

En reportant dans l'expression de R_p , on obtient :

$$R_p(z) = \sum_{k \geq 0} z^{(2k+1)2^p-1} (B(z))^{(2k+1)2^{p+1}} - 2 \sum_{k \geq 0} z^{(2k+1)2^p} (B(z))^{(2k+1)2^{p+1}+1}.$$

Rappelons le développement des puissances de B :

$$B^m(z) = \sum_{r \geq 0} B_r^{(m)} z^r \quad \text{avec} \quad B_r^{(m)} = \binom{2r+m-1}{r} - \binom{2r+m-1}{r-1};$$

appliqué à l'expression développée de R_p , il montre que

$$R_{p,n} = \sum_{k \geq 0} \left[\binom{2n+1}{n+1-2^p(2k+1)} - \binom{2n+1}{n-2^p(2k+1)} \right] - 2 \sum_{k \geq 0} \left[\binom{2n}{n-2^p(2k+1)} - \binom{2n}{n-1-2^p(2k+1)} \right],$$

ce qui tenant compte de la récurrence de base des coefficients binomiaux se ramène à

$$R_{p,n} = \sum_{q \in I'} \left[\binom{2n}{n+1-q} - 2 \binom{2n}{n-q} + \binom{2n}{n-1-q} \right] \quad \text{où} \quad I' = \{2^p(2k+1) \mid k \geq 0\}.$$

Cette forme est clairement équivalente à celle de l'énoncé.

On traite le développement de la série S_p de manière tout à fait analogue. Enfin l'expression de $T(z)$ en fonction des puissances de B est donnée par le théorème 1^C ; il s'ensuit le développement de l'énoncé où la valuation dyadique est étendue naturellement de $\mathbb{N}^+$ à $\mathbb{Z} - \{0\}$ par $v_2(-k) = v_2(k)$. $\square$

Ces formes binômiales des coefficients sont équivalentes à des formes trigonométriques. Nous en avons donné précédemment une preuve, fondée sur la décomposition en éléments simples des fractions $R_p(z)$ et $B(z)-S_p(z)$.

THEOREME 2^B. Les nombres $R_{p,n}$ et $S_{p,n}$ ont les expressions explicites :

$$R_{p,n} = \frac{4^n}{2^{p-1}} \sum_{1 \leq k \leq 2^p} (-1)^{k+1} \sin^2 \frac{k\pi}{2^{p+1}} \cos^{2n} \frac{k\pi}{2^{p+1}}$$

$$B_n - S_{p,n} = \frac{4^{n+1}}{2^p} \sum_{1 \leq k < 2^{p+1}} \sin^2 \frac{k}{2^p} \cos^{2n} \frac{k\pi}{2^p}.$$

Preuve : Les sommes de coefficients binômiaux dont les indices inférieurs sont en progression arithmétique sont équivalentes à des sommes de puissances de racines de l'unité. Cette équivalence, parfois connue sous le nom de "formule de Ramus" est l'application au développement du binôme des procédés de multi-section de séries et généralise l'égalité classique

$$2^{n-1} = \frac{1}{2} [(1+1)^n + (1-1)^n] = \binom{n}{0} + \binom{n}{2} + \binom{n}{4} \dots$$

On part des sommes

$$M_{n,h} = \sum_{k \in \mathbb{Z}} \binom{2n}{n-kh},$$

et l'on pose $v = n \bmod h$; ainsi :

$$M_{n,h} = \sum_{k \geq 0} \binom{2n}{v+kh}.$$

Soit ω une racine primitive h -ième de l'unité, on considère les développements binomiaux de $(1+\omega^0)^{2n}$, $(1+\omega^1)^{2n}$, $(1+\omega^2)^{2n}$..., dont on fait la somme pondérée par les multiplicateurs ω^{0v} , ω^{1v} , ω^{2v} ..., selon le schéma :

développements binomiaux	multiplicateurs
$(1+\omega^0)^{2n} = \sum \binom{2n}{j} \omega^{0j}$	$\omega^{-0v} = \omega^{-0n}$
$(1+\omega^1)^{2n} = \sum \binom{2n}{j} \omega^{1j}$	$\omega^{-1v} = \omega^{-1n}$
$(1+\omega^2)^{2n} = \sum \binom{2n}{j} \omega^{2j}$	$\omega^{-2v} = \omega^{-2n}$
$\vdots$	$\vdots$
$(1+\omega^{h-1})^{2n} = \sum \binom{2n}{j} \omega^{(h-1)j}$	$\omega^{-(h-1)v} = \omega^{-(h-1)n}$

La sommation des développements binomiaux fait disparaître tous les $\binom{2n}{j}$ tels que j n'est pas de la forme $v+kh$; les coefficients $\binom{2n}{v+kh}$ apparaissent h fois. Ainsi :

$$\begin{aligned} h \sum_{k \geq 0} \binom{2n}{v+kh} &= \sum_{0 \leq j < h} (1+\omega^j)^{2n} \omega^{-jn} \\ &= \sum_{0 \leq j < h} (\omega^{j/2} + \omega^{-j/2})^{2n} \\ &= 4^n \sum_{0 \leq j < h} \cos^{2n} \frac{j\pi}{h}. \end{aligned}$$

De la même manière, on calcule les sommes $M_{n,h}(\alpha) = \sum_{k \in \mathbb{Z}} \binom{2n}{n+\alpha-kh}$ pour lesquelles on obtient l'expression :

$$h M_{n,h}(\alpha) = \sum_{0 \leq j < h} (\omega^{s/2} + \omega^{-j/2})^{2n} \cdot \omega^{-j\alpha}.$$

L'application à $R_{p,n}$, par exemple, s'effectue en prenant $h = 2^{p+1}$, $\alpha = 2^p - 1$ puis 2^p . On obtient ainsi :

$$2^{p+1} R_{p,n} = \sum_{0 \leq j < h} (\omega^{s/2} + \omega^{-j/2})^{2n} (\omega^j - 1) \omega^{-jh/2}.$$

Comme $\omega^{h/2} = -1$, on a encore

$$\begin{aligned} 2^{p+1} R_{p,n} &= \sum_{0 \leq j < h} (\omega^{j/2} + \omega^{-j/2})^{2n} (-1)^j (\omega^j - 1) \\ &= \sum_{0 \leq j < h/2} (\omega^{j/2} + \omega^{-j/2})^{2n} (-1)^j (\omega^{j-2+\omega^{-j}}) \end{aligned}$$

en tenant compte de la symétrie due à la parité de h . L'énoncé découle du remplacement de $\omega^{j-2+\omega^{-j}}$ par $(\omega^{j/2} - \omega^{-j/2})^2$ et du passage aux formes trigonométriques correspondantes. $\square$

3. ANALYSE ASYMPTOTIQUE : DISTRIBUTION ET MOYENNE.

Nous montrons dans cette section l'existence d'une infinité de lois limites pour la distribution du nombre de registres sur l'ensemble des arbres de taille n ; ces lois limites dépendent de la partie fractionnaire de $\log_4 n$ et mettent en évidence des distributions de même forme pour n et $4n$. Nous effectuons ensuite une analyse asymptotique détaillée du nombre moyen de registres $\bar{R}_n$, laquelle repose sur certaines expressions closes décrivant la distribution des 1 dans les représentations binaires d'entiers. Il apparaît que l'évaluation asymptotique de $\bar{R}_n$ fait intervenir un terme périodique en $\log_4 n$. Ce terme s'exprime à partir de la fonction de Delange qui est continue et nulle part dérivable. Enfin le terme périodique dans le développement de R_n a des coefficients de Fourier qui s'expriment simplement au moyen des valeurs des fonctions $\zeta(s)$ de Riemann et $\Gamma(s)$ d'Euler le long de l'axe imaginaire.

3.1. Distribution du nombre de registres.

Le point de départ des évaluations asymptotiques est l'approximation classique des coefficients binomiaux par une loi de Gauss. Ainsi pour k dans le domaine $[0 ; \sqrt{n} \ln n]$, a-t-on l'expression

$$\frac{\binom{2n}{n+k}}{\binom{2n}{n}} = e^{-k^2/n} (1 + O(\frac{(\ln n)^*}{\sqrt{n}})),$$

où $O(u^*)$ s'interprète comme $O(u^\alpha)$ pour un certain $\alpha > 0$. Ce résultat se vérifie simplement comme conséquence de la formule de Stirling ou encore en développant en série le logarithme du premier membre.

Cette approximation s'étend aux différences finies de coefficients binomiaux qu'on peut approcher par des dérivées de la loi de Gauss et un calcul simple montre :

PROPOSITION. Les différences[†] de coefficients binomiaux $\Delta^r \binom{2n}{n+k}$

[†] Cette écriture, systématisée en 3.2, est la notation classique des différences finies. Ici $\Delta^r \binom{2n}{n+k} = \binom{2n}{n+k+r} - \binom{r}{1} \binom{2n}{n+k+r-1} + \binom{r}{2} \binom{2n}{n+k+r-2} \dots$

admettent pour $k \leq \sqrt{n} \ln n$ l'estimation :

$$\Delta^r \binom{2n}{n+k} = \binom{2n}{n} \frac{e^{-k^2/n}}{n^{r/2}} \left[H_r\left(\frac{k}{\sqrt{n}}\right) + O\left(\frac{(\ln n)^*}{\sqrt{n}}\right) \right],$$

où $H_k(x) = e^{x^2} \frac{d^k}{dx^k} e^{-x^2}$ est le k -ième polynôme d'Hermite. $\square$

On montre encore de la même manière que pour $k > \sqrt{n} \ln n$, les coefficients $\binom{2n}{n+k}$ sont exponentiellement petits devant $\binom{2n}{n}$, la propriété valant également pour les différences.

THEOREME 3^A. Lorsque p est extérieur à l'intervalle

$[\log_4 n - \log_2 \ln n ; \log_4 n + \log_2 \ln n]$, la proportion d'arbres calculable avec p registres est exponentiellement petite :

$$\frac{R_{p,n}}{B_n} = O\left(\frac{1}{n^\alpha}\right) \quad \text{pour tout } \alpha \geq 0.$$

Preuve : Lorsque $2^p > \sqrt{n} \ln n$, chacun des termes qui figure dans l'expression binomiale de $R_{p,n}$ est exponentiellement petit avec un comportement en $O(e^{-\ln^2 n})$; comme ces termes binomiaux sont en nombre au plus $4n$, il s'ensuit que pour $p > \log_4 n + \log_2 \ln n$, $\frac{R_{p,n}}{B_n} = O(ne^{-\ln^2 n})$; de là découle la première partie du théorème.

Lorsque $2^p < \sqrt{n} \ln n$, on utilise au contraire l'expression trigonométrique

$$R_{p,n} = \frac{4^n}{2^{p-1}} \sum_{1 \leq k < 2^p} (-1)^{k+1} \sin^2 \frac{k\pi}{2^{p+1}} \cos^{2n} \frac{k\pi}{2^{p+1}}.$$

En majorant les sinus par 1, les cosinus par $\cos \frac{\pi}{2^{p+1}}$, on obtient :

$$\frac{R_{p,n}}{B_n} \leq \frac{4^n}{B_n} 2 \cos^{2n} \frac{\pi}{2^{p+1}} = (n^{3/2}) \cos^{2n} \frac{\pi}{2^{p+1}}.$$

Pour $x \in [-2; +2]$, on vérifie l'inégalité $\cos x < 1 - \frac{x^2}{4}$, d'où :

$$\ln \cos^{2n} \frac{\pi}{2^{p+1}} = 2n \ln \cos \frac{\pi}{2^{p+1}} < 2n \ln \left(1 - \frac{\pi^2}{2^{2p+4}}\right).$$

Comme $\ln(1-x) < -x$ pour $x < 1$, on en déduit la majoration

$$\ln \cos^{2n} \frac{\pi}{2^{p+1}} < -\frac{\pi^2}{8} \frac{n}{2^p} < -\frac{\pi^2}{8} \ln^2 n.$$

D'où finalement pour $p < \log_4 n - \log_2 \ln n$

$$\frac{R_{p,n}}{B_n} = O\left(n^{3/2} e^{-\frac{\pi^2}{8} \ln^2 n}\right) = O\left(\frac{1}{n^\alpha}\right) \quad \text{pour tout } \alpha > 0. \quad \square$$

Comme conséquence du caractère très pointu de la distribution des $R_{p,n}$ on obtient immédiatement :

PROPOSITION. (i) Le nombre moyen de registres relatif à l'ensemble des arbres de taille n vaut

$$\bar{R}_n = \log_4 n + O(\log_2 \ln n).$$

(ii) L'écart type du nombre de registres sur le même ensemble σ_n admet la majoration

$$\sigma_n < 2 \log_2 \ln n. \quad \square$$

On montre à la section suivante comment obtenir une expression asymptotique précise de $\bar{R}_n$ et on traite au chapitre suivant de l'évaluation de σ_n . Revenant au problème de distribution on a :

THEOREME 3^B. La distribution du nombre de registres admet la "loi-limite" :

$$\frac{R_{p,n}}{B_n} = \lambda\left(\frac{2^p}{\sqrt{n}}\right) + O\left(\frac{\ln^* n}{\sqrt{n}}\right) \quad \text{pour } p \in [\log_4 n - \log_2 \ln n; \log_4 n + \log_2 \ln n]$$

où la fonction λ , reliée aux fonctions θ de la théorie des fonctions elliptiques est donnée par :

$$\lambda(x) = \sum_{k \text{ impair}} e^{-k^2 x^2} (4k^2 x^2 - 2).$$

Preuve : On part de l'expression binomiale de $R_{p,n}$

$$R_{p,n} = \sum_{q \in I'} \left[\binom{2n}{n+1-q} - 2 \binom{2n}{n-q} + \binom{2n}{n-1-q} \right] \quad \text{où } I' = \{2^p(2k+1) | k \geq 0\}.$$

On sépare la somme selon l'ensemble des indices :

$$I_1 = I \cap [0; \sqrt{n} \ln n[; \quad I_2 = I \cap [\sqrt{n} \ln n; +\infty[,$$

puis on applique aux termes correspondant à I_1 l'approximation gaussienne. Les termes correspondant à I_2 sont exponentiellement petits, et ainsi pour tout

$\alpha > 0$ en posant $x = \frac{2^p}{\sqrt{n}}$

$$\begin{aligned} \frac{R_{p,n}}{B_n} &= \sum_{2^p(2k+1) \in I_1} e^{-(2k+1)^2 x^2} (4(2k+1)^2 x^2 - 2) \\ &\quad + \sum_{2^p(2k+1) \in I_1} e^{-2(2k+1)^2 x^2} (4(2k+1)^2 x^2 - 2) O\left(\frac{\ln^* n}{\sqrt{n}}\right) \\ &\quad + O\left(\frac{1}{\alpha}\right). \end{aligned}$$

Etant donné que $\sqrt{n}/\ln n < 2^p < \sqrt{n} \ln n$, le nombre de termes intervenant dans les deux premières sommes qui vaut $\text{card}(I_1)$ est borné par $2\ln^2 n$. Ainsi la seconde somme vaut elle globalement $O(\frac{\ln^* n}{\sqrt{n}})$, d'où :

$$\frac{R_{p,n}}{B_n} = \sum_{2^p(2k+1) \in I_1} e^{-(2k+1)^2 x^2 (4(2k+1)^2 x^2 - 2)} + O(\frac{\ln^* n}{\sqrt{n}}).$$

On peut enfin, en n'introduisant qu'un terme exponentiellement petit, remplacer l'intervalle de sommation par $[0; +\infty]$, d'où :

$$\frac{R_{p,n}}{B_n} = \sum_{k \text{ impair}} e^{-k^2 x^2 (4k^2 x^2 - 2)} + O(\frac{\ln^* n}{\sqrt{n}}),$$

ce qui donne le résultat de l'énoncé. $\square$

Le graphe de la fonction λ est donnée par la figure 6

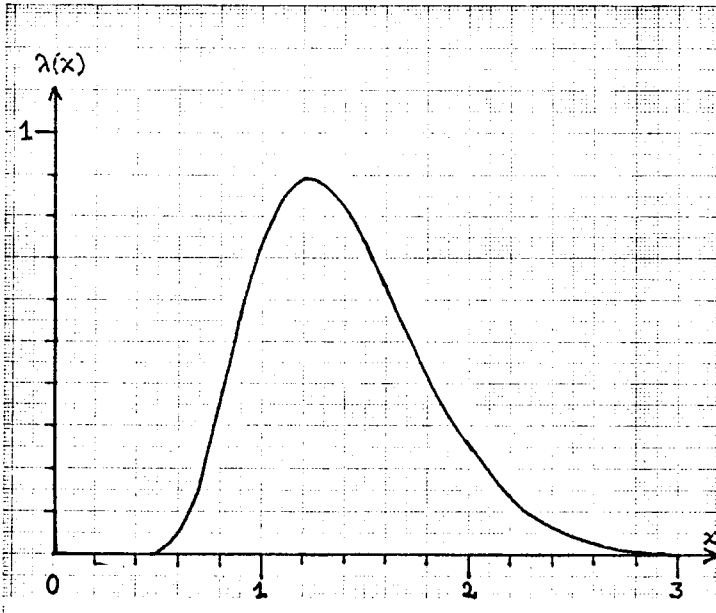


Figure 6 : Graphe de la fonction λ .

Soit $E(\alpha)$ l'ensemble des nombres qui s'écrivent sous la forme $4^{m+\alpha}$ avec m entier et $0 \leq \alpha < 1$:

$$E(\alpha) = \{n \in \mathbb{N} \mid \log_4 n - \lfloor \log_4 n \rfloor = \alpha\}.$$

Pour chaque α , lorsque n tend vers l'infini dans $E(\alpha)$, la distribution de la

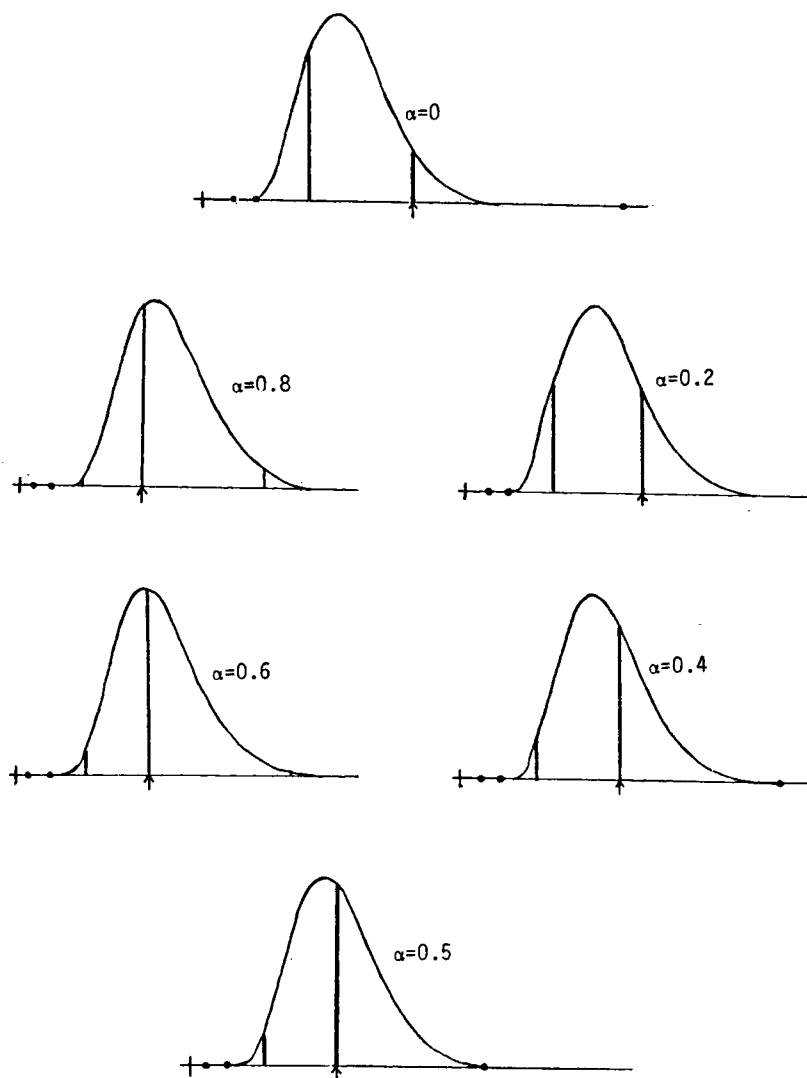


Figure 7 : Périodicité de la distribution limite du nombre de registres selon $\alpha = \log_4 n - \lfloor \log_4 n \rfloor$. La figure représente les valeurs $\alpha = 0 ; 0.2 ; 0.4 ; 0.5 ; 0.6 ; 0.8$. Le terme central correspondant à $\lfloor \log_4 n \rfloor$ est flêché.

variable $s = \text{Reg}(T) - \lfloor \log_4 |T| \rfloor$ tend vers une loi limite discrète donnée par $\{\lambda(2^{-\alpha} \cdot 2^s)\}_{s \in \mathbb{Z}}$. Ainsi recentrée autour de $\lfloor \log_4 n \rfloor$ la distribution du nombre de registres est donnée par un échantillonnage en des abscisses en progression géométrique de la fonction λ . Nous avons tracé les histogrammes correspondant aux valeurs de $\alpha = 0, 0.2, 0.4, 0.6, 0.8$ (figure 7). A noter que d'après l'origine du problème λ vérifie l'équation fonctionnelle

$$\sum_{s=-\infty}^{+\infty} \lambda(x2^{-s}) = 1 \quad \text{pour tout } x.$$

3.2. Evaluation asymptotique de la moyenne.

Une difficulté dans l'obtention d'une estimation précise de la moyenne provient de la présence de la valuation dyadique que multiplient des coefficients binomiaux. On peut lisser les termes de la somme donnant T_n , en utilisant la transformation d'Abel (ou sommation par parties discrète). Ceci conduit à une expression qui fait intervenir la fonction somme-des-chiffres pour laquelle existent des formules closes dues à Delange.

Rappelons ici la définition des opérateurs de base du calcul des différences. Pour une suite $\{a_n\}_{n \geq 0}$, on définit les opérateurs de

$$\text{différence : } \Delta a_n = a_{n+1} - a_n$$

$$\text{sommation}^\dagger : \Sigma a_n = a_0 + a_1 + \dots + a_{n-1}$$

$$\text{décalage : } E a_n = a_{n+1}.$$

Par réarrangement des termes, on a le lemme d'Abel ou formule de sommation par parties :

$$a_n(\Delta b_n) = [ab]_0^n - \Sigma(Eb_n)(\Delta a_n), \quad \text{où } [ab]_0^n = a_n b_n - a_0 b_0.$$

Partant de l'expression

$$T_n = \sum_{k>0} v_2(k) \left[\binom{2n}{n+k+1} - 2 \binom{2n}{n+k} + \binom{2n}{n+k-1} \right],$$

qu'on écrit avec un léger abus de notation

$$T_n = \sum_{k>0} v_2(k) \Delta^2 \binom{2n}{n+k-1},$$

[†] encore appelé opérateur de primitive discrète.

l'application du lemme d'Abel conduit à la forme équivalente

$$T_n = - \sum_{k>0} w(k) \Delta^3 \binom{2n}{n+k-1},$$

où w est la primitive discrète de v_2 :

$$w(k) = \sum_{j \leq k} v_2(j).$$

Cette expression se simplifie : considérons la représentation binaire de deux entiers consécutifs :

$$\text{BIN}(j-1) = u \ 0 \ 1 \ 1 \ \dots \ 1$$

$$\text{BIN}(j) = u \ 1 \ 0 \ 0 \ \dots \ 0,$$

avec $u \in \{0,1\}^*$. Le nombre de zéros de queue dans $\text{BIN}(j)$ est précisément $v_2(j)$.

Si l'on désigne par $v(j)$ le nombre de chiffres 1 dans la représentation binaire de j , on a ainsi l'égalité

$$v(j-1) - v(j) = v_2(j) - 1,$$

soit par sommation :

$$w(k) = \sum_{0 \leq j \leq k} v_2(j) = k - v(k).$$

Le comportement de w est moins erratique que celui de v_2 puisqu'on a $k - \lfloor \log_2(k+1) \rfloor \leq w(k) \leq k$. En remplaçant dans l'expression de T_n , on obtient

$$T_n = B_n + \sum_{k>0} v(k) \Delta^3 \binom{2n}{n+k-1}.$$

Une nouvelle sommation par parties conduit à l'expression

$$T_n = B_n - \sum_{k>0} S(k) \Delta^4 \binom{2n}{n+k-2},$$

où S est la primitive discrète de v :

$$S(k) = \sum_{0 \leq j < k} v(j) ;$$

en d'autres termes $S(k)$ vaut le nombre de chiffres 1 dans la représentation binaire des entiers $0, 1, 2, \dots, k-1$.

La fonction S est encore plus régulière ; un argument de comptage simple montre que

$$S(m) \sim \frac{1}{2} m \log_2 m,$$

ce qui est conforme à l'intuition : la représentation des entiers de 0 à $m-1$ comporte $1 + \lfloor \log_2 m \rfloor$ positions binaires comprenant asymptotiquement autant de 1

que de 0. Un raffinement de cet argument fondé sur le comportement de $S(2^n)$ montre l'expression plus précise

$$S(m) = \frac{1}{2} m \log_2 m + O(m).$$

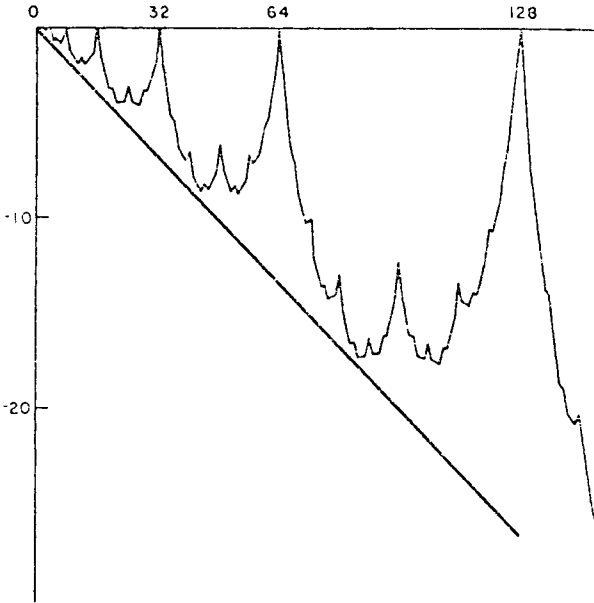


Figure 8 : Graphe des valeurs de $S(m) - \frac{1}{2} m \log_2 m$. La pente de la droite qui intervient comme borne inférieure est $\log_2 \frac{3}{4}$.

Le comportement de $S(m)$ tel que tabulé par [Mc Ilroy, 1974] est représenté à la figure 8.

Par des méthodes que nous discutons et étendons au chapitre suivant à d'autres systèmes de représentation d'entier, [Delange, 1975] a montré

THEOREME. Il existe une fonction F telle que pour tout $m \geq 0$

$$S(m) = \frac{1}{2} m \log m + mF(\log m)^\dagger$$

F étant continue, périodique de période 1, et nulle part dérivable.

[†] On notera désormais $\log$ pour $\log_2$.

La fonction F vaut

$$F(u) = \frac{1}{2} c(u) + 2^{c(u)} h(2^{-c(u)}) \quad \text{où } c(u) = 1 + [u] - u$$

est le complément de la partie fractionnaire de u , et où h est une superposition de fonctions triangulaires comme indiqué par la figure 9.

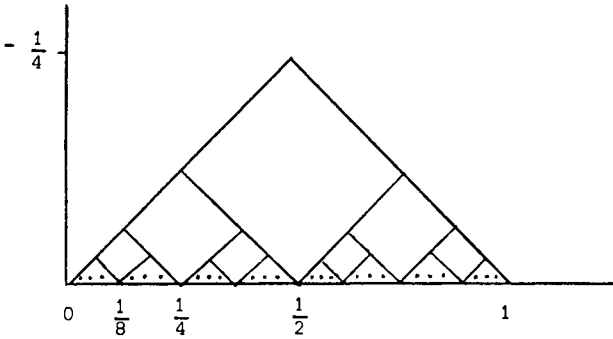


Figure 9 : La fonction h de Delange.

L'article de Delange contient encore le résultat de non différentiabilité suggéré par le diagramme de Mc Ilroy, ainsi que le développement en série de Fourier de la fonction F que nous utiliserons à la section suivante.

Revenant à l'expression de T_n

$$T_n = B_n - \sum_{k \geq 0} S(k) \Delta^4 \binom{2n}{n+k-2},$$

on utilise comme précédemment l'approximation gaussienne des termes centraux, ce qui conduit à :

$$T_n = B_n - \frac{1}{n^2} \binom{2n}{n} \sum_{0 \leq k < \sqrt{n}} \frac{S(k)}{\ln n} e^{-k^2/n} H_4\left(\frac{k}{\sqrt{n}}\right) + B_n O\left(\frac{\ln^* n}{\sqrt{n}}\right),$$

où le terme correcteur regroupe le terme d'erreur de l'approximation gaussienne et le terme d'erreur (beaucoup plus petit) provenant des coefficients binomiaux non centraux.

Afin d'approcher cette somme par une intégrale introduisons la

partition $P = \{0, \frac{1}{\sqrt{n}}, \frac{2}{\sqrt{n}}, \dots, \ln n\}$; on peut remplacer, dans la somme donnant T_n

l'entier k par $t\sqrt{n}$, avec t parcourant P . En utilisant l'expression de $S(k)$ qui résulte du théorème de Delange, on a ainsi :

$$T_n = B_n - \frac{1}{n^2} \sum_{t \in P} \left[\frac{t\sqrt{n}}{2} \log t\sqrt{n} + t\sqrt{n} F\left(\log t + \frac{1}{2} \log n\right) \right] e^{-t^2} H_4(t) + B_n O\left(\frac{\ln^* n}{\sqrt{n}}\right).$$

Soit pour $\frac{T_n}{B_n}$ l'expression

$$\frac{T_n}{B_n} = 1 - \log n I_1 - I_2 - I_3 + O\left(\frac{\ln^* n}{\sqrt{n}}\right),$$

avec

$$I_1 = \frac{1}{\sqrt{n}} \sum_{t \in P} \frac{t}{4} H_4(t) e^{-t^2}$$

$$I_2 = \frac{1}{\sqrt{n}} \sum_{t \in P} \frac{t}{2} \log t H_4(t) e^{-t^2}$$

$$I_3 = \frac{1}{\sqrt{n}} \sum_{t \in P} t F(\log t + \log_4 n) H_4(t) e^{-t^2}.$$

Chacune de ces sommes est une somme de Riemann d'une fonction continue sur l'intervalle $[0, \ln n]$. On peut remplacer ces sommes par les intégrales qu'elles approchent les termes d'erreur étant donnés par le lemme suivant :

LEMME. Soit f une fonction continue sur l'intervalle $[a; b]$ et soit $a = x_0, \dots, x_n = b$ une subdivision de l'intervalle. Si f est intégrable (au sens de Riemann) sur $[a, b]$ alors :

$$\left| \int_a^b f(x) dx - \sum_{0 \leq i \leq n} (x_i - x_{i-1}) f(x_i) \right| \leq (b-a) \sup_i \text{osc}(f; [x_i; x_{i+1}]),$$

où l'oscillation de f sur un intervalle I est définie par :

$$\text{osc}(f; I) = \sup f(I) - \inf f(I). \quad \square$$

En particulier, lorsque f admet une dérivée de module borné par M , $\text{osc}(f; I) \leq |I| \cdot M$. Ceci montre que pour les deux sommes I_1, I_2 , l'erreur résultant du remplacement des sommes par les intégrales vaut $O\left(\frac{\ln^* n}{\sqrt{n}}\right)$.

La somme I_3 requiert un traitement spécial puisque F est continue mais non dérivable.

LEMME. L'oscillation de F sur un intervalle d'amplitude h est en $O(h \ln \frac{1}{h})$.

Preuve : On part de la définition de F :

$$F(u) = \frac{1}{2} c(u) + 2^{c(u)} h(2^{-c(u)}).$$

Compte tenu de la dérivabilité des fonctions $\frac{1}{2} c(u)$, $2^{c(u)}$ et $2^{-c(u)}$ sur l'intervalle $[0; 1[$, on s'assure que l'oscillation de ces trois fonctions sur un intervalle de longueur h est en $O(h)$. On se trouve ainsi ramené à déterminer l'oscillation de $h(x)$ sur un intervalle d'amplitude $O(h)$. On démontre que cette oscillation est en $O(h \ln \frac{1}{h})$ en partageant la somme donnant $h(x)$:

$$h(x) = \sum_{0 \leq j < \log \frac{1}{h^2}} 2^{-j} g(2^j x) + \sum_{j \geq \log \frac{1}{h^2}} 2^{-j} g(2^j x) ;$$

la seconde fonction est en $O(h^2)$ donc son oscillation est aussi en $O(h^2)$; la première fonction est la somme de $2 \log \frac{1}{h}$ fonctions dont l'oscillation est $O(h)$ et a donc une oscillation en $O(h \log \frac{1}{h})$. $\square$

Ces résultats montrent :

$$\begin{aligned} I_1 &= \int_0^{\ln n} \frac{t}{4} H_4(t) e^{-t^2} dt + O\left(\frac{\ln^* n}{\sqrt{n}}\right) \\ I_2 &= \int_0^{\ln n} \frac{t}{2} \log t H_4(t) e^{-t^2} dt + O\left(\frac{\ln^* n}{\sqrt{n}}\right) \\ I_3 &= \int_0^{\ln n} t F(\log t + \log_4 n) H_4(t) e^{-t^2} dt + O\left(\frac{\ln^* n}{\sqrt{n}}\right). \end{aligned}$$

Etant donné la décroissance rapide de e^{-t^2} dans les intégrandes, on peut remplacer pour chacune de ces trois intégrales les limites d'intégration par $[0; +\infty]$ en n'introduisant que des termes d'erreur exponentiellement petits. De sorte qu'en posant

$$\begin{aligned} J_1 &= \int_0^{\infty} \frac{t}{4} H_4(t) e^{-t^2} dt \\ J_2 &= \int_0^{\infty} \frac{t}{2} \log t H_4(t) e^{-t^2} dt \\ J_3 &= \int_0^{\infty} t F(\log t + \log_4 n) H_4(t) e^{-t^2} dt, \end{aligned}$$

on a :

$$I_1 = J_1 + O\left(\frac{\ln^* n}{\sqrt{n}}\right) ; I_2 = J_2 + O\left(\frac{\ln^* n}{\sqrt{n}}\right) ; I_3 = J_3 + O\left(\frac{\ln^* n}{\sqrt{n}}\right).$$

Ainsi l'expression de $\bar{R}_n = \frac{T_n}{B_n}$ devient

$$\bar{R}_n = -\log n \cdot J_1 + 1 - J_2 - J_3 + O\left(\frac{\ln^* n}{\sqrt{n}}\right).$$

Un calcul élémentaire montre que $J_1 = -\frac{1}{2}$ et $J_2 = \frac{\gamma}{2 \ln 2}$ [†]. Quant à J_3 , si l'on définit $J_3(u) = \int_0^\infty t F(\log t + u) e^{-t^2} H_4(t) dt$, on voit que la périodicité de F entraîne celle de $J_3(u)$. D'où

THEOREME 4^A. Le nombre moyen de registres $\bar{R}_n$ vérifie

$$\bar{R}_n = \log_4 n + D(\log_4 n) + O(1),$$

où $D(u) = 1 - \frac{\gamma}{2 \ln 2} - \int_0^\infty t H_4(t) e^{-t^2} F(\log t + u) dt$ est une fonction continue périodique de période 1. □

La fonction D est indéfiniment dérivable. La périodicité de la distribution lorsqu'on passe de n à $4n$ se retrouve ainsi de manière explicite dans la valeur moyenne du nombre de registres.

3.3. Le terme périodique du nombre moyen de registres.

Les résultats de la section précédente montrent la présence d'un terme périodique dans l'expression de $\bar{R}_n$. Ce terme périodique y est représenté comme une intégrale de convolution. Nous montrons ici qu'il est possible d'obtenir une expression explicite des coefficients de Fourier de ce terme périodique^{††}. La dérivation part du calcul par Delange des coefficients de Fourier de la fonction F .

La fonction F est une somme de termes dont chacun est exprimé au moyen de la fonction triangulaire périodique $g(x)$, et possède donc un développement de Fourier.

LEMME ([Delange, 1975]). La fonction F a un développement de Fourier

$F(u) = \sum f_k e^{-i2k\pi u}$ dont les coefficients sont donnés par :

$$\begin{cases} f_0 = \frac{1}{2} \log \pi - \frac{\gamma}{2 \ln 2} - \frac{1}{4} \\ f_k = -\frac{1}{\ln 2} x_k^{-1} (1+x_k)^{-1} \zeta(x_k) \quad \text{où } x_k = \frac{2ik\pi}{\ln 2} \end{cases} \quad \square$$

[†] Le symbole γ désigne ici la constante d'Euler $\gamma = 0.57721\dots$

^{††} L'auteur est redevable à H. Delange de lui avoir indiqué la possibilité de ce développement.

Nous en déduisons

THEOREME 4^B. Le nombre moyen de registres relatif à l'ensemble des arbres de taille n a comme expression :

$$\bar{R}_n = \log_4 n + D(\log_4 n) + o(1),$$

où D est développable en série de Fourier : $D(u) = \sum d_k e^{2ik\pi u}$

les coefficients d_k étant donnés par :

$$d_0 = \left(\frac{1}{2} - \frac{\gamma+2}{2 \ln 2} + \log \pi\right) = 0.292...$$

$$d_k = \frac{1}{\ln 2} (x_k - 1) \Gamma\left(\frac{x_k}{2}\right) \zeta(x_k) \quad \text{avec} \quad x_k = \frac{2ik\pi}{\ln 2}.$$

Preuve : On part de l'expression de $D(u)$:

$$D(u) = 1 - \frac{\gamma}{2 \ln 2} - \int_0^\infty t H_4(t) e^{-t^2} F(u + \log t) dt.$$

L'intégrale est une variante d'un produit de convolution. Les coefficients sont déterminés au moyen des intégrales

$$a_k = \int_0^1 e^{-2ik\pi u} \int_0^\infty F(\log t + u) t H_4(t) e^{-t^2} dt du,$$

par $d_k = a_k$ si $k \neq 0$ et $d_0 = a_0 + 1 - \frac{\gamma}{2 \ln 2}$.

L'intégrale double correspondant à a_k existe, ce qui permet d'échanger l'ordre de sommation :

$$a_k = \int_0^\infty t H_4(t) e^{-t^2} \int_0^1 e^{-2ik\pi u} F(\log t + u) du dt.$$

On pose $v = \log t + u$ et grâce à la périodicité de F , on obtient :

$$a_k = \int_0^\infty t H_4(t) e^{-t^2} e^{2ik\pi \log t} \int_0^1 e^{-2ik\pi v} F(v) dv dt.$$

Le calcul de a_k se décompose donc en celui des quantités

$$f_k = \int_0^1 e^{-2ik\pi v} F(v) dv : \text{donné par le lemme de Delange ;}$$

et $c_k = \int_0^\infty t H_4(t) e^{-t^2} e^{2ik\pi \log t} dt$: qui se ramène à des intégrales Eulériennes.

Posant $u = t^2$ dans l'expression de c_k :

$$c_k = \frac{1}{2} \int_0^\infty (16u^2 - 48u + 12) e^{-u} u^{ik\pi/\ln 2} du = 8\Gamma\left(3 + \frac{ik\pi}{\ln 2}\right) - 24\Gamma\left(2 + \frac{ik\pi}{\ln 2}\right) + 6\Gamma\left(1 + \frac{ik\pi}{\ln 2}\right).$$

Pour $k=0$, c_k vaut -2. Pour $k \neq 0$, la propriété fonctionnelle de Γ :

$\Gamma(z+1) = z\Gamma(z)$, conduit après simplification à l'expression

$$c_k = x_k (x_k - 1) \Gamma\left(\frac{x_k}{2}\right) \quad \text{où} \quad x_k = \frac{2ik\pi}{\ln 2}.$$

Combinant cette expression avec l'expression de f_k fournie par le lemme de Delange, on en déduit le développement de l'énoncé. Enfin la série de Fourier de D est absolument (et rapidement) convergente puisque $z(it) = O(|t|^{\frac{1}{2}+\epsilon})$ et $\Gamma(it) = O(e^{-\frac{\pi}{2}|t|})$ lorsque $t \rightarrow \infty$ (cf. [Whittaker, Watson ; 1902]). $\square$

Il reste à effectuer l'analyse numérique des coefficients de Fourier de la fonction $D(u)$. La décroissance exponentielle de la fonction Γ le long de l'axe imaginaire se déduit de la formule des compléments

$$\Gamma(z)\Gamma(-z) = \frac{-\pi}{z \sin \pi z},$$

d'où, comme $\Gamma(\bar{z}) = \overline{\Gamma(z)}$, la relation

$$|\Gamma(it)|^2 = \frac{\pi}{t \operatorname{sh} \pi t} \cdot \frac{1}{2^{1+\epsilon}}$$

Cette décroissance entraîne puisque $|z(it)| = O(|t|^{\frac{1}{2}+\epsilon})$ que numériquement la fonction D est approchée de manière très étroite par les premiers termes de son développement de Fourier : ainsi les trois premiers termes donnent une précision meilleure que 10^{-10} .

Le calcul de Γ s'effectue (cf. [Edwards, 1974]) par application de la formule d'Euler-Mac Laurin à la représentation sous forme de somme de $\ln \Gamma(s)$. On obtient ainsi la série de Stirling :

$$\ln \Gamma(s) = (s - \frac{1}{2}) \ln s - s + \frac{1}{2} \ln 2\pi + \frac{B_2}{1.2s} + \frac{B_4}{3.4s^3} + \dots + \frac{B_{2m}}{(2m-1)2m s^{2m-1}} + R_{2m}$$

dans laquelle $B_2, B_4, \dots$ sont les nombres de Bernoulli, et où l'on montre que R_{2m} est borné par :

$$R_{2m} \leq \left(\frac{1}{\cos \theta/2} \right)^{2m+2} \left| \frac{B_{2m+2}}{(2m+1)(2m+2)s^{2m+1}} \right| \text{ avec } \theta = \arg s.$$

L'application de ce calcul aux valeurs $\frac{x_k}{2} = \left(\frac{ik\pi}{\ln 2} \right)_{k=1,2,3}$ montre qu'en prenant seulement 5 termes de la série de Stirling l'erreur est inférieure à 10^{-10} .

On calcule ainsi

$$\Gamma\left(\frac{x_1}{2}\right) \cong 5.47320334 \cdot 10^{-5} + i9.51234347 \cdot 10^{-4}$$

$$\Gamma\left(\frac{x_2}{2}\right) \cong -4.17675105 \cdot 10^{-7} - i3.50438021 \cdot 10^{-7}$$

$$\Gamma\left(\frac{x_3}{2}\right) \cong -2.25805054 \cdot 10^{-10} + i2.80694697 \cdot 10^{-10}.$$

De même, l'application de la formule d'Euler-Maclaurin au reste de la série donnant

$\zeta(s)$ (cf. également [Edwards, 1974]) montre :

$$\begin{aligned}\zeta(s) &= \sum_{1 \leq k < N} \frac{1}{k^s} + \sum_{N < k} \frac{1}{k^s} \\ \zeta(s) &= \sum_{1 \leq k < N} \frac{1}{k^s} + \frac{N^{1-s}}{s-1} + \frac{1}{2} N^{-s} + \frac{B_2}{2!} s N^{-s-1} + \frac{B_4}{4!} s(s+1)(s+2) N^{-s-3} + \dots \\ &\quad + \frac{B_{2m}}{2m!} s(s+1) \dots (s+2m-2) N^{-s-2m+1} + R_{2m}\end{aligned}$$

où $R_{2m} \leq \left| \frac{s+2m+1}{s+2m+1} \right| \cdot \left| \frac{B_{2m+2}}{2m+2!} s(s+1) \dots (s+2m) N^{-s-2m-1} \right|$ avec $\sigma = \text{Re}(s)$.

La formule d'Euler-Maclaurin initialement applicable lorsque $\text{Re}(s) > 1$ reste valide par prolongement analytique pour $\text{Re}(s) > -2m$ (m étant le nombre de termes du développement). Ainsi avec $N=40$ et $m=5$, obtient-on

$\{\zeta(\chi_k)\}_{k=1,2,3}$ avec une précision meilleure que 10^{-10} . De la sorte, on calcule :

$$\begin{aligned}\zeta(\chi_1) &\approx 1.59698524 + i0.27944039 \\ \zeta(\chi_2) &\approx 3.07592654 - i0.53845269 \\ \zeta(\chi_3) &\approx 2.39137398 - i3.27326909.\end{aligned}$$

Soit finalement pour c_1, c_2, c_3 les valeurs :

$$\begin{aligned}c_1 &\approx -1.98089443 \cdot 10^{-2} - i4.54681440 \cdot 10^{-3} \\ c_2 &\approx 2.44367551 \cdot 10^{-5} - i3.73073420 \cdot 10^{-5} \\ c_3 &\approx -2.24451566 \cdot 10^{-8} - i5.31770423 \cdot 10^{-8}.\end{aligned}$$

Ainsi l'amplitude de $D(u)$ est majorée par $4.1 \cdot 10^{-2}$. Le graphe de $\bar{R}_n - \log_4 n$ est donné par la figure 10, et montre l'accord très rapide de $\bar{R}_n$ et des termes $\log_4 n + D(\log_4 n)$.

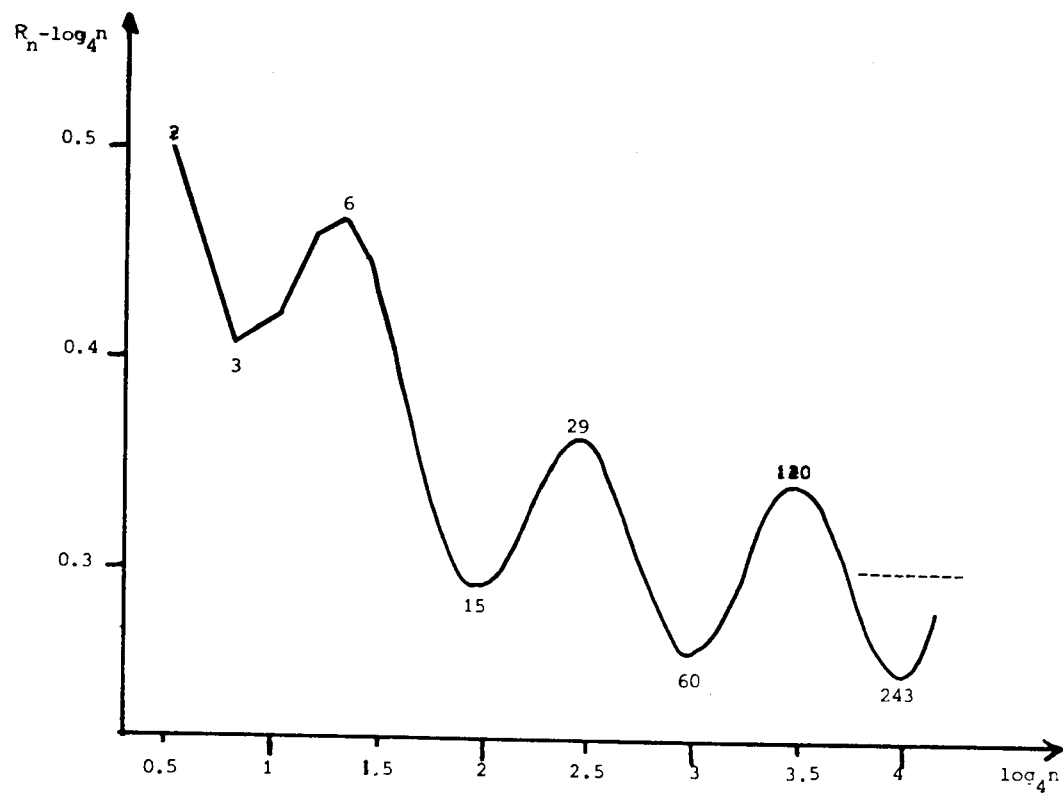


Figure 10 : Valeurs de $\bar{R}_n - \log_4 n$.

4. ORDRE DES SOMMETS D'UN ARBRE BINAIRE.

Les applications statistiques de la fonction nombre de registres, ou nombre de Strahler, sont relatives à l'ensemble des sous-arbres d'un arbre. Etant donné un arbre T et un sommet $s \in T$, on appelle ordre de s dans T noté $\omega(s; T)$ le nombre de registres associé au sous-arbre de T de racine s . Nous calculons ici l'ordre moyen d'un sommet d'un arbre de taille n .

On définit les quantités :

$$\Omega_{p,n} = \text{card}\{s \in T \mid |T| = n \text{ et } \omega(s; T) = p\}.$$

$$\Omega_p(z) = \sum_{n \geq 0} \Omega_{p,n} z^n$$

$$U_n = \sum_{p \geq 0} p \Omega_{p,n};$$

enfin la statistique est effectuée sur l'ensemble des sommets internes des arbres de taille n , qui sont en nombre $nB_n = \binom{2n}{n-1}$, d'où la valeur moyenne

$$\bar{\Omega}_n = \frac{1}{\binom{2n}{n-1}} U_n = \frac{1}{\binom{2n}{n-1}} \sum_{p \geq 0} p \Omega_{p,n}.$$

D'après les méthodes développées au chapitre I, la série génératrice

$\Omega_p(z)$ vaut :

$$\Omega_p(z) = \frac{1}{\sqrt{1-4z}} R_p(z),$$

ce qui, compte tenu de l'expression de $R_p(z)$ donnée à la section 2, montre que

$$\Omega_p(z) = z^{2^{p-1}} \frac{B^{2^{p+1}}}{1 - z^{2^{p+1}} B^{2^{p+2}}},$$

d'où pour $\Omega_{p,n}$ des formes analogues à celles de $R_{p,n}$

$$\Omega_{p,n} = \sum_{k \geq 0} [\binom{2n+1}{n+1-2^p(2k+1)} - \binom{2n+1}{n-2^p(2k+1)}].$$

Il en va de même pour la série génératrice $U(z) = \sum U_n z^n$:

$$U(z) = \frac{1}{\sqrt{1-4z}} T(z) = \sum_{k \geq 1} v_2(k) z^{k-1} B^{2^k},$$

d'où la formule

$$U_n = \sum_{k \geq 1} [\binom{2m+1}{m+1-k} - \binom{2m+1}{m-k}] v_2(k) = \frac{1}{m+1} \sum_{k \geq 1} \binom{2m+2}{m+1-k} k v_2(k).$$

En particulier les premières valeurs de U_n sont

$$U_1 = 1 \quad U_2 = 4 \quad U_3 = 16 \quad U_4 = 64 \quad U_5 = 256, \text{ soit} \\ \bar{\Omega}_1 = 1 \quad \bar{\Omega}_2 = 1 \quad \bar{\Omega}_3 = 1.066... \quad \bar{\Omega}_4 = 1.1428... \quad \bar{\Omega}_5 = 1.2095... .$$

Cette somme se calcule par les méthodes de la section 3 et l'on montre ainsi :

PROPOSITION. *L'ordre moyen d'un sommet dans un arbre de taille n a pour estimation*

$$\bar{\Omega}_n = 2 - O\left(\frac{\ln n}{\sqrt{n}}\right). \quad \square$$

Cette estimation est voisine de l'ordre moyen des sommets d'un arbre parfait, puisque pour un arbre de taille h , celui-ci vaut

$$\frac{2^{h-1}}{2^h - 1} \left(1 + \frac{2}{2} + \frac{3}{2^2} + \dots + \frac{h}{2^{h-1}}\right) \approx 2.$$

Au contraire pour le réseau de la Upper Hiwassee River étudié par Shreve (de taille 89) cette quantité vaut ~ 1.25 et pour le peuplier courant (de taille 339, cf. Mc Mahon) elle vaut ~ 1.33 .

Ce résultat est aussi à comparer à la hauteur moyenne du plus grand arbre parfait qui occure en un sommet d'un arbre de taille n , laquelle tend vers la constante transcendante

$$\alpha = \sum_{h \geq 1} 2^{-(2^h - 2)} = 1.282352...,$$

ce qui s'établit facilement par le compte d'occurences de motifs du chapitre I (cf. également [de Bruijn, 1975]). Comme il est prévisible $\alpha \leq 2 = \lim_{n \rightarrow \infty} \bar{\Omega}_n$.

CHAPITRE III

SYSTEMES DE NUMERATION ET PHENOMENES DE PERIODICITE EN ANALYSE D'ALGORITHMES.

Différents systèmes de numération apparaissent dans les problèmes d'analyse d'algorithmes. Nous avons vu au chapitre précédent, le rôle que joue le système binaire vis-à-vis de l'allocation de registres. Certaines structures de données, comme les files binomiales de [Vuillemin ; 1978] et la généralisation que constituent les structures dynamiques de [Bentley ; 1978] sont même construites explicitement sur la représentation binaire des entiers. Les périodicités inhérentes aux distributions de chiffres dans ces systèmes de numération conduisent à l'apparition de termes périodiques dans les expressions de coûts moyens d'algorithmes.

Dans ce chapitre, nous étendons d'abord le résultat de Delange sur la représentation binaire, à d'autres systèmes de numération et nous traitons en détail le code binaire réfléchi ou code Gray (section 1). Nous obtenons pour le nombre total de 1 dans la représentation en code Gray des entiers $1, 2, \dots, m-1$ l'expression

$$T(m) = \frac{1}{2} m \log m + m G(\log m)$$

où G est continue, périodique, nulle part dérivable ; cette fonction est construite à partir d'une superposition de fonctions triangulaires comme illustré par la figure 1a ; la figure 1b montre par contraste la fonction homologue relative au système binaire classique.

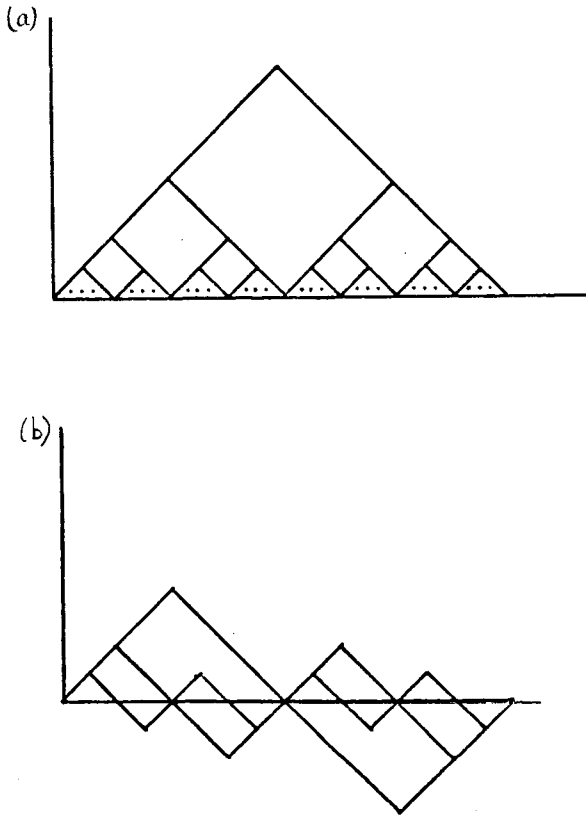


Figure 1 : Fonctions triangulaires associées au code binaire (1a) et au code Gray (1b).

Nous appliquons ensuite (section 2) ces résultats à l'analyse du comportement des réseaux de tri par fusion pair-impair, qui sont dûs à [Batcher ; 1968] : le code Gray joue vis-à-vis de l'analyse un rôle très semblable à celui de la représentation binaire pour l'allocation de registres. Faisant apparaître géométriquement la représentation Gray des entiers dans l'analyse, nous en déduisons une preuve élémentaire par l'analyse réelle du résultat de [Sedgewick ; 1978] concernant le nombre moyen d'échanges dans ces réseaux, lequel vaut

$$\bar{E}_n = \frac{1}{4} n \log n + n M(\log_4 n) + o(n)$$

où M est continue, périodique et indéfiniment dérivable.

Dans une dernière partie (section 3), nous discutons une alternative de traitement de ces problèmes par transformation intégrale dans le plan complexe fondée sur la transformation de Mellin-Fourier. Cette méthode initialement employée en analyse d'algorithme par [de Bruijn, Knuth, Rice ; 1972] dans l'évaluation de la hauteur des arbres est à la base de la preuve originelle du résultat de Sedgewick ; elle a également été appliquée par Knuth à divers problèmes algorithmiques et par Kemp à l'étude du nombre de registres. Nous l'utilisons ici pour traiter des analyses non immédiatement accessibles à l'analyse réelle :

- le calcul de la variance du nombre de registres sur l'ensemble des arbres de taille n , pour lequel nous obtenons l'expression

$$\sigma_n^2 = V(\log_4 n) + o(1),$$

où V est continue et périodique ;

- la détermination de la taille moyenne du plus grand arbre parfait plongé dans un arbre binaire pour lequel nous obtenons l'expression

$$\bar{P}_n = \sqrt{n} W(\log_4 n) + O(1),$$

où W est continue et périodique.

Ces deux résultats montrent, comme dans le cas des arbres 2-3 équilibrés traité par [Odlyzko ; 1979] l'apparition de fonctions périodiques comme termes principaux de développements asymptotiques.[†]

Les résultats des sections 1 et 2 que recourent des résultats indépendants de Ramshaw font l'objet d'une publication commune [Flajolet, Ramshaw ; 1979]. Certains développements géométriques et les résultats de la section 3 sont nouveaux.

+ On pourra également consulter [Knuth ; 1973] pour d'autres exemples de phénomènes de périodicité.

Soit $\gamma_k(n)$ le k -ième chiffre de la représentation Gray de n , et soit $\gamma(n) = \sum_k \gamma_k(n)$ le nombre total de 1 dans la représentation de l'entier n . Avec les notations du chapitre précédent, on a $\gamma(n) = v(g(n))$.

Le problème que nous considérons consiste à calculer le nombre total de 1 contenus dans les n premières colonnes de la table donnée par la figure 2, c'est-à-dire à évaluer la somme :

$$T(n) = \sum_{0 \leq j < n} \gamma(j).$$

Nous montrons le résultat suivant :

THEOREME 1. Soit $\gamma(n)$ le nombre de chiffres 1 dans la représentation en code Gray de n . Il existe une fonction $G: \mathbb{R} \rightarrow \mathbb{R}$, continue, périodique et de période 1 telle que :

$$\sum_{0 \leq j < n} \gamma(j) = \frac{1}{2} n \log n + n G(\log n).^{\dagger}$$

La série de Fourier de $G: \sum g_k e^{2ik\pi x}$ converge absolument et ses coefficients g_k sont donnés par

$$g_0 = 2 \log \Gamma\left(\frac{1}{4}\right) - \log \pi - \frac{1}{2\ln 2} - \frac{5}{4} = 0.093604 \dots$$

$$g_k = \frac{2}{\ln 2} x_k^{-1} (1 + x_k)^{-1} \zeta(x_k; \frac{1}{4}) \quad \text{où } x_k = \frac{2k\pi i}{\ln 2} \text{ pour } k \neq 0.$$

Preuve : La preuve est voisine de celle de Delange. Considérant les lignes successives du tableau des représentations d'entier, on observe que la k -ième ligne consiste en la répétition infinie de blocs de bits du type $0^{2^k} 1^{2^k} 0^{2^k} 1^{2^k}$. Ainsi chaque ligne comporte asymptotiquement autant de 0 que de 1, et cet argument de comptage simple complété par une estimation grossière de l'erreur conduit à l'équivalence :

$$T(n) \sim \frac{1}{2} n \log n.$$

On affine cet argument en comparant le nombre exact de 1 dans la k -ième ligne, soit $\sum_{0 \leq k < n} \gamma_k(n)$ à l'estimation par $\frac{1}{2} \log n$; soit $e_k(n)$ l'écart entre ces deux quantités :

$$e_k(n) = \left(\sum_{0 \leq k < n} \gamma_k(n) \right) - \frac{n}{2}.$$

[†] Nous continuons de suivre la convention habituelle en informatique que $\log u = \log_2 u$.

La valeur de $e_k(n)$ lorsque n est multiple de 2^k se détermine exactement, car pour $n=p2^k$ la k -ième ligne consiste en un nombre entier de blocs de taille 2^k . Ainsi

$$e_k(p2^k) = \begin{cases} 0 & \text{si } p \equiv 0 \pmod{4} \\ -2^{k-1} & \text{si } p \equiv 1 \pmod{4} \\ 0 & \text{si } p \equiv 2 \pmod{4} \\ 2^{k-1} & \text{si } p \equiv 3 \pmod{4}. \end{cases}$$

Pour p entre ces valeurs particulières, $e_k(n)$ est une fonction linéaire. On peut donc, à un changement d'échelle près, exprimer e_k au moyen d'une fonction triangulaire $t : \mathbb{R} \rightarrow \mathbb{R}$ définie par :

$$t(x) = \begin{cases} -\frac{x}{2} & \text{si } 0 \leq x < \frac{1}{4} \\ x - \frac{1}{4} & \text{si } \frac{1}{4} \leq x < \frac{3}{4} \\ \frac{1}{2} - \frac{x}{2} & \text{si } \frac{3}{4} \leq x < 1 \end{cases} \quad \text{et } t(x+1) = t(x) \text{ pour tout } x.$$

La valeur des premières fonctions d'écart et le graphe de la fonction $t(x)$ sont représentés aux figures 3 et 4.

0	$-\frac{1}{2}$	-1	$-\frac{3}{2}$	-2	$-\frac{5}{2}$	-3	$-\frac{7}{2}$	-4	$-\frac{9}{2}$	-5	$-\frac{11}{2}$	-6	$-\frac{13}{2}$	-7	$-\frac{15}{2}$	-8	$-\frac{15}{2}$	-7
0	$-\frac{1}{2}$	-1	$-\frac{3}{2}$	-2	$-\frac{5}{2}$	-3	$-\frac{7}{2}$	-4	$-\frac{7}{2}$	-3	$-\frac{5}{2}$	-2	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1
0	$-\frac{1}{2}$	-1	$-\frac{3}{2}$	-2	$-\frac{3}{2}$	-2	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1
0	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1
0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 : n

Figure 3 : Valeur des fonctions d'écart $e_k(n)$ pour $k=0,1,2,3$ et $n=0\dots 18$.

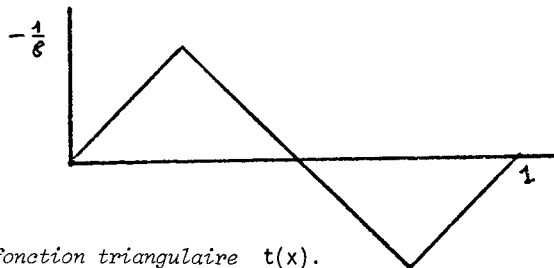


Figure 4 : La fonction triangulaire $t(x)$.

On a ainsi

$$e_k(n) = 2^{k+2} t(n/2^{k+2}).$$

Soit $\ell = \lfloor \log n \rfloor$. Comme toutes les lignes du tableau des représentations Gray dont l'index est plus grand que ℓ commencent par un zéro au moins, on a :

$$T(n) = \sum_{0 \leq i < n} \gamma(i) = \sum_{0 \leq k < \ell} \left(\frac{n}{2} + e_k(n) \right) = \frac{1}{2}(\ell+1)n + \sum_{0 \leq k \leq \ell} 2^{k+2} t(n/2^{k+2}).$$

Comme $t(x)$ est nul pour x entier, la borne supérieure sur l'indice de sommation peut être remplacée par $-\infty$:

$$\sum_{0 \leq k \leq \ell} 2^{k+2} t(n/2^{k+2}) = \sum_{-\infty < k \leq \ell} 2^{k+2} t(n/2^{k+2}) = \sum_{j \geq 0} 2^{\ell-j+2} t(n/2^{\ell-j+2}),$$

la dernière ligne étant obtenue par le changement d'indice $j = \ell - k$. Cette expression suggère de considérer la fonction

$$h(x) = \sum_{j \geq 0} 2^{-j} t(2^j x)$$

laquelle, à cause de la convergence uniforme de la somme est une fonction continue.

Cette fonction est une superposition de fonctions triangulaires comme l'indique

la figure 5.

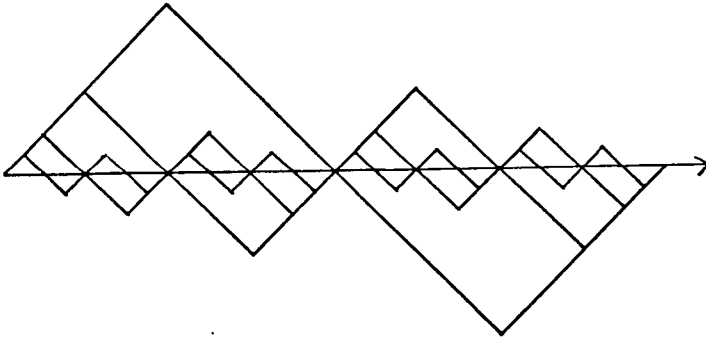


Figure 5 : La superposition de fonctions triangulaires donnant la fonction h .

Cette fonction est voisine des exemples standards dus à Bolzano et Weierstrass d'une fonction continue nulle part dérivable. Utilisant cette fonction h , l'expression

de $T(n)$ se met sous la forme

$$T(n) = \frac{1}{2}(\ell+1)n + 2^{\ell+2}h(n/2^{\ell+2}).$$

Puisque $\ell = \lfloor \log n \rfloor$, on peut faire apparaître la partie fractionnaire de $\log n$:

$$T(n) = \frac{1}{2}n \log n + \frac{1}{2}n(\lfloor \log n \rfloor - \log n + 1) + n2^{\lfloor \log n \rfloor - \log n + 2}h(2^{\lfloor \log n \rfloor - \log n - 2}).$$

En utilisant la notation $\{x\} = x - \lfloor x \rfloor$, on définit la fonction $G : \mathbb{R} \rightarrow \mathbb{R}$ par l'égalité :

$$G(x) = \frac{1}{2}(1 - \{x\}) + 2^{2-\{x\}}h(2^{\{x\}-2}),$$

et l'on a la formule

$$T(n) = \frac{1}{2}n \log n + nG(\log n).$$

La fonction $G(x)$ est périodique car seulement fonction de $\{x\}$; d'après la continuité de h , elle est continue sur $[0;1[$. On calcule directement

$$G(0) = \frac{1}{2} + 4h\left(\frac{1}{4}\right) = \frac{1}{2} + 4t\left(\frac{1}{4}\right) = \frac{1}{2} - \frac{1}{2} = 0 ;$$

et comme $G(x)$ tend vers 0 lorsque x tend vers 1, on en déduit la continuité de $G(x)$ en $x=1$, donc finalement en tout x réel.

Il reste à déterminer la série de Fourier de G , c'est-à-dire exprimer G sous la forme

$$G(x) = \sum_{k \in \mathbb{Z}} g_k e^{2ik\pi x},$$

avec

$$g_k = \int_0^1 G(u) e^{-2ik\pi u} du.$$

La fonction G est somme de deux fonctions :

$$G(x) = \frac{1}{2}(1 - \{x\}) + 2^{2-\{x\}}h(2^{\{x\}-2}) ;$$

la première fonction triangulaire a une série de Fourier qui se détermine directement. La seconde est une superposition de fonctions triangulaires, et ses coefficients de Fourier s'expriment sous forme de série.

On a donc $g_k = c_k + d_k$, avec

$$c_k = \int_0^1 \left(\frac{1-u}{2}\right) e^{-2ik\pi u} du$$

$$d_k = \int_0^1 2^{2-u} h(2^{u-2}) e^{-2ik\pi u} du.$$

On trouve immédiatement les valeurs $c_0 = \frac{1}{4}$ et $c_k = \frac{1}{4ik\pi}$ lorsque $k \neq 0$.

La convergence uniforme de la somme donnant $h(x)$ permet d'échanger somme et

intégrale, soit

$$d_k = \sum_{j \geq 0} \int_0^1 2^{2-u-j} t(2^{u+j-2}) e^{-2ik\pi u} du.$$

Le changement de variable $u = 2-j + \log v$ conduit à l'expression :

$$d_k = \frac{1}{\ln 2} \sum_{j \geq 0} \int_{2^{j-2}}^{2^{j-1}} \frac{t(v) dv}{v^{2+2ik\pi/\ln 2}},$$

soit encore, en posant $x_k = \frac{2ik\pi}{\ln 2}$:

$$d_k = \frac{1}{\ln 2} \int_{\frac{1}{4}}^{\infty} \frac{t(v) dv}{v^{2+x_k}} = \frac{1}{\ln 2} H(1+x_k),$$

où $H(z)$ représente l'intégrale :

$$H(z) = \int_{\frac{1}{4}}^{\infty} \frac{t(v)}{v^{z+1}} dv.$$

Nous allons évaluer cette intégrale sous l'hypothèse de départ que $\operatorname{Re}(z) > 2$, ce qui rend faciles certaines manipulations algébriques. Le caractère analytique des expressions obtenues montre que l'évaluation reste valable pour $\operatorname{Re}(z) > 0$, domaine qui constitue le demi-plan d'absolue convergence de l'intégrale. Partant de la définition de la fonction triangulaire $t(x)$, on vérifie l'équivalence de la forme

$$t(v) = \int_0^v \left(\left\lfloor -x + \frac{3}{4} \right\rfloor - \left\lfloor x + \frac{1}{4} \right\rfloor - \frac{1}{2} \right) dx.$$

Utilisant cette écriture, l'intégration par partie de l'intégrale qui donne $H(z)$ conduit à l'expression :

$$H(z) = -\frac{4z-1}{2z} + \frac{1}{z} \int_{\frac{1}{4}}^{\infty} \left(\left\lfloor x + \frac{3}{4} \right\rfloor - \left\lfloor x + \frac{1}{4} \right\rfloor - \frac{1}{2} \right) \frac{dv}{v^z}.$$

L'hypothèse $\operatorname{Re}(z) > 2$ permet de séparer en trois parties cette intégrale. La troisième partie s'évalue directement :

$$\int_{\frac{1}{4}}^{\infty} \frac{dv}{v^z} = \frac{4^{z-1}}{z-1}.$$

Les deux premières parties font intervenir la fonction $\zeta(z, a)$ d'Hurwitz, qui généralise la fonction $\zeta(z)$ de Riemann :

$$\zeta(z, a) = \sum_{j \geq 0} \frac{1}{(j+a)^z} \quad \text{pour } \operatorname{Re}(z) > 1.$$

Lorsque $0 < a \leq 1$, on calcule comme suit les intégrales définies intervenant dans l'expression de H :

$$(z-1) \int_a^{J+a} [v+1-a] \frac{dv}{v^z} = \sum_{1 \leq j \leq J} j \int_{a+j-1}^{a+j} \frac{(z-1)dv}{v^z} = \sum_{1 \leq j \leq J} j \left(\frac{1}{(a+j-1)^{z-1}} - \frac{1}{(a+j)^{z-1}} \right) \\ = \left(\sum_{0 \leq j < J} \frac{1}{(j+a)^{z-1}} \right) - \frac{J}{(J+a)^{z-1}}.$$

Comme $\operatorname{Re}(z) > 2$, on peut faire tendre J vers l'infini, et obtenir ainsi l'expression :

$$(z-1) \int_a^{\infty} [v+1-a] \frac{dv}{v^z} = \zeta(z-1, a).$$

Revenant à $H(z)$, on en déduit :

$$H(z) = -\frac{4^{z-1}}{2(z-1)} + \frac{\zeta(z-1, \frac{1}{4}) - \zeta(z-1, \frac{3}{4})}{z(z-1)},$$

égalité qui reste valable par prolongement analytique pour $\operatorname{Re}(z) > 0$, $z \neq 1$.

Cette expression peut se transformer : pour tout entier $\ell \geq 2$, on a en effet l'égalité

$$\zeta(z, 1/\ell) + \zeta(z, 2/\ell) + \dots + \zeta(z, (\ell-1)/\ell) = \ell^z \zeta(z),$$

qui s'obtient en regroupant les termes dans la somme donnant ζ selon les classes résiduelles de leur index modulo ℓ .

Utilisant cette récurrence avec $\ell = 2$ et $\ell = 4$, on obtient

$$\zeta(z, \frac{3}{4}) = (4^z - 2^z) \zeta(z) - \zeta(z, \frac{1}{4});$$

soit finalement

$$H(z) = -\frac{4^{z-1}}{2(z-1)} + \frac{2\zeta(z-1, \frac{1}{4})}{z(z-1)} - \frac{2(4^{z-1} - 2^{z-1})\zeta(z-1)}{z(z-1)}.$$

Cette expression permet le calcul des quantités d_k pour $k \neq 0$; il suffit de substituer $(1+x_k)$ à z et d'observer les simplifications résultant de l'égalité $2^{x_k} = 1$:

$$d_k = \left(-\frac{1}{4k\pi i} + \frac{2\zeta(x_k, \frac{1}{4})}{(\ln 2)x_k(1+x_k)} \right) \text{ pour } k \neq 0.$$

Quant à d_0 , il s'obtient à partir de $H(1)$ dont la valeur est calculable par les développements :

$$4^{z-1} = 1 + 2 \ln 2 (z-1) + O((z-1)^2) \\ 2^{z-1} = 1 + \ln 2 (z-1) + O((z-1)^2) \\ \frac{1}{z} = 1 - (z-1) + O((z-1)^2) \\ \zeta(z-1, \frac{1}{4}) = \frac{1}{4} + (\ln \Gamma(\frac{1}{4}) - \frac{\ln(2\pi)}{2})(z-1) + O((z-1)^2),$$

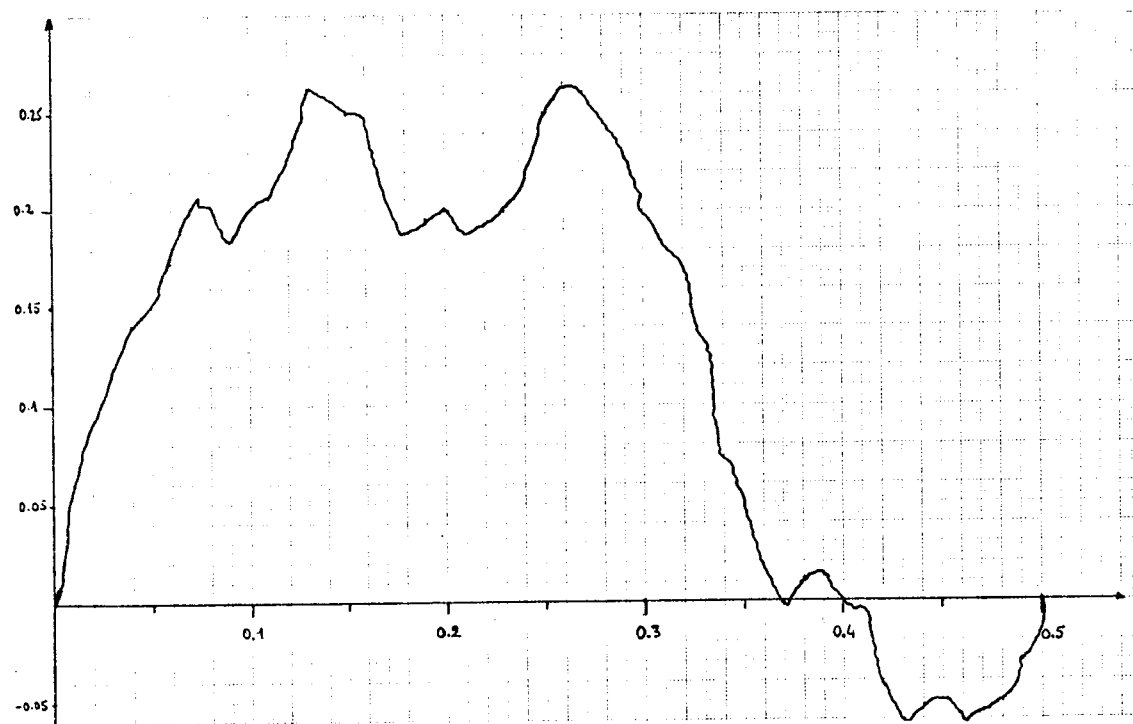


Figure 6 : le graphe de la fonction $-2h(x)$ relative au code Gray.

ce dernier résultat étant classique (cf. [Whittaker, Watson ; 1927]).

De ces développements, on déduit

$$H(z) = 2 \ln \Gamma\left(\frac{1}{4}\right) - \frac{3 \ln 2}{2} - \ln \pi - \frac{1}{2} + O(z-1)$$

ce qui implique

$$d_0 = 2 \log \Gamma\left(\frac{1}{4}\right) - \log \pi - \frac{1}{2 \ln 2} - \frac{3}{2}.$$

Ceci permet finalement le calcul des coefficients de Fourier de G puisque

$$g_k = c_k + d_k :$$

$$g_0 = 2 \log \Gamma\left(\frac{1}{4}\right) - \log \pi - \frac{1}{2 \ln 2} - \frac{5}{4}$$

$$g_k = \frac{2\zeta(x_k, \frac{1}{4})}{(\ln 2)^{x_k} (1+x_k)} \text{ pour } x_k = \frac{2k\pi i}{\ln 2}, k \neq 0.$$

Enfin, comme $\zeta(it, \frac{1}{4}) = O(|t|^{\frac{1}{2}+\epsilon})$, la série de Fourier de G est absolument convergente, ce qui termine la preuve du théorème. $\square$

A titre d'illustration, nous avons représenté à la figure 6, le tracé approximatif du graphe de la fonction $-2h(x)$ sur l'intervalle $[0; \frac{1}{2}]$.

Cette méthode d'estimation de distribution de chiffres s'étend à d'autres systèmes de numération (cf. [Flajolet, Ramshaw ; 1979]), la condition sur ces systèmes étant simplement que la même régularité d'apparition des chiffres se reproduise en chaque position. C'est par exemple le cas du système ternaire équilibré dans lequel les chiffres sont $1, 0, -1$ (encore noté $\bar{1}$) et la base est 3. On sait que chaque entier positif ou négatif possède une représentation unique sous cette forme (cf. [Knuth ; 1969]) ; par exemple

$$20 = 3^3 - 3^2 + 3 - 1 \text{ est noté dans ce système } 1\bar{1}11.$$

Du fait du rôle symétrique des chiffres 1 et $\bar{1}$, on démontre par exemple que la somme des chiffres des représentations des entiers $0, 1, \dots, m-1$ soit $U(m)$ a une expression dans laquelle n'apparaît pas de termes en $O(m \log m)$, et qui peut se

mettre sous la forme

$$U(m) = m K(\log_3 m)$$

où K est continue, périodique et de période 1.

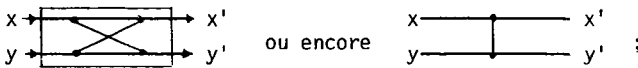
Des résultats du même type valent pour les systèmes $\langle q, r \rangle$ dans lesquels la base est q , et les chiffres sont pris contigûment dans l'intervalle $r, r+1, \dots, r+q-1$. Ils font encore intervenir dans les développements de Fourier, les valeurs de la fonction ζ d'Hurwitz pour des valeurs rationnelles du paramètre. Ceci permet par exemple de traiter le système binaire négatif dont la base est -2 (cf. l'article déjà cité, écrit en collaboration avec L. Ramshaw).

2. ANALYSE DU TRI PAR FUSION PAIR-IMPAIR.

Nous décrivons ici brièvement la méthode du tri par fusion pair-impair, due à [Batcher ; 1968]. Partant de la présentation de l'algorithme donnée par [Knuth ; 1972] et [Sedgewick ; 1978], nous faisons apparaître géométriquement la représentation Gray des entiers comme l'un des paramètres essentiel de l'algorithme. Utilisant les résultats de la section précédente nous en déduisons une preuve élémentaire de l'analyse primitivement effectuée par Sedgewick. La démarche suivie au cours de l'analyse asymptotique est voisine du traitement que nous avons donné de l'évaluation de l'allocation de registres au chapitre précédent.

2.1. Réseaux de tri et fusion pair-impair.

Un réseau de tri de n éléments est formé de n lignes d'entrées et n lignes de sorties ; la structure interne est une interconnection de modules de comparaison-échange, qu'on peut représenter graphiquement sous la forme



l'entrée d'un module s'effectue à gauche, la sortie à droite et la spécification de la fonction calculée est

$$x' = \min(x, y) \quad y' = \max(x, y).$$

D'une manière équivalente, un réseau de tri de n éléments équivaut à un programme séquentiel dont les variables communes à l'entrée et à la sortie sont $x_1, x_2, \dots, x_n$ et dont l'unique instruction est une procédure de comparaison-échange.

$X(x; y)$ définie par : si $x > y$ alors $x \leftrightarrow y$ fsi.[†]

La figure 7 représente deux réseaux de tri de 4 éléments, ainsi que les programmes associés.

[†]

L'instruction $x \leftrightarrow y$ signifie l'échange du contenu de x et y , c.à.d. la suite d'instructions $t \leftarrow x ; x \leftarrow y ; y \leftarrow t$.

Les réseaux de tri se distinguent des algorithmes de tri classiques par la rigidité de la structure de contrôle. En revanche une réalisation câblée peut permettre la parallélisation des opérations portant sur des lignes distinctes. Il est concevable de réaliser physiquement certains de ces réseaux par des mémoires à bulles (cf.e.g. [Chung, Luccio, Wong ; 1979]).

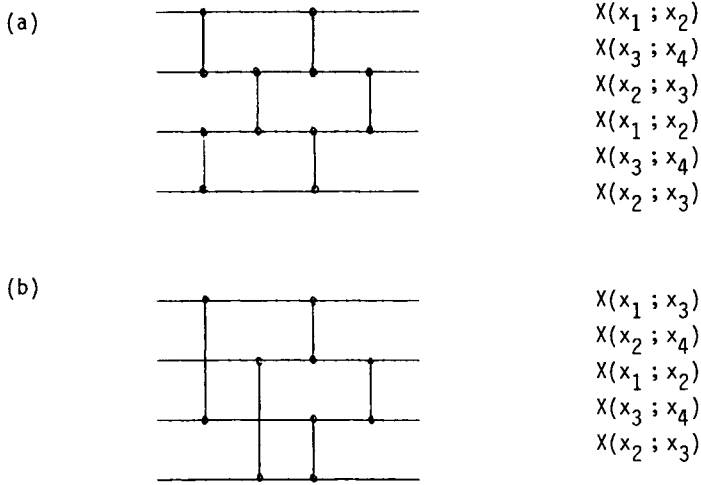


Figure 7 : Deux réseaux de tri de 4 éléments ; le réseau (b) est optimal vis-à-vis du nombre de comparateurs (les entrées et sorties sont étiquetées de haut en bas $x_1, x_2, \dots$).

L'algorithme de tri par fusion pair-impair de Batcher généralise la structure donnée par la fig. 7 (b). Il est défini récursivement par :

$$\begin{aligned}
 \text{TRI}(x_1, x_2, x_3, \dots) &\equiv \text{TRI}(x_1, x_3, x_5, \dots); \\
 &\quad \text{TRI}(x_2, x_4, x_6, \dots); \\
 &\quad \text{FUSION}(x_1, x_3, x_5, \dots; x_2, x_4, x_6, \dots).
 \end{aligned}$$

La procédure clef en est la procédure de fusion qui opère sur les positions paires et impaires selon le schéma donné figure 8.

(a) Fusion dans l'algorithme de Batcher

$$E_0 : X(1;2) X(3;4) X(5;6) \dots$$

$$E_1 : \delta = 2^\ell \quad 2^\ell \leq n < 2^{\ell+1}$$

$$X(2, 2+\delta-1) X(4, 4+\delta-1) X(6, 6+\delta-1) \dots$$

$$E_2 : \delta = 2^{\ell-1}$$

$$X(2, 2+\delta-1) X(4, 4+\delta-1) X(6, 6+\delta-1) \dots$$

...

$$E_{\ell-1} : \delta = 2$$

$$X(2, 2+\delta-1) X(4, 4+\delta-1) X(6, 6+\delta-1) \dots$$

(b) Le réseau de Fusion correspondant à $n = 16$

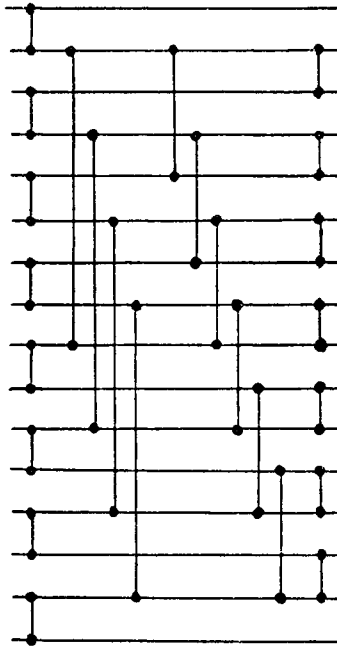


Figure 8 : la Fusion pair-impair.

2.2. L'analyse de la fusion.

On montre que l'analyse du tri sous la statistique des permutations[†] se ramène à l'analyse de la fusion pair-impair sur l'ensemble des suites 2-triées. C'est cette dernière analyse que nous présentons ici. Une suite finie $a_1, a_2, a_3, \dots$ est dite 2-triée si ses éléments vérifient les conditions

$$a_1 < a_3 < a_5 < a_7 \dots ; \text{ et } a_2 < a_4 < a_6 < a_8 \dots$$

Le nombre de suites 2-triées formées sur l'ensemble $[1..2n]$, vaut le coefficient binomial $\binom{2n}{n}$ puisqu'une telle suite est déterminée par le choix des éléments apparaissant en position paire (ce qui revient à choisir n éléments parmi $2n$). Le nombre de comparaisons effectuées par l'algorithme est indépendant de la nature des données et se détermine facilement en fonction de n . (cf. les références déjà citées). Le paramètre qui nous intéresse ici est le nombre moyen d'échanges $\bar{E}_n$ effectué lors de la fusion des $\binom{2n}{n}$ suites 2-triées de longueur $2n$. Si E_n est le nombre total d'échanges, on a ainsi :

$$\bar{E}_n = \frac{1}{\binom{2n}{n}} E_n$$

Knuth a observé qu'on peut donner de l'algorithme une présentation géométrique qui fournit la preuve la plus élégante de correction. Il existe en effet une correspondance naturelle entre les suites 2-triées sur $[1..2n]$ et les chemins reliant les sommets opposés d'un treillis $n \times n$.

Correspondance entre suites 2-triées et chemins :

Soit la suite $s = s_1 s_2 \dots s_{2n}$ où les s_j tous différents sont pris dans l'intervalle $[1..2n]$ et où

$$s_1 < s_3 < s_5 < \dots < s_{2n-1} \text{ et } s_2 < s_4 < s_6 < \dots < s_{2n}$$

On considère dans le plan cartésien les chemins à étapes horizontales $h|_0^1$ ou verticales $v|_1^0$ joignant l'origine au point $M|_n^n$. A la suite s , on associe un chemin $\alpha(s)$ considéré comme un mot de longueur $2n$ sur l'alphabet $\{h, v\}$ défini par

[†] analyse dans laquelle on considère comme également probables toutes les permutations des entrées de l'algorithme.

$\alpha(s) = c_1 c_2 \dots c_{2n}$ où c_j est une lettre h si j est en position paire dans s et où c_j est une lettre v si j est en position impaire dans s . De manière équivalente, le chemin est caractérisé par la suite de ses sommets

$$M_0 \begin{vmatrix} x_0=0 \\ y_0=0 \end{vmatrix}, M_1 \begin{vmatrix} x_1 \\ y_1 \end{vmatrix}, M_2 \begin{vmatrix} x_2 \\ y_2 \end{vmatrix} \dots M_{2n} \begin{vmatrix} x_{2n}=n \\ y_{2n}=n \end{vmatrix};$$

les x_k, y_k y sont définis de proche en proche par

$$\begin{cases} x_{k+1}=x_k \\ y_{k+1}=y_{k+1} \end{cases} \text{ si } k \text{ est en position impaire} \quad \begin{cases} x_{k+1}=x_k \\ y_{k+1}=y_{k+1} \end{cases} \text{ si } k \text{ est en}$$

position paire.

Ainsi la suite triée $1,2,3,\dots,2n$ est représentée par le mot (chemin) $(vh)^n$; la suite $2,1,4,3,6,5,\dots$ par le mot $(hv)^n$ etc...

Les étapes successives $E_0, E_1, \dots, E_{\ell-1}$ de l'algorithme de Batcher transforment la suite 2-triée s en une séquence de suites également 2-triées :

$$s = s^{(0)}; s^{(1)}; \dots s^{(k)}.$$

Il correspond par la bijection α une suite de chemins

$$c^{(0)} = \alpha(s^{(0)}); c^{(1)} = \alpha(s^{(1)}); \dots; c^{(k)} = \alpha(s^{(k)}),$$

et l'on démontre que $c^{(j+1)}$ se déduit de $c^{(j)}$ par une transformation de pliage P_j ce qu'on peut représenter par le schéma

$$\begin{array}{ccccccc} s^{(0)} & \xrightarrow{E_0} & s^{(1)} & \xrightarrow{E_1} & s^{(2)} & \dots & \xrightarrow{E_{\ell-1}} s^{(\ell)} \\ \alpha \downarrow & & \alpha \downarrow & & \alpha \downarrow & & \alpha \downarrow \\ c^{(0)} & \xrightarrow{P_0} & c^{(1)} & \xrightarrow{P_1} & c^{(2)} & \dots & \xrightarrow{P_{\ell-1}} c^{(\ell)} \end{array}$$

Géométriquement, la transformation P_0 transforme les parties d'un chemin en-dessus de la diagonale principale en leurs symétriques, par rapport à la diagonale et laisse invariantes les parties situées en-dessous de la diagonale. De

même la transformation P_j pour $j > 0$ transforme un chemin par pliage autour de la parallèle $\Delta_{\ell-j}$ à la diagonale principale située en dessous de celle-ci à distance $2^{\ell-j}$; en sorte que les points situés sous $\Delta_{\ell-j}$ sont transformés par symétrie autour de $\Delta_{\ell-j}$, alors que les segments situés au-dessus sont laissés invariants. En d'autres termes :

- P_0 transforme un point $\begin{vmatrix} x \\ y \end{vmatrix}$ en $\begin{vmatrix} x' \\ y' \end{vmatrix}$ avec
 $x' = y$ et $y' = x$ si $y < x$; $x' = x$ et $y' = y$ si $y \geq x$;
- P_j transforme un point $\begin{vmatrix} x \\ y \end{vmatrix}$ en $\begin{vmatrix} x' \\ y' \end{vmatrix}$ avec
 $x' = y - \delta$ et $y' = x - \delta$ si $y > x + \delta$; $x' = x$ et $y' = y$ si $y \leq x + \delta$.

La suite de transformations $P_0, P_1, \dots, P_{\ell-1}$ transforme tout chemin en le chemin associé à la suite totalement triée. Les pliages successifs réduisent progressivement l'écart du chemin à la diagonale comme représenté à la figure 9 reprise de l'article de Sedgewick.

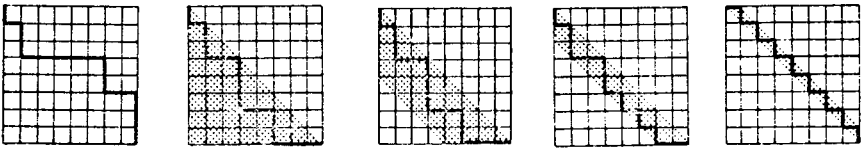


Figure 9 : La suite des pliages d'un chemin.

On vérifie de surcroît (cf. les références déjà citées) que chaque segment transformé par pliage en un segment distinct -on dira qu'un tel segment est renversé- correspond au déplacement d'un élément par l'algorithme de fusion (soit un "demi-échange"). La détermination de E_n (nombre total d'échanges) se trouve ainsi ramenée à l'évaluation sur l'ensemble des $\binom{2n}{n}$ chemins du nombre de segments transformés par pliage au cours de la suite d'opérations $P_0, P_1, \dots, P_{\ell-1}$.

Il est clair qu'un segment situé à distance $d \geq 0$ de la première diagonale -c'est-à-dire un segment horizontal joignant des points $\begin{vmatrix} y+d \\ y \end{vmatrix}$ et $\begin{vmatrix} y+d+1 \\ y \end{vmatrix}$ ou bien un segment vertical joignant des points $\begin{vmatrix} y+d+1 \\ y \end{vmatrix}$ et $\begin{vmatrix} y+d+1 \\ y+1 \end{vmatrix}$ - subit un nombre de renversements qui est une fonction de d .

On peut alors définir le code de réflexion d'un tel segment s lequel est une fonction de d seulement comme un mot $r_{\ell-2} r_{\ell-3} \dots r_0$ où $r_{\ell-1-j}$ vaut 1 si l'application de $P_{\ell-1-j}$ après $P_{\ell-1}, P_{\ell-2}, \dots, P_{\ell-j}$ conduit à un renversement et où r_j vaut 0 sinon. En d'autres termes :

$$r_j = 1 \text{ si } P_{\ell-1-j} \circ P_{\ell-j} \circ \dots \circ P_{\ell-1}(s) \neq P_{\ell-j} \circ P_{\ell-j+1} \circ \dots \circ P_{\ell-1}(s)$$

$$r_j = 0 \text{ sinon.}$$

Ainsi lorsque $n=16$, un segment à distance 9 comme le segment s d'extrémités $\begin{smallmatrix} 9 \\ 0 \end{smallmatrix}, \begin{smallmatrix} 10 \\ 0 \end{smallmatrix}$ se transforme

- par P_1 en un segment d'extrémités $\begin{smallmatrix} 1 \\ 8 \end{smallmatrix}, \begin{smallmatrix} 2 \\ 8 \end{smallmatrix}$
- par P_2 en un segment d'extrémités $\begin{smallmatrix} 4 \\ 5 \end{smallmatrix}, \begin{smallmatrix} 4 \\ 6 \end{smallmatrix}$
- par P_3 en un segment d'extrémités $\begin{smallmatrix} 4 \\ 5 \end{smallmatrix}, \begin{smallmatrix} 4 \\ 6 \end{smallmatrix}$
- par P_4 en un segment d'extrémités $\begin{smallmatrix} 4 \\ 5 \end{smallmatrix}, \begin{smallmatrix} 5 \\ 5 \end{smallmatrix}$.

Ceci montre que le code de réflexion de s est 1101 (cf. figure 10).

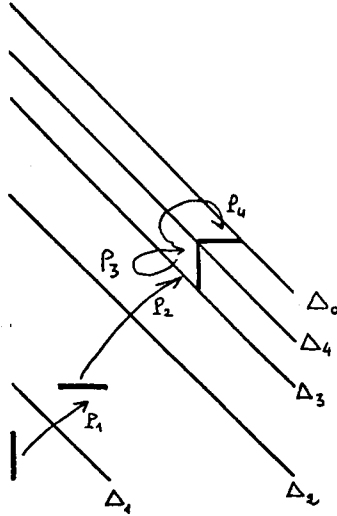


Figure 10 : *Pliages successifs d'un segment.*

On vérifie facilement en comparant les procédés de construction, que le code de réflexion d'un segment situé à distance d de la diagonale coïncide avec la représentation en code Gray de l'entier d .

Soit alors $N_{d,n}$ le nombre de segments verticaux dans l'ensemble des $\binom{2n}{n}$ chemins, situés à distance d de la diagonale principale sous cette diagonale. Par symétrie, $N_{d,n}$ est aussi le nombre de segments horizontaux situés à distance d

et au dessus de la diagonale. Avec, comme précédemment, $\gamma(n)$ qui désigne le nombre de 1 dans la représentation Gray de l'entier n , le nombre d'échanges associés aux $N_{d,n}$ segments sous la diagonale à distance d est $\gamma(d) N_{d,n}$; la quantité homologue relative aux $N_{d,n}$ segments au-dessus de la diagonale et à distance d est $(1+\gamma(d)) N_{d,n}$ (le terme 1 traduit le coût du pliage P_0). Nous avons ainsi

$$\bar{E}_n = \frac{1}{\binom{2n}{n}} \sum_d (1+2\gamma(d)) N_{d,n}.$$

Nous rencontrerons au chapitre suivant des problèmes d'énumération voisins du calcul des $N_{d,n}$. Ici les $N_{d,n}$ se calculent comme une somme de produits de nombres de chemins selon le diagramme de la figure 11; de sorte que

$$N_{d,n} = \sum_j \binom{2j+d-1}{j} \binom{2n+d-2j}{n-j}.$$

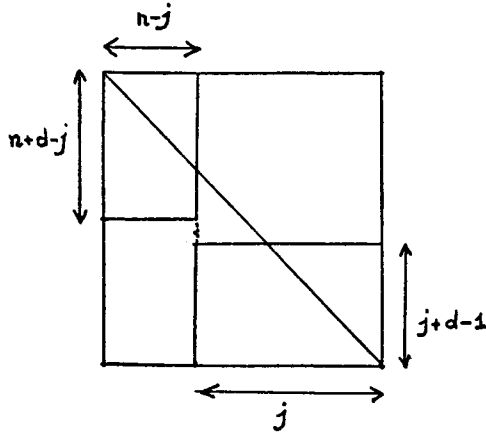


Figure 11 : Détermination des $N_{d,n}$.

Cette somme se réduit (cf. [Sedgewick, 1978] ; [Knuth, 1968]) de sorte que finalement

$$N_{d,n} = \sum_{0 \leq k < d} \binom{2n}{n-k}.$$

Substituant dans l'expression de $\bar{E}_n$ et après sommation par partie, l'on obtient l'expression :

$$\bar{E}_n = \frac{1}{\binom{2n}{n}} \sum_{k \geq 1} \binom{2n}{n-k} (2T(k)+k),$$

avec comme précédemment $T(m) = \sum_{j < m} \gamma(j)$.

2.3. L'analyse asymptotique.

Le point de départ est l'expression du nombre moyen d'échanges

$$\bar{E}_n = \frac{n}{2} + 2 \sum_{k \geq 1} T(k) \frac{\binom{2n}{n-k}}{\binom{2n}{n}},$$

où $T(k)$ désigne la fonction somme des chiffres des représentations Gray des entiers $1, 2, \dots, k-1$. La situation est ici très voisine du problème du nombre moyen de registres, bien que les deux analyses ne soient pas réductibles l'une à l'autre. Le résultat de l'analyse asymptotique est le suivant ([Sedgewick ; 1978]).

THEOREME 2. Soit $\bar{E}_n$ le nombre moyen d'échanges effectués lors de la fusion par l'algorithme de Batcher de deux suites triées de taille n . Il existe une fonction $M : \mathbb{R} \rightarrow \mathbb{R}$ continue, périodique et de période 1, telle que

$$\bar{E}_n = \frac{1}{4} n \log n + n M(\log_4 n) + O(\sqrt{n} \log^2 n).$$

La fonction M est développable en série de Fourier $M(x) = \sum_{k \in \mathbb{Z}} m_k e^{2ik\pi x}$, et les coefficients sont donnés par :

$$m_0 = 2 \log \Gamma\left(\frac{1}{4}\right) - \log \pi - \frac{\gamma + 2}{4 \ln 2} - \frac{3}{4} = 0.385417\dots$$

$$m_k = \frac{2\Gamma(x_k/2)\zeta(x_k, \frac{1}{4})}{(\ln 2)^{x_k}(1 + x_k)} \quad \text{pour} \quad x_k = \frac{2k\pi i}{\ln 2}, \quad k \neq 0.$$

Preuve : On commence en limitant dans la somme l'indice k à $k < \sqrt{n} \ln n$

$$\bar{E}_n = \frac{n}{2} + 2 \sum_{1 \leq k < \sqrt{n} \ln n} T(k) \frac{\binom{2n}{n-k}}{\binom{2n}{n}} + O(ne^{-\ln^2 n}),$$

puis en utilisant l'approximation gaussienne des coefficients binomiaux ;

$$\bar{E}_n = \frac{n}{2} + 2 \sum_{1 \leq k < \sqrt{n} \ln n} T(k) e^{-k^2/n} (1 + O(n^{-1} \log^4 n)),$$

les termes d'erreur introduits absorbant l'erreur en $O(ne^{-\ln^2 n})$. Le théorème 1 fournit de $T(k)$ une expression explicite qui, par substitution dans la somme, donne :

$$\bar{E}_n = \frac{n}{2} + 2 \sum_{1 \leq k < \sqrt{n} \ln n} \left(\frac{k \log k}{2} + kG(\log k) \right) e^{-k^2/n} (1 + O(n^{-1} \log^4 n)),$$

où G est continue, périodique et possède une série de Fourier explicitement

déterminée.

On scinde la somme en deux parties :

$$\begin{aligned}\bar{E}_n &= \frac{n}{2} + (C_n + 2D_n)(1 + O(n^{-1} \log^4 n)) \\ C_n &= \sum_{1 \leq k < \sqrt{n} \ln n} k \log k e^{-k^2/n} \\ D_n &= \sum_{1 \leq k < \sqrt{n} \ln n} k G(\log k) e^{-k^2/n}.\end{aligned}$$

La somme C_n apparaît comme une somme des valeurs de la fonction

$f(x) = x \log x e^{-x^2/n}$ et la formule sommatoire d'Euler-Mac-Laurin prise sous la forme

$$\sum_{1 \leq k < m} f(k) = \int_1^m f(x) dx - \frac{f(m) - f(1)}{2} + \frac{f'(m) - f'(1)}{12} + O\left(\int_1^m f''(x) dx\right),$$

avec $m = \sqrt{n} \ln n$ montre l'estimation

$$C_n = \int_1^{\sqrt{n} \ln n} x \log x e^{-x^2/n} dx + O(\sqrt{n} \ln n).$$

L'extension des limites d'intégration à l'intervalle $[0; +\infty]$ n'introduit qu'un terme d'erreur en $O(1)$; l'intégrale résultante s'évalue par le changement de variable $y = x/\sqrt{n}$, d'où par sommation la valeur :

$$\begin{aligned}C_n &= \frac{n}{2} \frac{1}{\ln 2} \int_0^\infty y e^{-y^2} dy + \frac{n}{\ln 2} \int_0^\infty y \ln(y) e^{-y^2} dy + O(\sqrt{n} \ln n) \\ &= \frac{n}{2} \frac{1}{\ln 2} \left(\frac{1}{2}\right) + \frac{n}{\ln 2} \left(-\frac{\gamma}{4}\right) + O(\sqrt{n} \ln n) = \frac{n \log n}{4} - \frac{\gamma n}{4 \ln 2} + O(\sqrt{n} \ln n).\end{aligned}$$

La somme donnant D_n fait intervenir la fonction $G(x)$ qui n'est pas dérivable.

On reprend la majoration, utilisée au chapitre 2, de l'écart entre somme et intégrale par l'oscillation de la fonction sommée :

$$\left| f(j) - \int_j^{j+1} f(x) dx \right| \leq \text{osc}(f, [j, j+1]).$$

Ici $f(x) = x G(\log x) e^{-x^2/n}$ où

$$\begin{aligned}G(x) &= \frac{1 - \{x\}}{2} + 2^{2-\{x\}} h(2^{\{x\}} - 2) \\ h(x) &= \sum_{j \geq 0} 2^{-j} t(2^j x).\end{aligned}$$

Comme la fonction triangulaire $t(x)$ est lipschitzienne de rapport $\frac{1}{2}$, la décomposition de la somme donnant $h(x)$ montre comme précédemment que l'oscillation de h sur un intervalle d'amplitude $k < 1$ est en $O(k \log \frac{1}{k})$; la même

estimation vaut pour l'oscillation de G sur un intervalle d'amplitude k ; de la sorte, on a

$$\text{osc}(G(x), [\log j ; \log(j+1)]) = O\left(\frac{\log j}{j}\right)$$

puisque l'amplitude de l'intervalle $[\log j ; \log(j+1)]$ est essentiellement $k = \frac{1}{j}$.

Quant à $f(x)$, il vient :

$$\text{osc}(f(x), [j ; j+1]) = O(\log j),$$

d'où finalement

$$|D_n - \int_1^{\sqrt{n} \ln n} x G(\log x) e^{-x^2/n} dx| \leq \sum_{1 \leq k < \sqrt{n} \ln n} O(\log k) = O(\sqrt{n} \log^2 n).$$

Comme précédemment, on ramène les limites d'intégration à $[0 ; +\infty]$, ce qui n'introduit qu'un terme d'erreur en $O(\sqrt{n} \log^2 n)$ et l'on en déduit que

$$D_n = \int_0^\infty x G(\log x) e^{-x^2/n} dx + O(\sqrt{n} \log^2 n) = n \int_0^\infty y G(\log y + \frac{1}{2} \log n) e^{-y^2} dy$$

en posant $y = x/\sqrt{n}$.

Ainsi en regroupant les termes obtient-on pour $\bar{E}_n$ l'expression

$$\bar{E}_n = \frac{1}{4} n \log n + n M(\log_4 n) + O(\sqrt{n} \log^2 n),$$

où la fonction M continue, périodique et de période 1, est définie par :

$$M(u) = -\frac{\gamma}{4 \ln 2} + \int_0^\infty y G(\log y + u) e^{-y^2} dy.$$

Le calcul de la série de Fourier de $M(u)$ s'effectue de manière analogue au développement du chapitre II. On écrit la série de Fourier de M sous la forme

$$\sum_{k \geq 0} m_k e^{-2ik\pi u},$$

avec

$$m_0 = -\frac{\gamma}{4 \ln 2} + \int_0^1 \int_0^\infty y G(\log y + u) e^{-y^2} dy du$$

$$m_k = \int_0^1 \int_0^\infty y G(\log y + u) e^{-y^2} dy e^{-2ik\pi u} du \text{ pour } k \neq 0.$$

Le changement de variable $v = \log y + u$ permet, en utilisant la périodicité de G de décomposer les intégrales doubles :

$$m_0 = -\frac{\gamma}{4 \ln 2} + \int_0^\infty y e^{-y^2} dy \int_0^1 G(v) dv$$

$$m_k = \int_0^\infty y e^{-y^2} dy e^{2ik\pi \log y} \int_0^1 e^{-2ik\pi v} G(v) dv \text{ pour } k \neq 0.$$

Les intégrales portant sur la variable y sont des intégrales Eulériennes exprimables au moyen de la fonction Γ :

$$\int_0^{\infty} y e^{-y^2} dy = \frac{1}{2} \Gamma(1) = \frac{1}{2}$$

$$\int_0^{\infty} y e^{-y^2} e^{2ik\pi \log y} dy = \frac{1}{2} \Gamma\left(\frac{ik\pi}{1n2}\right) ;$$

les intégrales portant sur la variable y sont les coefficients de Fourier de G calculés à la section précédente, de sorte que, toutes réductions faites, on obtient les valeurs

$$m_0 = 2 \log \Gamma\left(\frac{1}{4}\right) - \log \pi - \frac{\gamma+2}{4 \ln 2} - \frac{3}{4}$$

$$m_k = \frac{2}{\ln 2} \frac{\Gamma(x_k/2) \zeta(x_k/4)}{x_k(x_k+1)} \quad \text{avec} \quad x_k = \frac{2ik\pi}{\ln 2} \quad \square$$

Cette alternative à la dérivation de Sedgewick repose sur des méthodes analytiques plus élémentaires, et montre la relation entre la périodicité de la fonction arithmétique qui apparaît dans l'analyse, et le résultat de l'évaluation asymptotique. A titre d'illustration, la figure 12 (reprise de l'article de Sedgewick) indique la comparaison des valeurs exactes et estimées du coefficient du terme linéaire intervenant dans $\bar{E}_n$.

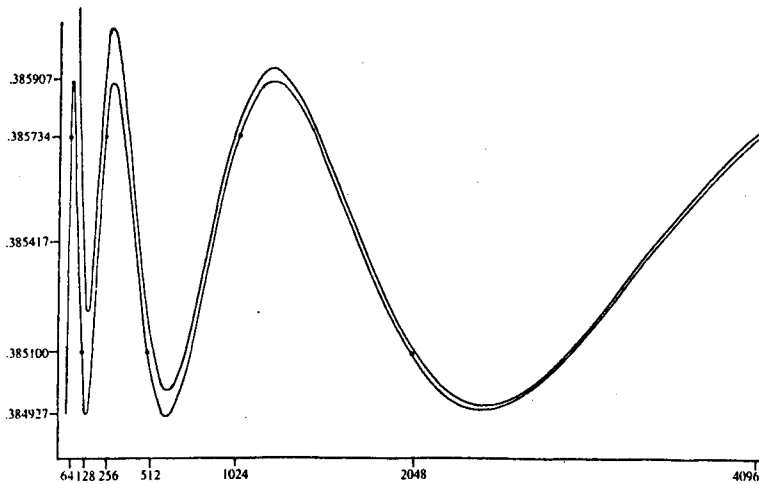


Figure 12 : Le terme linéaire dans le nombre moyen d'échanges :
valeur exacte et valeur asymptotique.

3. TRANSFORMATION DE MELLIN : VARIANCE DU NOMBRE DE REGISTRES, TAILLE D'ARBRE PARFAIT PLONGE.

Nous présentons ici l'estimation par transformation de Mellin de sommes combinatoires ou apparaissent des convolutions de coefficients binomiaux et de quantités arithmétiques qui ne sont pas directement liées à des systèmes de numération. C'est notamment le cas des fonctions non-linéaires de la valuation dyadique (carré, exponentielle) qui apparaissent dans l'estimation de la variance du nombre de registres ou de la taille moyenne du plus grand arbre parfait plongé dans un arbre de taille n .

La méthode utilisée par [de Bruijn, Knuth, Rice ; 1972] pour traiter le problème de la hauteur des arbres et par [Sedgewick ; 1978] pour l'analyse des réseaux de tri, est en fait très générale. Soit à évaluer une somme combinatoire

$$A_n = \sum_{k \geq 0} \alpha_k \Delta^r \binom{2n}{n-k}$$

où l'on suppose que les quantités α_k sont de l'ordre d'une puissance de $\ln k$:

$$\alpha_k = O(\ln^* k) \text{ pour tout } k.$$

On introduit une fonction B définie par

$$B(x) = \sum_{k \geq 1} \alpha_k e^{-k^2 x^2} H_r(kx)$$

où $H_r(t)$ est le r -ième polynôme d'Hermite. L'approximation classique des coefficients binomiaux et de leurs différences montre que

$$\frac{A_n}{\binom{2n}{n}} = B(n^{-1/2}) n^{-r/2} + O(n^{-(r+1)/2} \ln^* n).$$

Le problème est ainsi ramené à l'estimation du comportement de $B(x)$ lorsque $x \rightarrow 0$.

La méthode d'évaluation consiste à calculer la transformée de Mellin de $B(x)$ en utilisant diverses propriétés algébriques élémentaires de la transformation. On obtient ainsi une fonction $B^*(s)$ qui permet de retrouver $B(x)$ par transformation inverse. La transformation inverse exprimée par une intégrale dans le plan complexe s'évalue par le calcul des résidus lorsque sont satisfaites certaines conditions analytiques, et il apparaît que chaque résidu fournit un terme du développement asymptotique de $B(x)$ au voisinage de l'origine.

Rappelons brièvement les propriétés élémentaires de la transformation de Mellin.

La transformation de Mellin d'une fonction $f(x)$ définie sur $[0; +\infty]$ est une fonction $f^*(s)$ donnée par l'intégrale :

$$(M_0) \quad f^*(s) = \int_0^{\infty} x^{s-1} f(x) dx,$$

et est encore notée $\mathcal{M}[f(x); s]$. Si f est continue, possède un ordre en zéro et un ordre à l'infini :

$$f(x) = O(x^\alpha) \text{ au voisinage de } 0^+; f(x) = O(x^{-\beta}) \text{ au voisinage de } +\infty,$$

la transformée $f^*(s)$ est définie dans la bande

$$-\alpha < \operatorname{Re}(s) < \beta.$$

D'autre part, on a élémentairement (par changement de variables et intégration par parties) :

$$(M_1) \quad \mathcal{M}[f(x^m); s] = \frac{1}{m} \mathcal{M}[f(x); s/m], \quad m \geq 0;$$

$$(M_2) \quad \mathcal{M}[f(ax); s] = a^{-s} \mathcal{M}[f(x); s];$$

$$(M_3) \quad \mathcal{M}\left[\frac{df}{dx}; s\right] = -(s-1) \mathcal{M}[f(x); s-1];$$

cette dernière égalité étant valide dans un domaine où

$$\lim_{x \rightarrow 0^+} x^{s-1} f(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow \infty} x^{s-1} f(x) = 0.$$

L'intégrale Eulérienne définissant pour $\operatorname{Re}(s) > 0$ la fonction gamma

$$\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx,$$

peut-être considérée comme une transformée de Mellin :

$$(M_4) \quad \mathcal{M}[e^{-x}; s] = \Gamma(s) \text{ pour } \operatorname{Re}(s) > 0.$$

Enfin, il existe pour la transformée de Mellin un théorème d'inversion de même nature que pour la transformée de Fourier (cf. e.g. [Widder; 1971]) :

Si l'intégrale $\int_0^{\infty} x^{s-1} f(x) dx$ converge absolument pour $\operatorname{Re}(s) = c$ et si $f(x)$ est à variation bornée au voisinage de x , alors ⁺

$$(M_5) \quad \frac{f(x+) + f(x-)}{2} = \lim_{R \rightarrow \infty} \frac{1}{2i\pi} \int_{c-R_i}^{c+R_i} f(s) x^{-s} ds.$$

Revenant à notre problème, on part de l'intégrale donnant la transformée de Mellin qu'on applique à la fonction B . On obtient la suite de dérivations :

⁺ $f(x+) = \lim_{\substack{\varepsilon \rightarrow 0 \\ \varepsilon > 0}} f(x+\varepsilon); f(x-) = \lim_{\substack{\varepsilon \rightarrow 0 \\ \varepsilon < 0}} f(x-\varepsilon)$

$$\begin{aligned}
 B^*(s) &= \int_0^\infty x^{s-1} \sum_{k \geq 1} \alpha_k e^{-k^2 x^2} H_r(kx) dx & (M_0) \\
 &= \int_0^\infty x^{s-1} \sum_{k \geq 1} \alpha_k \frac{d^r(e^{-u^2})}{du^r} \Big|_{u=kx} dx \\
 &= \sum_{k \geq 1} \alpha_k \int_0^\infty x^{s-1} \frac{d^r(e^{-u^2})}{du^r} \Big|_{u=kx} dx \\
 &= \sum_{k \geq 1} \alpha_k k^{-s} \int_0^\infty x^{s-1} \frac{d^r(e^{-x^2})}{dx^r} dx & (M_2) \\
 &= \sum_{k \geq 1} \alpha_k k^{-s} (-1)^r (s-1)(s-2) \dots (s-r) \int_0^\infty x^{s-r-1} e^{-x^2} dx & (M_3) \\
 &= \sum_{k \geq 1} \alpha_k k^{-s} (-1)^r (s-1)(s-2) \dots (s-r) \frac{1}{2} \Gamma\left(\frac{s-r}{2}\right) & (M_1, M_4).
 \end{aligned}$$

Ainsi finalement :

$$B^*(s) = \frac{(-1)^r}{2} \xi(s)(s-1)(s-2) \dots (s-r) \Gamma\left(\frac{s-r}{2}\right),$$

où $\xi(s) = \sum_{k \geq 1} \alpha_k k^{-s}$ est la série génératrice de Dirichlet de la suite $\{\alpha_k\}_{k \geq 1}$.

La condition $\operatorname{Re}(s) > \max(1, r)$ est suffisante pour justifier cette dérivation. En effet, dans la zone $\operatorname{Re}(s) > 1$ la série génératrice de Dirichlet de ξ est définie[†] et normalement convergente dans $\operatorname{Re}(s) \geq 1 + \delta > 1$. La condition $\operatorname{Re}(s) > r$ permet ici l'utilisation de la formule de transformation des dérivées. Enfin la décroissance rapide à l'infini de e^{-x^2} permet dans la région $\operatorname{Re}(s) > \max(1, r)$ l'intervention de la sommation avec l'intégration.

Pour $\operatorname{Re}(s) > \max(1, r)$, on vérifie que $B^*(s)$ est continue et satisfait les conditions du théorème d'inversion (M_5) de sorte que l'on a :

$$B(x) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \xi(s) \Gamma\left(\frac{s-r}{2}\right) (1-s)(2-s) \dots (r-s) x^{-s} ds \quad \text{où } c \geq \max(1, r)$$

Supposons $\xi(s)$ prolongeable analytiquement et méromorphe dans $\mathbb{C}$. Sans rechercher les conditions analytiques les plus générales de validité, nous observons que le résidu de l'intégrande autour d'un pôle $s_0 = \sigma + i\tau$ de multiplicité k se calcule par les développements locaux

$$\begin{aligned}
 \frac{1}{2} \xi(s) \Gamma\left(\frac{s-r}{2}\right) (1-s) \dots (r-s) &= \frac{1}{(s-s_0)^k} [a_0 + a_1(s-s_0) + \dots + a_{k-1}(s-s_0)^{k-1} + o((s-s_0)^k)] \\
 x^{-s} &= x^{-s_0} (1 + (s-s_0) \ln x + \frac{(s-s_0)^2}{2!} \ln^2 x + \dots + \frac{(s-s_0)^{k-1}}{k!} \ln^k x + o((s-s_0)^k))
 \end{aligned}$$

[†] On a fait l'hypothèse que $\alpha_k = O(\ln^* k)$.

et se trouve ainsi être de la forme :

$$x^{-\sigma} P_{k-1}(\ln x) [\cos(\tau \ln x) + i \sin(\tau \ln x)],$$

où P_{k-1} est un polynôme de degré $k-1$ exactement.

Ainsi lorsqu'il est justifié de calculer l'intégrale donnant $B(x)$ par résidus, l'on obtient un développement asymptotique de $B(x)$ au voisinage de zéro dont les termes sont tels que :

- la partie réelle d'un pôle donne l'ordre principal $x^{-\sigma}$ du terme ;
- la partie imaginaire, si elle est non-nulle, conduit à un terme périodique fonction de $\ln x$;
- la multiplicité du pôle, suivant $k \geq 1$, conduit à un polynôme de degré $(k-1)$ en $(\ln x)$.

Cette méthode est particulièrement adaptée à l'évaluation de sommes où les fonctions arithmétiques sont des fonctions de la valuation dyadique. Soit en effet $\theta : \mathbb{N} \rightarrow \mathbb{N}$ une fonction définie sur les entiers, et soit

$$\theta(x) = \sum_{n \geq 0} \theta(n) x^n$$

sa série génératrice ordinaire. L'évaluation de la somme

$$B(x) = \sum_{k \geq 1} \theta(v_2(k)) e^{-k^2 x^2} H_r(kx)$$

par transformation de Mellin conduit, comme on l'a vu, à la transformée

$$B^*(s) = \frac{1}{2} \xi(s) \Gamma\left(\frac{s-r}{2}\right) (1-s)(2-s) \dots (r-s)$$

où $\xi(s)$ est la série génératrice de Dirichlet de $\{\theta(v_2(k))\}_{k \geq 1}$. Une suite

$\{\theta(v_2(k))\}_{k \geq 1}$ a une série génératrice de Dirichlet qui s'exprime au moyen de la série génératrice ordinaire $\theta(x)$. En effet, par définition

$$\xi(s) = \sum_{k \geq 1} \frac{\theta(v_2(k))}{k^s} = \frac{\theta(0)}{1^s} + \frac{\theta(1)}{2^s} + \frac{\theta(0)}{3^s} + \frac{\theta(2)}{4^s} + \frac{\theta(0)}{5^s} + \frac{\theta(1)}{6^s} + \dots$$

En décomposant $\theta(n)$:

$$\theta(n) = \theta(0) + (\theta(1) - \theta(0)) + (\theta(2) - \theta(1)) + \dots + (\theta(n) - \theta(n-1))$$

et en regroupant les termes dans ξ , on obtient :

$$\begin{aligned}\xi(s) = & \frac{\theta(0)}{1^s} + \frac{\theta(0)}{2^s} + \frac{\theta(0)}{3^s} + \frac{\theta(0)}{4^s} + \frac{\theta(0)}{5^s} + \frac{\theta(1)}{6^s} + \frac{\theta(0)}{7^s} + \frac{\theta(3)}{8^s} + \dots \\ & + \frac{\theta(1)-\theta(0)}{2^s} + \frac{\theta(1)-\theta(0)}{4^s} + \frac{\theta(1)-\theta(0)}{6^s} + \frac{\theta(1)-\theta(0)}{8^s} + \dots \\ & + \frac{\theta(2)-\theta(1)}{4^s} + \frac{\theta(2)-\theta(1)}{8^s} + \dots \\ & + \frac{\theta(3)-\theta(2)}{8^s} + \dots\end{aligned}$$

Cette manipulation purement algébrique montre l'égalité formelle :

$$\begin{aligned}\xi(s) = & \zeta(s) [\theta(0) + \frac{\theta(1)-\theta(0)}{2^s} + \frac{\theta(2)-\theta(1)}{4^s} + \frac{\theta(3)-\theta(2)}{8^s} + \dots] \\ = & \zeta(s) [\theta(\frac{1}{2^s}) - \frac{1}{2^s} \theta(\frac{1}{2^s})] \\ = & \zeta(s) \theta(\frac{1}{2^s}) (1 - \frac{1}{2^s}),\end{aligned}$$

égalité dont la justification analytique ne présente en principe pas de difficulté.

APPLICATION 1. Variance du nombre de registres.

Soit V_n la variance du nombre de registres sur l'ensemble des arbres de taille n , et soient $M_n^{(\alpha)}$ pour $\alpha = 1, 2, \dots$ les moments de la distribution du nombre de registres. On a classiquement

$$V_n = M_n^{(2)} - (M_n^{(1)})^2,$$

et ici les $M_n^{(\alpha)}$ sont donnés par :

$$M_n^{(\alpha)} = \frac{1}{B_n} \sum_{k \geq 1} (v_2(k))^\alpha \Delta^2 \binom{2n}{n+k-1}.$$

Le calcul du moment d'ordre 1 (moyenne) a déjà été effectué. Il suffit donc d'évaluer le moment d'ordre 2. Ceci conduit à introduire la fonction

$$C(x) = \sum_{k \geq 1} (v_2(k))^2 e^{-k^2 x^2} H_2(kx),$$

et l'on a :

$$M_n^{(2)} = C\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{1}{\sqrt{n}}\right).$$

La transformée de Mellin de $C(x)$ est

$$C^*(s) = \frac{1}{2} \zeta(s) \Gamma\left(\frac{s}{2} - 1\right) (s-1)(s-2) \left(1 - \frac{1}{2^s}\right) \theta\left(\frac{1}{2^s}\right)$$

où :

$$\theta(x) = \sum_{n \geq 0} n^2 x^n = \frac{d^2}{dx^2} \left(\frac{1}{1-x}\right) - \frac{d}{dx} \left(\frac{1}{1-x}\right) = \frac{x(1+x)}{(1-x)^3};$$

d'où finalement pour $\text{Re}(s) \geq 2$, en tenant compte de la propriété fonctionnelle de la fonction gamma :

$$C^*(s) = \frac{2^s+1}{(2^s-1)^2} (s-1) \zeta(s) \Gamma\left(\frac{s}{2}\right).$$

Cette fonction possède des singularités sur le seul axe imaginaire (incluant 0) :

- un pôle triple en $s=0$
- des pôles doubles en $s = \frac{2ik\pi}{\ln 2}$ pour $k \neq 0$.

On sait en effet que la fonction

$$\eta(s) = s(s-1) \zeta(s) \Gamma\left(\frac{s}{2}\right)$$

est entière et à partir des valeurs classiques

$$\zeta(0) = -\frac{1}{2}; \quad \zeta'(0) = -\frac{1}{2} \ln 2\pi; \quad \Gamma(1) = 1; \quad \Gamma'(1) = -\gamma,$$

on calcule facilement :

$$\eta(0) = 1; \quad \eta'(0) = -\frac{\gamma+2}{2} + \ln 2\pi.$$

Introduisons également la notation

$$\epsilon(s) = (s-1) \zeta(s) \Gamma\left(\frac{s}{2}\right).$$

Le calcul par résidus de $C(x)$ comme transformée inverse de $C^*(s)$ donne :

- pour le résidu en 0, une contribution

$$c_1(x) = \log^2 x - 2 \log x \left(\frac{\eta'(0)}{\ln 2} - \frac{1}{2} \right) + \frac{1}{3} - \frac{\eta'(0)}{\ln 2} + \frac{\eta''(0)}{(\ln 2)^2}$$

- pour la somme des résidus en $x_k = \frac{2ik\pi}{\ln 2}$, une contribution de :

$$c_2(x) = -2 \log x \pi(x) + \rho(x) - \pi(x)$$

où
$$\pi(x) = \frac{1}{\ln 2} \sum_{k \geq 0} \epsilon(x_k) e^{-x_k \ln x} \quad \text{et} \quad \rho(x) = \frac{2}{(\ln 2)^2} \sum_{k \geq 0} \epsilon'(x_k) e^{-x_k \ln x}.$$

Chacune des séries π et ρ est une série de Fourier absolument convergente en $\ln x$, dont les coefficients décroissent exponentiellement.

Le choix d'un contour rectangulaire joignant les points $2+i\rho$; $-N-\frac{1}{2}+i\rho$;

$-N-\frac{1}{2}-i\rho$; $2-i\rho$ où N est entier montre lorsqu'on fait tendre N et ρ vers

l'infini que la fonction $C(x)$ est égale à la somme de $c_1(x)$, $c_2(x)$ et d'un

terme d'erreur en $O(x^m)$ où l'on peut rendre m arbitrairement grand. On en déduit ainsi pour le moment $M_n^{(2)}$ l'estimation

$$M_n^{(2)} = c_1\left(\frac{1}{\sqrt{n}}\right) + c_2\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{\ln^* n}{\sqrt{n}}\right).$$

Reprenant l'expression de la moyenne, telle que nous l'avons calculée au chapitre

précédent, on a $M_n^{(1)} = D\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{\ln^* n}{\sqrt{n}}\right)$ où l'on peut donner à $D(x)$ la forme

$$D(x) = -\log x + \left(\frac{\eta'(0)}{\ln 2} - \frac{1}{2}\right) + \pi(\ln x).$$

Ainsi la variance $V_n = M_n^{(2)} - (M_n^{(1)})^2$ peut elle s'écrire sous la forme

$$V_n = C_1\left(\frac{1}{\sqrt{n}}\right) + C_2\left(\frac{1}{\sqrt{n}}\right) - \left(D\left(\frac{1}{\sqrt{n}}\right)\right)^2 + O\left(\frac{\ln^* n}{\sqrt{n}}\right).$$

Le calcul de $C_1(x) + C_2(x) - D(x)$ montre que les termes en $\log^2 x$ et $\log x$ s'éliminent, ce qui est cohérent avec l'estimation de la variance donnée au chapitre précédent ; de fait :

$$C_1(x) + C_2(x) - (D(x))^2 = \frac{1}{12} + \frac{(\eta'(0))^2 + \eta''(0)}{(\ln 2)^2} - \pi^2(\ln x) - \frac{2\eta'(0)}{\ln 2} \pi(\ln x) + \frac{2}{(\ln 2)^2} \rho(\ln x)$$

D'où finalement :

PROPOSITION. La variance du nombre de registres est une fonction continue périodique et de période 1, de $\log_4 n$ qui s'écrit sous la forme

$$V_n = \alpha + \beta \pi^2(-\log_4 x) + \gamma \pi(-\log_4 x) + \delta \rho(-\log_4 x)$$

où les fonctions π et ρ sont développables en séries de Fourier et où les constantes $\alpha, \beta, \gamma, \delta$ s'expriment au moyen des valeurs de $\Gamma(s)$ $\zeta(s)$ et deux de leurs dérivées en $s=1$ et $s=0$.

APPLICATION 2. Taille moyenne du plus grand arbre parfait plongé.

Ainsi que nous l'avons vu au chapitre II, la taille du plus grand sous-arbre parfait plongé dans un arbre binaire dont le nombre de registres est p , vaut $2^p - 1$. Ainsi, désignant par $\bar{p}_n$ la taille moyenne de ce plus grand sous-arbre parfait plongé, la moyenne étant calculée sur l'ensemble des B_n arbres binaires de taille n , nous avons l'égalité

$$\bar{p}_n + 1 = \frac{1}{B_n} \sum_{k \geq 1} 2^{v_2(k)} \Delta^2 \binom{2n}{n+k}.$$

A cette somme se trouve associée la fonction

$$B(x) = \sum_{k \geq 1} 2^{v_2(k)} e^{-k^2 x^2} H_2(kx),$$

et l'on a la relation :

$$\bar{p}_n + 1 = B(n^{-1/2}) n^{1/2} + O(\ln^* n).$$

La transformée de Mellin de $B(x)$, soit $B^*(s)$, se détermine à partir de la série génératrice ordinaire de la suite $\{2^n\}_{n \geq 0}$:

$$\Theta(z) = \sum 2^n z^n = \frac{1}{1-2z},$$

de sorte que, pour $\operatorname{Re}(s) \geq 2$:

$$\begin{aligned} B^*(s) &= \frac{1}{2}(s-1)(s-2) \Gamma\left(\frac{s-2}{2}\right) \zeta(s) \frac{1}{1-2^{-s+1}} (1-2^{-s}) = \frac{1}{4}(s-1)(s-2) \Gamma\left(\frac{s-2}{2}\right) \zeta(s) \frac{2^s-1}{2^{s-1}-1} \\ &= \frac{1}{2}(s-1) \Gamma\left(\frac{s}{2}\right) \zeta(s) \frac{2^s-1}{2^{s-1}-1}. \end{aligned}$$

La fonction $B^*(s)$ possède des pôles simples en $s=1$, $s=1+\chi_k$ où $\chi_k = \frac{2ik\pi}{\ln 2}$.

Les résidus de $B^*(s)x^{-s}$ valent :

$$\begin{aligned} - \quad \text{en } s=1 : & \frac{1}{2} \Gamma\left(\frac{1}{2}\right) x^{-1} = \frac{\sqrt{\pi}}{2x} \\ - \quad \text{en } s=1+\chi_k : & -\frac{1}{4} \chi_k \Gamma\left(\frac{1+\chi_k}{2}\right) \zeta(1+\chi_k) x^{-1} e^{-(1+\chi_k)\ln x}. \end{aligned}$$

Ici encore, la décroissance très rapide de la fonction gamma à l'infini montre que la somme des résidus en les points $(1+\chi_k)$ est une série de Fourier qui converge absolument, en sorte que si l'on définit la fonction $\omega(u)$ par :

$$\omega(u) = \frac{\sqrt{\pi}}{2} - \frac{1}{4e} \sum_{k \neq 0} \chi_k \Gamma\left(\frac{1+\chi_k}{2}\right) \zeta(1+\chi_k) e^{-\chi_k u},$$

ω est continue périodique de période 1 et indéfiniment dérivable. Une estimation des termes d'erreurs montre alors l'égalité :

$$B(x) = \frac{1}{x} \omega(\ln x) + O(x^m) \quad \text{pour } m > 0 \text{ arbitraire.}$$

Revenant à l'expression de la quantité $\bar{P}_n$, il vient en posant $W(u) = \omega(u \ln 2)$:

PROPOSITION. La taille moyenne du plus grand arbre parfait plongé dans un

arbre binaire satisfait à l'estimation :

$$\bar{P}_n = \sqrt{n} W(\log_4 n) + O(\ln^* n)$$

où W est continue, périodique de période 1 et développable en série de Fourier :

$$W(u) = \frac{\sqrt{\pi}}{2} - \frac{1}{4e} \sum_{k \neq 0} \chi_k \Gamma\left(\frac{1+\chi_k}{2}\right) \zeta(1+\chi_k) e^{-2ik\pi u},$$

formule dans laquelle $\chi_k = \frac{2ik\pi}{\ln 2}$.

CHAPITRE IV

EVALUATION DES STRUCTURES DE DONNEES DYNAMIQUES.

Ce chapitre est consacré à l'évaluation d'une large classe d'algorithmes opérant sur des structures de données dynamiques. Cette question a été abordée sous des angles divers[†] :

- l'allocation dynamique de mémoire séquentielle est discutée par [Knuth ; 1968] et rejoint le cas particulier du type de données "pile" de ce chapitre.
- le type de données "partition d'ensemble" dans lequel on réalise la gestion de classes d'équivalences est utile à divers algorithmes de théorie des graphes ; l'analyse de l'algorithme de Aho-Hopcroft-Ullman est réalisée par [Knuth, Schönhage ; 1978].
- le type de données "dictionnaire" a été étudié généralement sous l'effet d'une suite d'adjonctions (cf.[Knuth ; 1973] pour une discussion des algorithmes); [Snyder ; 1977] a également considéré le coût maximal d'une suite d'opérations. Notre approche s'appuie sur la problématique de J. Françon ([Françon ; 1978]). Il s'agit d'estimer le coût des structures dont la taille varie avec le temps -c'est notamment le cas des fichiers- et qui sont soumises de diverses manières à des opérations d'adjonction, interrogation et suppression. La comparaison de deux structures différentes, ou encore de deux implantations d'une même structure ne peut habituellement se limiter à la comparaison des coûts individuels de chaque opération. En effet certains algorithmes privilégient certaines opérations : dans une liste non triée l'adjonction est plus simple à réaliser que dans une liste

[†] Nous renvoyons à [Flajolet, Françon, Vuillemin ; 1979 b] pour une bibliographie plus complète.

triée (il suffit de placer la clef en tête de la structure) ; en revanche, la recherche d'une clef absente de la structure y est plus coûteuse (on doit explorer toute la structure). Il arrive même que les fonctions de coût relatives à une opération ne soient pas des fonctions monotones de la taille de la structure : c'est notamment le cas des structures binômiales de J. Vuillemin, dont on a déjà dit que le comportement était lié à certaines propriétés de la représentation binaire des entiers.

Pour ces deux raisons, on s'efforce d'évaluer non les coûts individuels d'opérations -appelés coûts unitaires- mais au contraire les coûts de suites d'opérations -appelés coûts intégrés-. Il existe pour l'évaluation des coûts intégrés une approche probabiliste qui fait certaines hypothèses sur l'arrivée, la durée de vie ou la distribution des clefs ; ainsi par exemple : arrivées obéissant à une loi de Poisson, clefs dont la durée de vie dans le fichier est donnée par une loi exponentielle, clefs tirées uniformément sur un intervalle.

Nous suivons dans ce travail l'approche de nature combinatoire de Françon. Dans cette approche, l'évaluation d'un algorithme s'effectue sur l'ensemble de toutes les évolutions possibles de la structure. On considèrera ainsi pour un dictionnaire toutes les positions relatives possibles en lesquelles une clef peut arriver. D'une certaine manière la démarche prolonge la méthode classique d'analyse des algorithmes de tri lorsqu'on l'effectue sur l'ensemble des $n!$ permutations des clefs d'entrées.

Utilisant des résultats de combinatoire des permutations de [Françon, Viennot ; 1979], Françon a le premier montré que les comptes d'histoires, relatifs à certains types de données usuels, conduisent à des dénombrements en termes de quantités combinatoires connues : nombres factoriels, nombres d'involutions, nombres d'Euler. Ceci a permis les premiers calculs explicites de coûts intégrés. Notre travail continue celui de Françon : en nous plaçant dans un cadre plus général, nous montrons que l'expression naturelle des séries génératrices de dénombrements d'histoires est en termes de fractions continues. Ce résultat ouvre

aux histoires de fichier l'accès à l'algèbre classique des fractions continues. Nous obtenons ainsi (section 2) des expressions de séries génératrices de dénombrements concernant :

(i) les histoires des structures classiques de dictionnaires, files de priorité, etc... pour lesquels nous retrouvons par l'algèbre les résultats de Françon.

(ii) les histoires de hauteur bornée : évaluation de coûts sous l'hypothèse que la taille du fichier ne dépasse jamais une certaine taille maximum.

(iii) les histoires relatives à une population bornée : évaluation de coûts de transactions qui s'effectuent par rapport à un fichier de référence fixe.

Nous montrons (section 2) que les dénombrements d'histoires de hauteur bornée et populations bornées conduisent dans les cas classiques à des expressions où interviennent les polynômes orthogonaux de l'analyse : polynômes de Laguerre, Hermite, Poisson-Charlier et Meixner. Nous poursuivons (section 3) cette étude des séries génératrices en montrant que le passage des coûts unitaires d'opérations individuelles aux coûts intégrés de suites d'opérations se traduit sur les séries génératrices par une transformation intégrale linéaire. Ainsi par exemple, pour les dictionnaires, la série génératrice exponentielle des coûts intégrés $\hat{K}(z)$ est reliée à la série génératrice ordinaire des coûts unitaires $c(u)$ par la relation :

$$\hat{K}(z) = \frac{2}{2-z} \int_0^t \frac{c(u) du}{\sqrt{(1-u)(t-u)}}, \quad \text{où } t = \left(\frac{z}{2-z}\right)^2.$$

Cette transformation se simplifie dans le cas des files de priorité et, entre une série génératrice $\bar{K}(z)$ de coûts intégrés, et la série génératrice ordinaire $c(u)$ des coûts unitaires, on a la relation

$$\bar{K}(z) = \frac{1}{\sqrt{1-2t}} \frac{1}{1-t} c\left(\frac{t}{1-t}\right).$$

Ceci conduit pour les implantations courantes à des expressions explicites très simples des coûts intégrés permettant la comparaison effective des structures.

Il existe un lien de nature informatique entre ce chapitre et les précédents. Un grand nombre d'implantations efficaces de structures dynamiques,

reposent sur l'utilisation des arbres : arbres binaires de recherche, arbres binomiaux, arbres tournois, arbres pagodes... . Par ailleurs, le type de donnée pile, qui est du domaine de la statistique Catalan, apparaît comme cas limite particulièrement simple.

L'introduction des fractions continues et l'obtention des séries génératrices ont été présentées dans un rapport préliminaire ([Flajolet ; 1978]) ; ceci forme la base des sections 1 et 2. Les résultats de la section 3 concernant les formes intégrales des coûts intégrés résultent d'un travail commun effectué avec J. Françon et J. Vuillemin ([Flajolet, Françon, Vuillemin ; 1979 a et b]). Les aspects plus spécifiquement combinatoires sont discutés au chapitre suivant.

1. FICHIERS ET FRACTIONS CONTINUES.

Dans cette partie, nous rappelons brièvement la problématique informatique des histoires de fichiers, renvoyant à [Françon, 1978] pour certains compléments de présentation. Nous montrons ensuite comment à chaque type de donnée -dictionnaire, file de priorité...- déterminé par une fonction de possibilité arbitraire, se trouve associée une fraction continue équivalente (au sens des séries formelles) à la série génératrice des histoires relatives à cette structure.

1.1. La problématique informatique.

Nous considérons ici les algorithmes de manipulation de fichiers qui opèrent des adjonctions, interrogations ou suppressions, sur une structure dont la taille varie ainsi avec le temps. L'accès aux "clefs" peut se faire par position ou par valeur. Nous considérerons tout particulièrement les cinq classes d'algorithmes suivantes, dites encore types de données⁺ :

a. le type de donnée "dictionnaire" : les clefs sont accessibles par leur valeur et pour chaque opération, toutes les valeurs des clefs sont permises ; en particulier les interrogations peuvent porter sur des clefs présentes ou absentes du fichier. Ces structures sont classiquement implantées en mémoire centrale, sous forme de listes triées ou non, et d'arbres équilibrés ou non. Elles sont utilisées pour maintenir à jour des tables d'identificateurs, des ensembles d'objets dans des fichiers

b. le type de donnée "file de priorité" : les clefs sont accessibles par valeur ; à chaque instant peut s'opérer une adjonction d'une clef à valeur quelconque, ou une suppression portant sur la clef de plus petite valeur ("de priorité minimale"). Les files de priorité peuvent s'implanter comme des dictionnaires, mais il est avantageux de tirer partie des restrictions d'accès lors de la suppression en

⁺ cf. [Knuth ; 1973] pour une présentation des principaux algorithmes.

utilisant les structures d'arbre tournoi, tas, arbre pagode ([Françon, Viennot, Vuillemin ; 1978]) ou encore arbre binomial ([Vuillemin ; 1978]). Elles sont utilisées lorsqu'on doit gérer sous forme de file d'attente -par exemple dans un système d'exploitation d'ordinateur- un ensemble d'individus obéissant à un système de priorités. On les retrouve également dans les algorithmes de plus court chemin dans les graphes.

c. le type de donnée "table de symboles" : les clefs y sont accessibles par valeur, les adjonctions sont quelconques, les suppressions portent sur la dernière clef entrée, et les interrogations portent sur des clefs présentes dans le fichier. Ces structures s'implantent comme des dictionnaires. L'absence d'interrogation négative (en utilisation normale) est typique de l'organisation de la table des symboles par un compilateur d'un langage dont tous les identificateurs sont déclarés (ALGOL, PASCAL,...).

d. le type de donnée "pile" : les clefs y sont accessibles par position ; les opérations permises sont l'adjonction et la suppression, et elles portent toujours sur l'élément en première position, le "sommet de pile". Cette structure très simple s'implante sous forme de liste chaînée ou de vecteur.

e. le type de donnée "liste linéaire" : les clefs y sont accessibles par position ; les opérations permises sont l'adjonction et la suppression lesquelles peuvent prendre à chaque instant toutes les valeurs compatibles avec la taille de la structure. Cette structure peut s'implanter sous forme d'arbre de positions.

Les problèmes pratiques qui peuvent se poser concernant les implantations de l'un de ces types de donnée sont nombreux :

- estimation de l'utilité de maintenir une structure triée, par opposition à une structure non triée, une structure d'arbre par opposition à une structure de liste ; les résultats peuvent varier selon que l'utilisation est de type dictionnaire ou file de priorité ;
- évaluation de l'influence sur les performances des algorithmes du choix

de certains paramètres : par exemple la taille d'une mémoire tampon dans un système de mémoire hiérarchisée, lorsqu'on plante une pile ou une file de priorité etc..

Nous nous proposons de montrer que moyennant certaines hypothèses simples, on peut apporter une réponse à de nombreuses questions de ce type.

On part d'un ensemble $\mathcal{K}$ de clefs dont on suppose, comme pour l'ensemble des nombres réels ou des chaînes de caractères, qu'il est infini, totalement ordonné et dense. Pour la structure de dictionnaire qui donne accès aux clefs par leur valeur, on définit une suite d'opérations comme une séquence de la forme

$$O_1(k_1) O_2(k_2) \dots O_n(k_n)$$

où $n \in \mathbb{N}$ et où pour chaque i , $1 \leq i \leq n$, on a $k_i \in \mathcal{K}$ et $O_i \in \{A, S, Q\}$. L'élément $O(k)$ s'interprète comme effectuer l'opération d'adjonction (A), interrogation (Q) ou suppression (S) sur la clef k . Une suite d'opérations est licite si les adjonctions portent sur des clefs non présentes dans le fichier, et si les suppressions portent sur des clefs présentes dans le fichier ; plus précisément, on définit la suite d'états de fichiers $F_0, F_1, \dots, F_n$ avec $F_i \subset \mathcal{K}$, attachée à une suite d'opération, par les règles

- (i) $F_0 = \emptyset$; et pour tout i tel que $1 \leq i \leq n$:
- (ii) $F_i = F_{i-1} \cup \{k_i\}$ si $O_i = A$ et $k_i \notin F_{i-1}$;
- (iii) $F_i = F_{i-1} \setminus \{k_i\}$ si $O_i = S$ et $k_i \in F_{i-1}$;
- (iv) $F_i = F_{i-1}$ si $O_i = Q$.

La suite d'opérations est licite si F_n est bien défini par les règles ci-dessus. On modifie cette définition pour les files de priorité en imposant que la suppression porte sur l'élément minimal du fichier courant, soit pour la suite d'états de fichiers les règles :

- (i) $F_0 = \emptyset$, et pour tout i tel que $1 \leq i \leq n$:
- (ii) $F_i = F_{i-1} \cup \{k_i\}$ si $O_i = A$ et $k_i \in F_{i-1}$;
- (iii) $F_i = F_{i-1} \setminus \{k_i\}$ si $O_i = S$ et $k_i = \min(F_{i-1})$.

Enfin, il est commode de considérer les tables de symboles comme des files de priorité sur lesquelles on n'effectue que des interrogations positives :

- (i) $F_0 = \emptyset$, et pour $1 \leq i \leq n$:
- (ii) $F_i = F_{i-1} \cup \{k_i\}$ si $O_i = A$ et $k_i \in F_{i-1}$
- (iii) $F_i = F_{i-1} \setminus \{k_i\}$ si $O_i = S$ et $k_i = \min(F_{i-1})$
- (iv) $F_i = F_{i-1}$ si $O_i = Q$ et $k_i = F_{i-1}$.

Exemple : Une suite d'opérations du type dictionnaire avec la suite des états de fichiers correspondant est :

A(2.7)	A(3.1)	Q(0.57)	A(1.4)	S(3.1)	Q(2.7)
$\emptyset$	{2.7}	{2.7;3.1}	{2.7;3.1}	{2.7;3.1;1.4}	{2.7;3.1}
					{2.7;3.1}.

Pour le type file de priorité où les suppressions portent toujours sur la clef de valeur minimale présente dans le fichier

A(2.7)	A(3.1)	A(10)	S(2.7)	S(3.1)	A(99)	S(10)
$\emptyset$	{2.7}	{2.7;3.1}	{2.7;3.1;10}	{3.1;10}	{10}	{10;99}
						{99}. $\square$

Pour la structure de liste linéaire, une suite d'opérations est une suite de la forme

$$O_1(p_1, k_1) O_2(p_2, k_2) \dots O_n(p_n, k_n)$$

où $n \in \mathbb{N}$ et où pour chaque i , $1 \leq i \leq n$, on a $k_i \in \mathcal{K}$, $p_i \in \mathbb{N}$ et $O_i \in \{A, S\}$. Chaque instruction de type $O(k, p)$ s'interprète comme : "effectuer l'opération O d'adjonction ou suppression de la clef k en p -ième position⁺ dans le fichier. Plus précisément on définit pour une telle suite d'opérations la suite des états de fichiers associée $S_0, S_1, \dots, S_n$ où $S_i \in \mathcal{K}^*$ est une séquence de clefs (non un ensemble comme dans le cas précédent) donnée par les règles :

- (i) $S_0 = \epsilon$ (la suite vide de $\mathcal{K}^*$) et pour tout i tel que $1 \leq i \leq n$:
- (ii) si $S_{i-1} = S_{i-1,1} S_{i-1,2} \dots S_{i-1,p_i-1} t_{i-1}$ et si $O_i = A$
alors $S_i = S_{i-1,1} S_{i-1,2} \dots S_{i-1,p_i-1} k_i S_{i-1,p_i} \dots S_{i-1,t_{i-1}}$
- (iii) si $S_{i-1} = S_{i-1,1} S_{i-1,2} \dots S_{i-1,t_{i-1}}$ et si $O_i = S$
alors $S_i = S_{i-1,1} S_{i-1,2} \dots S_{i-1,p_i-1} S_{i-1,p_i+1} \dots S_{i-1,t_{i-1}}$;

⁺

On suppose les positions numérotées à partir de zéro.

l'application de cette règle est soumise à la condition que $S_{i-1}, p_i = k_i$.

Une suite d'opérations de type pile est une suite d'opérations de type liste linéaire dans laquelle chacune des position p_i vaut 0.

Il convient d'observer que cette définition des suites d'opérations n'est qu'une représentation commode de la manière dont ces types de données sont habituellement utilisés. Ainsi par exemple la suppression dans une file de priorité s'effectue généralement par une commande $S_{\min}$ (suppression du minimum) sans qu'il soit nécessaire que l'utilisateur spécifie la valeur de ce minimum ; il en va de même d'une suppression dans une liste linéaire Il est clair que notre définition n'est qu'un raccourci de présentation et qu'on pourrait donner des définitions équivalentes plus proches des conditions pratiques d'utilisation mais plus hétérogènes.

Nous définissons maintenant des représentations canoniques de suites d'opérations encore appelées histoires. Ces suites canoniques sont en nombre fini pour chaque longueur n et ainsi permettent de ramener l'analyse à des problèmes combinatoires de dénombrement.

Les histoires relatives aux types de données dans lesquels l'accès aux clefs se fait par valeur, s'obtiennent en ne retenant que les informations relatives à l'ordre relatif des clefs dans le fichier courant. On introduit d'abord la notion de rang d'un élément x dans un ensemble F , quantité notée $\rho(x; F)$ et donnée par :

$\rho(x; F) = i$ si $f_{i-1} < x \leq f_i$ où $F = \{f_0; f_1; \dots; f_r\}$ avec $f_0 < f_1 < f_2 < \dots < f_r$ (par convention $f_{-1} = -\infty$ et $f_{r+1} = +\infty$).

Etant donnée une suite d'opérations pour un type de donnée dictionnaire, file de priorité ou table de symboles :

$$O_1(k_1) O_2(k_2) \dots O_n(k_n),$$

avec la suite d'états de fichiers $F_0, F_1, F_2 \dots F_n$, on appelle histoire la suite

$$h = O'_1(r_1) O'_2(r_2) \dots O'_n(r_n)$$

où pour $1 \leq i \leq n$ $r_i = \rho(k_i; F_{i-1})$ et

$$\begin{aligned} O'_i &= O_i \quad \text{si } O_i \in \{A; S\} ; O'_i = Q^+ \quad \text{si } O_i = Q \text{ et } k_i \in F_{i-1} ; \\ O'_i &= Q^- \quad \text{si } O_i = Q \text{ et } k_i \notin F_i . \end{aligned}$$

On compense ainsi la perte d'information sur la valeur des clefs en indiquant dans l'histoire le caractère positif ou négatif de l'interrogation. Les histoires relatives aux types de données dans lesquels l'accès aux clefs se fait par position, s'obtiennent en ne conservant que les informations relatives à la position des clefs dans le fichier courant. Ainsi pour le type de donnée liste linéaire ou pile, à une suite d'opérations

$$O_1(p_1; k_1) O_2(p_2; k_2) \dots O_n(p_n; k_n),$$

on associe l'histoire

$$O_1(p_1) O_2(p_2) \dots O_n(p_n).$$

Exemple : A la suite d'opérations de type file de priorité

A(2.7) A(3.1) A(10) S(2.7) S(3.1) A(99) S(10)

correspond l'histoire :

A(0) A(1) A(2) S(0) S(0) A(1) S(0) ;

à la suite de type dictionnaire

A(2.7) A(3.1) Q(0.57) A(1.4) S(3.1) Q(2.7) :

A(0) A(1) Q⁻(0) A(0) S(2) Q⁺(1) ;

et à la suite de type liste linéaire

A(0;2.7) A(0;1.7) A(2;2.6) S(1;2.7) A(2;0.57) :

A(0) A(0) A(2) S(1) A(2). $\square$

Le postulat de base de la théorie des histoires de fichiers proposée par J. Françon consiste à évaluer les algorithmes sur l'ensemble des histoires.

Une telle démarche n'est **a priori justifiée** que si les **coûts de suites d'opérations** ayant même histoire associée sont égaux. Désignant par $c(h)$ le coût d'une histoire -c'est-à-dire le coût de n'importe quelle suite d'opérations ayant cette histoire comme représentant canonique- on définit le coût intégré de n opérations comme la moyenne

$$K_n = \frac{1}{|X_n|} \sum_{h \in X_n} c(h)$$

des coûts rapportée à l'ensemble (fini) X_n des histoires de longueur n .

Nous n'entrerons pas ici dans tous les détails de la problématique pour lesquels nous renvoyons à ([Françon ; 1978] et [Flajolet, Françon, Vuillemin ; 1979]).

Cette méthode d'analyse se justifie pour les algorithmes de dictionnaire, file de priorité ou table de symboles qui opèrent par comparaison entre les clefs.

L'intérêt de cette réduction à un ensemble fini de cas est que les histoires peuvent être définies indépendamment comme des objets combinatoires, ce qui en rend possible le dénombrement.

Définition combinatoire des histoires.

(i) Un schéma s est un mot sur l'alphabet $\{A, S, Q^+, Q^-\}$ tel que pour tout préfixe gauche t de s , on ait[†]

$$|t|_A \geq |t|_S.$$

On note S l'ensemble des schémas.

(ii) A tout schéma $s = 0_1 0_2 \dots 0_n$ se trouve associée la suite des hauteurs $h_0 h_1 \dots h_n$ définie par

$$h_0 = 0 ; h_{i+1} = h_i + \begin{cases} 1 & \text{si } 0_{i+1} = A \\ 0 & \text{si } 0_{i+1} = Q^+, Q^- \\ -1 & \text{si } 0_{i+1} = S. \end{cases}$$

(iii) A chaque type de donnée Σ (dictionnaire, file de priorité...) se trouve associée une fonction de possibilité $\text{pos}^\Sigma : \{A, S, Q^+, Q^-\} \times \mathbb{N} \rightarrow \mathbb{N}$ où $\text{pos}(\omega; k)$ est donnée par la table suivante :

Σ	ω			
	A	S	Q^+	Q^-
Dictionnaire	k+1	k	k	k+1
File de priorité	k+1	1	0	0
Table des symboles	k+1	1	k	0
Liste linéaire	k+1	k	0	0
Pile	1	1	0	0

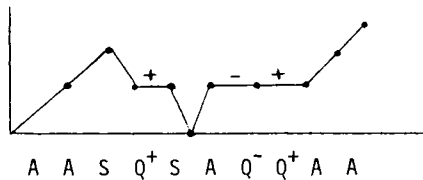
[†] $|w|_x$ désigne le nombre d'occurrences de la lettre x dans le mot w .

Une histoire de type Σ est un couple $h = (s, v)$ où $s = 0_1 0_2 \dots 0_n$ est un schéma et $v = v_1 v_2 \dots v_n$ est une suite d'entiers satisfaisant aux inégalités

$$0 \leq v_i < \text{pos}^\Sigma(0_i ; h_{i-1}).$$

Une telle histoire est encore notée $0_1(v_1) 0_2(v_2) \dots 0_n(v_n)$. La suite d'entiers $v_1 v_2 \dots v_n$ s'appelle la valuation de l'histoire, et la suite $0_1 0_2 \dots 0_n$ est le schéma de l'histoire. $\square$

Il est facile de vérifier que cette définition combinatoire des histoires coïncide pour chaque type de donnée avec la définition des histoires comme représentations canoniques de suites d'opérations : le schéma représente la suite des opérations lorsqu'on ne prend pas en compte les arguments (clefs ou positions) ; la condition (i) signifie qu'aucune suppression ne peut avoir lieu sur le fichier vide ; la hauteur h_i est égale à la taille du fichier courant après les opérations $0_1 0_2 \dots 0_{i-1}$. Un schéma est représentable par une ligne brisée dans le plan cartésien dont les sommets sont les points $(0, h_0) (1, h_1) (2, h_2) \dots (n, h_n)$, et par exemple au schéma $s = \text{AASQ}^+ \text{SAQ}^- \text{Q}^+ \text{AA}$ correspond le diagramme :

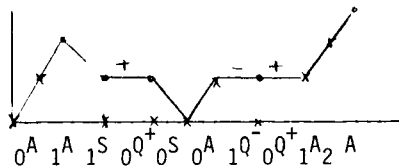


(on étiquette éventuellement les paliers horizontaux par + et -).

On représente les histoires par des diagrammes plans consistant en la représentation cartésienne du schéma complétée des valeurs successives de la valuation représentées par les points $(0; v_1) ; (1; v_2) ; (2; v_3) \dots (n-1; v_n)$ (les croix sur la figure). Ainsi à l'histoire de dictionnaire

$$h = (\text{AASQ}^+ \text{SAQ}^- \text{Q}^+ \text{AA}, 0110001012)$$

correspond le diagramme



La hauteur du schéma $s = 0_1 0_2 \dots 0_n$ auquel se trouve associée la suite des hauteurs $u = h_1 h_2 \dots h_n$ est par définition $\max_{1 \leq j \leq n} \{h_j\}$. On dit que le schéma termine à niveau (hauteur) h_n . L'opération 0_j est dite, dans ce contexte, opérer à niveau (hauteur) h_{j-1} . Enfin on étend aux histoires la terminologie relative aux schémas et l'on parlera ainsi de hauteur d'une histoire, de niveau d'une opération dans une histoire

Sous des hypothèses† de "stationnarité" (ou "aléatoirité") satisfaites par de nombreuses structures de données, on peut globalement décomposer le coût d'une histoire en une somme de coûts d'opérations individuelles. Plus précisément, on considère les structures pour lesquelles existe une famille de constantes C_{ω_k} dites coûts unitaires où $\omega \in \{A, S, Q^+, Q^-\}$ et $k \in \mathbb{N}$ telles qu'on ait l'égalité pour tout n :

$$\sum_{h \in \mathcal{H}_n} c(h) = \sum_{\substack{\omega \in \{A, S, Q^+, Q^-\} \\ k \geq 0}} C_{\omega_k} N_{\omega_k, n}$$

où $\mathcal{H}_n$ représente l'ensemble des histoires de longueur n et $N_{\omega_k, n}$ vaut le nombre d'opérations de type ω effectuées à niveau k dans l'ensemble de ces histoires. Moyennant cette **condition**, on peut formellement évaluer les coûts intégrés comme si les coûts de chaque opération ne dépendaient que de la nature de l'opération et de la taille du fichier ; on a alors pour la valeur du coût intégré de la structure sur l'ensemble $\mathcal{H}_n$ l'expression :

$$K_n = \frac{1}{|\mathcal{H}_n|} \sum_{\omega, k} C_{\omega_k} N_{\omega_k, n}.$$

Notons que les structures de liste triées ou non, les arbres binaires de recherche (sans équilibrage), les arbres tournois et arbres pagodes, les files binomiales vérifient cette hypothèse et les coûts unitaires s'évaluent eux-mêmes comme certains coûts moyens. Le reste de ce chapitre est consacré aux méthodes d'évaluation de coûts intégrés de telles structures.

† cf. [Françon ; 1978 et 1979].

1.2. Histoires de fichiers et fractions continues : généralités.

Nous traitons dans cette section les histoires de fichiers d'un type Σ tout à fait général données par une fonction de possibilités quelconque

$$\text{pos}(A ; k) = \alpha_k \quad \text{pos}(Q ; k) = \kappa_k \quad \text{pos}(S ; k) = \sigma_k.$$

Nous montrons l'expression en termes de fractions continues[†] des séries génératrices associées aux histoires de hauteur bornée aux histoires de hauteur quelconque et aux histoires vérifiant certaines conditions initiales ou finales.

Dans cette partie de l'exposé, il n'est pas nécessaire de distinguer entre interrogations positives et négatives. Il est commode ici de considérer les histoires comme des mots sur un alphabet infini $X = \{a_j^{(i)} ; s_j^{(i)} ; q_j^{(i)}\}_{i,j \geq 0}$ où $\omega_j^{(i)}$ représente la i^{e} possibilité de l'opération ω sur un fichier de taille j . Ainsi à toute histoire $h = (s, v)$ avec $s = \omega_1 \dots \omega_n$ et $v = v_1 \dots v_n$ dont la suite des hauteurs est $h_0 \dots h_n$ associe-t-on le mot

$$\begin{matrix} (v_1) & (v_2) & & (v_n) \\ \omega_{h_0} & \omega_{h_1} & \dots & \omega_{h_{n-1}} \end{matrix}.$$

On considère d'abord l'ensemble $\mathcal{H}$ des histoires dont la hauteur finale est nulle^{††}; cet ensemble est encore noté $\Sigma \mathcal{H}$ lorsqu'on veut mettre en évidence le type de structure Σ auquel les histoires sont relatives.

THEOREME 1^A. La série caractéristique non-commutative des histoires $\chi(\mathcal{H})$ admet le développement en fractions continues

$$\chi(\mathcal{H}) = \frac{1}{1 - (\Sigma q_0^{(i)}) - \frac{(\Sigma a_0^{(i)}) | (\Sigma s_1^{(i)})}{1 - (\Sigma q_1^{(i)}) - \frac{(\Sigma a_1^{(i)}) | (\Sigma s_2^{(i)})}{1 - (\Sigma q_2^{(i)}) - \frac{(\Sigma a_2^{(i)}) | (\Sigma s_3^{(i)})}{\dots}}}}$$

où $\frac{u|v}{w}$ dénote $uw^{-1}v$, et où dans chaque somme $\Sigma \omega_j^{(i)}$, l'indice de sommation i est astreint à la condition $0 \leq i < \text{pos}(\omega ; j)$.

[†] Pour les propriétés classiques des fractions continues, nous renvoyons à [Perron ; 1954] , [Wall ; 1967]

^{††} Ces histoires font jouer un rôle symétrique aux adjonctions et suppressions.

Preuve : La preuve de ce résultat utilise des notions de la théorie des séries formelles en indéterminées non commutatives, dont l'introduction dans les problèmes d'énumération combinatoire est due à Schützenberger (cf. les références données au chapitre I).

On part d'un codage des schémas qui étiquette chaque lettre $\{A, Q, S\}$ d'un schéma par sa hauteur dans le schéma. Soit Y l'alphabet infini $\{a_j; q_j; s_j\}_{j \geq 0}$. Ce codage est une application c de $\{A, Q, S\}^*$ dans Y^* qu'on définit par : $c(s_1 s_2 \dots s_n) = u_1 u_2 \dots u_n$ ssi est satisfaite la condition suivante pour tout m tel que $1 \leq m \leq n$:

- chaque lettre est étiquetée par sa hauteur dans le schéma, i.e. si $s_m = \omega$ où $\omega = A, S$ ou Q et si $|s_1 s_2 \dots s_{m-1}|_A - |s_1 s_2 \dots s_{m-1}|_S = \lambda$ alors $u_m = \omega_\lambda$.

Ainsi au schéma $s = AAASAQSAQS$ correspond le codage $c(s) = a_0 a_1 a_2 s_3 a_2 q_3 s_3 a_3 q_4 s_4$.

Soit $\mathcal{S}$ l'ensemble des schémas et $\mathcal{S}^{[h]}$ l'ensemble des schémas de hauteur $\leq h$. Nous allons donner une expression de la série caractéristique non commutative de l'ensemble $c(\mathcal{S}^{[h]})$ des schémas étiquetés de hauteur $\leq h$.

Rappelons les principes suivants : l'on se place dans l'algèbre large $C\langle\langle T \rangle\rangle$ d'un monoïde libre T^* construite sur le corps des nombres complexes ; en d'autres termes $C\langle\langle T \rangle\rangle$ est l'algèbre des séries formelles en les indéterminées non-commutatives T à coefficients complexes. Un élément $C\langle\langle T \rangle\rangle$ s'écrit

$$a = \sum_{u \in T^*} a_u \cdot u,$$

et les opérations de somme et produit de Cauchy sont classiquement définies par :

$$a + b = \sum_{u \in T^*} (a_u + b_u) \cdot u$$

$$a \cdot b = \sum_{u \in T^*} \left(\sum_{vw=u} s_v t_w \right) \cdot u.$$

La valuation d'une série a est

$$\text{val}(a) = \min\{|u|; a_u \neq 0\} \text{ où } |u| \text{ désigne la longueur du mot } u.$$

Par convention $\text{val}(0) = +\infty$. On dispose alors sur $C\langle\langle T \rangle\rangle$ d'une notion de convergence (formelle) : une suite $\{a_n\}_{n \geq 0}$ d'éléments de $C\langle\langle T \rangle\rangle$ a comme limite a ssi

$$\lim_{n \rightarrow \infty} \text{val}(a - a_n) = +\infty.$$

En d'autres termes a_n tend vers a ssi les coefficients des séries a_n se stabilisent progressivement, en partant des termes d'ordre le plus bas. Cette notion de convergence induit une notion de sommabilité pour les suites infinies. Les séries dont le terme constant est non nul possèdent un inverse multiplicatif. En particulier, on peut toujours supposer ce terme constant égal à 1 ; le quasi-inverse de la série a où $\text{val}(a) > 0$ est par définition la série $(1-a)^{-1}$ qui est encore donnée par

$$(1-a)^{-1} = \sum_{k \geq 0} a^k.$$

Dans le cadre de l'énumération d'ensembles de mots, on utilise la notion de série caractéristique. La série caractéristique de l'ensemble E est notée $\chi(E)$ ou $\text{car}(E)$ et est définie par

$$\chi(E) = \text{car}(E) = \sum_{u \in E} u.$$

Soient E et F des sous-ensembles de T^* (encore appelés langages). La réunion de E et F est notée indifféremment $E \cup F$ ou $E + F$. Le produit de concaténation de E et F est noté $E.F$ ou simplement EF . Enfin avec la notation du mot vide par ϵ , l'étoile de E , notée E^* , est définie par $E^* = \epsilon + E + E.E + E.E.E + \dots$. Sous des conditions générales de non-ambiguïté les opérations sur les ensembles se traduisent sur les séries caractéristiques.

LEMME. Soient E, F deux sous-ensembles de T^*

(i) $\text{car}(E+F) = \text{car}(E) + \text{car}(F)$ si $E \cap F = \emptyset$

(ii) $\text{car}(E.F) = \text{car}(E) \cdot \text{car}(F)$ si

$\forall u, u' \in E \quad \forall v, v' \in F : uv = u'v' \text{ implique } u = u' \text{ et } v = v'$

(on dit dans ce cas que le produit $E.F$ est non ambigu)

(iii) $\text{car}(E^*) = (1 - \text{car}(E))^{-1}$ si sont remplies les deux conditions

- $E^j \cap E^k = \emptyset$ pour tout couple i, j tel que $i \neq j$
- chaque produit E^k est non ambigu.

(on dit que l'étoile de E est non ambiguë).

□

Ici, chacun des ensembles $\mathcal{Y}^{[h]}$ peut être décrit par une expression régulière, c.à.d. une expression dans laquelle n'interviennent que les opérations d'union, de produit de concaténation et d'étoile. En effet :

$$\mathcal{Y}^{[0]} = q_0^*$$

$$\mathcal{Y}^{[1]} = (q_0 + a_0(q_1)^*s_1)^*$$

$$\mathcal{Y}^{[2]} = (q_0 + a_0(q_1 + a_1(q_2)^*s_2)^*s_1)^*,$$

et l'on vérifie par récurrence que

$$\mathcal{Y}^{[h+1]} = \sigma^{[h]}(\mathcal{Y}^{[h]})$$

où $\sigma^{[h]}$ est la substitution : $q_h \rightarrow (q_h + a_h(q_h)^*s_{h+1})$.

On obtient la série caractéristique de $\mathcal{Y}^{[h]}$ en remplaçant chaque opération sur les ensembles par son homologue sur les séries selon le lemme ci-dessus.

Il faut pour cela s'assurer que les opérations qui apparaissent sont non ambiguës. La non ambiguïté est trivialement vérifiée pour les unions et produits, lesquels portent sur des ensembles de mots formés de lettres différentes. Si l'on définit les ensembles

$$\mathcal{Y}^{[h,h]} = q_h^*$$

$$\mathcal{Y}^{[h,h-1]} = (q_{h-1} + a_{h-1}(q_h)^*s_h)^* \dots,$$

on vérifie que chaque opération étoile est non ambiguë : les couples $a_{h-1} - s_h$ constituent en effet une parenthétisation non ambiguë.

Ainsi, en introduisant la notation

$$\frac{a|c}{b} \text{ pour représenter } ab^{-1}c,$$

l'on obtient pour les séries caractéristiques $\text{car}(\mathcal{Y}^{[h]})$ les expressions

$$\text{car}(\mathcal{Y}^{[0]}) = \frac{1}{1-q_0}$$

$$\text{car}(\mathcal{Y}^{[1]}) = \frac{1}{1-q_0 - \frac{a_0|s_1}{1-q_1}}$$

$$\text{car}(\mathcal{Y}^{[2]}) = \frac{1}{1-q_0 - \frac{a_0|s_1}{1-q_1 - \frac{a_1|s_2}{1-q_2}}}$$

et en général
$$\text{car}(\mathcal{Y}^{[h]}) = \frac{1}{1-q_0 - \frac{a_0 | s_1}{1-q_1 - \frac{a_1 | s_2}{1-q_2 - \frac{a_2 | s_3}{\ddots \frac{a_{h-1} | s_h}{1-q_h}}}}}$$

Cette écriture montre clairement l'analogie avec les réduites d'une fraction continue.

D'après l'origine combinatoire du problème, on a la chaîne d'inclusions

$$\mathcal{Y}^{[0]} \subset \mathcal{Y}^{[1]} \subset \mathcal{Y}^{[2]} \dots \subset \mathcal{Y}^{[h]} \dots \subset \mathcal{Y},$$

avec pour $S^{[h]} = \text{car}(\mathcal{Y}^{[h]})$ la propriété :

$$\text{val}(S^{[h]} - S^{[h-1]}) \geq 2h.$$

Cette propriété traduit simplement le fait qu'un schéma de hauteur h a une longueur supérieure ou égale à $2h$.

On définit par analogie avec le cas classique, une fraction continue non commutative comme limite de réduites, en posant

$$\frac{1}{1-q_0 - \frac{a_0 | s_1}{1-q_1 - \frac{a_1 | s_2}{\ddots}}} = \lim_{h \rightarrow \infty} \frac{1}{1-q_0 - \frac{a_0 | s_1}{1-q_1 - \frac{a_1 | s_2}{\ddots \frac{a_{h-1} | s_h}{1-q_h}}}},$$

et la discussion précédente montre l'identité

$$\text{car}(S) = \frac{1}{1-q_0 - \frac{a_0 | s_1}{1-q_1 - \frac{a_1 | s_2}{\ddots}}}.$$

On passe enfin d'un schéma à l'ensemble des histoires correspondantes en remplaçant un symbole d'opération à niveau k soit ω_k par la somme de toutes les possibilités correspondantes. Pour cela on définit la substitution (morphisme) $\tau : Y \rightarrow C \langle\langle X \rangle\rangle$ par :

$$\tau(\omega_k) = \sum_i \omega_k^{(i)},$$

où $\omega \in \{A, S, Q\}$, $k \in \mathbb{N}$ et la somme porte sur les valeurs de l'index i telles que $0 \leq i < \text{pos}(\omega; k)$. Le morphisme τ est continu et l'on a les égalités

$$\text{car}(\mathcal{H}^{[h]}) = \tau(\text{car}(\mathcal{J}^{[h]})) \text{ d'où } \text{car}(\mathcal{H}) = \tau(\text{car}(\mathcal{J}));$$

la notation $\mathcal{H}^{[h]}$ désigne l'ensemble des histoires de hauteur $\leq h$. Cette égalité est équivalente à l'énoncé du théorème. $\square$

Soit H_n le nombre d'histoires revenant au niveau 0 relatives au système de possibilité Σ

$$\text{pos}(A; k) = \alpha_k \quad \text{pos}(S; k) = \sigma_k \quad \text{pos}(Q; k) = \kappa_k.$$

On obtient la série génératrice des quantités H_n , $H(z) = \sum H_n z^n$, à partir de $\text{car}(\mathcal{H})$ au moyen du morphisme qui envoie chaque variable de X sur la variable z . Ceci conduit au résultat suivant :

THEOREME 1^B. La série génératrice des nombres H_n comptant les histoires⁺ de longueur n admet le développement en fraction continue

$$\sum_{n \geq 0} H_n z^n = \frac{1}{1 - \kappa_0 z - \frac{\alpha_0 \sigma_1 z^2}{1 - \kappa_1 z - \frac{\alpha_1 \sigma_2 z^2}{1 - \kappa_2 z - \frac{\alpha_2 \sigma_3 z^2}{\dots}}}}.$$

La série génératrice des nombres $H_n^{[h]}$ comptant les histoires de longueur n et de hauteur $\leq h$ admet l'expression rationnelle

$$\sum_{n \geq 0} H_n^{[h]} z^n = \frac{1}{1 - \kappa_0 z - \frac{\alpha_0 \sigma_1 z^2}{1 - \kappa_1 z - \frac{\alpha_1 \sigma_2 z^2}{\dots}} \over 1 - \kappa_h z} \quad \square$$

⁺ Rappelons qu'il s'agit des histoires commençant et finissant à hauteur 0.

Les développements de ce type sont classiquement connus sous le nom de fractions continues de Jacobi ou J-fractions. Lorsque chacun des coefficients κ_j est nul, la fraction est dite de Stieltjes ou encore S-fraction.

1.3. Histoires de hauteur bornée ; passages à niveaux.

Les polynômes numérateurs et dénominateurs des réduites apparaissent également dans l'expression des séries génératrices d'histoires terminant à une hauteur donnée, ainsi que dans diverses autres séries génératrices. Nous présentons ici ces calculs valables pour un système de possibilités quelconque.

Nous considérons les classes $\mathcal{H}_{0,\ell}^{[h]}$ composées des histoires de hauteur $\leq h$ partant du niveau 0 et finissant au niveau ℓ ; avec les notations précédentes $\mathcal{H}_{0,0}^{[h]} = \mathcal{H}^{[h]}$ et $\mathcal{H} = \mathcal{H}_{0,0}^{[\infty]}$. Plus généralement, nous introduisons l'ensemble $\mathcal{H}_{k,\ell}^{[h]}$ des histoires en un sens étendu, dont la hauteur est $\leq h$, qui partent du niveau k et finissent au niveau ℓ ; de telles histoires étendues sont définies par le fait qu'en leur préfixant une suite de k adjonctions, on obtient une histoire qui part de 0 :

$$u \in \mathcal{H}_{k,\ell}^{[h]} \text{ ssi } a_0^{(0)} a_1^{(0)} \dots a_{k-1}^{(0)} u \in \mathcal{H}_{0,\ell}^{[h]}.$$

Avec des conventions évidentes, on pose $\mathcal{H}_{k,\ell} = \mathcal{H}_{k,\ell}^{[\infty]}$ etc On ajoute éventuellement un indice supplémentaire pour désigner la longueur de sorte que

$$\mathcal{H}_{k,\ell,n}^{[h]} = \{u \mid u \in \mathcal{H}_{k,\ell}^{[h]} \text{ et } |u| = n\}$$

et $\mathcal{H}_n^{[h]} = \mathcal{H}_{0,0,n}^{[h]}$ etc... .

On désigne enfin systématiquement les dénombrements par des lettres droites :

$H_{k,\ell,n}^{(h)} = \text{card}(\mathcal{H}_{k,\ell,n}^{[h]})$ etc..., et l'on utilise la notation classique des séries génératrices ordinaires : $H_{k,\ell}^{[h]}(z) = \sum_{n \geq 0} H_{k,\ell,n}^{[h]} z^n$ etc... .

Nous avons vu l'expression sous forme de réduites, des séries génératrices d'histoires de hauteur bornées

$$H^{[h]}(z) = \sum H_n^{[h]} z^n = \frac{1}{1 - \kappa_0 z - \frac{\alpha_0 \sigma_1 z^2}{1 - \kappa_1 z - \frac{\alpha_1 \sigma_2 z^2}{\dots}}}$$

Cette fraction peut se mettre sous la forme d'un quotient de deux polynômes, et l'on obtient :

PROPOSITION : La série génératrice des histoires de hauteur $\leq h$ vaut

$$H^{[h]}(z) = \frac{P_h(z)}{Q_h(z)},$$

où les P_h, Q_h vérifient la récurrence linéaire :

$$\begin{aligned} P_{-1}(z) &= 0 ; P_0(z) = 1 ; & P_h(z) &= (1 - \kappa_h z) P_{h-1}(z) - \alpha_{h-1} \sigma_h z^2 P_{h-2}(z) \\ Q_{-1}(z) &= 1 ; Q_0(z) = 1 - \kappa_0 z ; & Q_h(z) &= (1 - \kappa_h z) Q_{h-1}(z) - \alpha_{h-1} \sigma_h z^2 Q_{h-2}(z). \end{aligned}$$

Preuve : Cette proposition exprime seulement la classique récurrence linéaire des numérateurs et dénominateurs de réduites de fractions continues ; nous en rappelons ici le principe.

On considère pour tout h la fonction

$$g^{[h]}(y) = \frac{1}{1 - \kappa_0 z - \frac{\alpha_0 \sigma_1 z^2}{\dots \frac{1 - \kappa_h z - y}{1 - \kappa_h z - y}}}$$

obtenue en "marquant" par y le dernier niveau de la h -ième réduite. Cette fonction est homographique en y , et l'on a

$$g^{[h+1]}(y) = g^{[h]} \left(\frac{\alpha_h \sigma_{h+1} z^2}{1 - \kappa_h z - y} \right).$$

Ainsi existe sur les fonctions $g^{[h]}$ une récurrence qui n'est pas apparente sur les réduites, et l'on vérifie l'égalité

$$g^{[h]}(y) = \frac{P_h - P_{h-1}y}{Q_h - Q_{h-1}y},$$

où les P_h et Q_h sont donnés par la récurrence linéaire ci-dessus. En $y=0$, $g^{[h]}(y)$ vaut la h -ième réduite. $\square$

Ce même développement montre :

PROPOSITION. La série génératrice des histoires allant de la hauteur h à la hauteur h , et dont tous les sommets ont une hauteur non inférieure à h , vaut

$$H_{h,h}^{[\geq h]}(z) = \sum_n H_{h,h,n}^{[\geq h]} z^n = \frac{1}{\alpha_{h-1} \sigma_h z^2} \frac{Q_{h-1}(z) \cdot H(z) - P_{h-1}(z)}{Q_{h-2}(z) H(z) - P_{h-2}(z)},$$

où $H_{h,h,n}^{[\geq h]}$ dénote le nombre de telles histoires de longueur n .

Preuve : le théorème 1 montre que la série génératrice de ces histoires, $H_{h,h}^{[\geq h]}(z)$ vaut la fraction continue :

$$t = \frac{1}{1 - \kappa_h z - \frac{\alpha_h \sigma_{h+1} z^2}{1 - \kappa_{h+1} z - \frac{\alpha_{h+1} \sigma_{h+2} z^2}{\dots}}}$$

qu'on appelle encore h -ième troncature de la fraction continue donnant $H(z)$. D'après la preuve de la proposition précédente cette troncature t vérifie l'équation

$$g^{[h]}(\alpha_{h-1} \sigma_h z^2 t) = H(z),$$

d'où, en résolvant, la preuve de l'énoncé. $\square$

PROPOSITION. La série génératrice des histoires allant de la hauteur 0 à la hauteur k vaut

$$H_{0,k}(z) = \sum_n H_{0,k,n} z^n = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_k z^k} [Q_{k-1}(z) H(z) - P_{k-1}(z)],$$

où $H_{0,k,n}$ est le nombre de telles histoires ayant longueur n .

Preuve : La décomposition des histoires de $\mathcal{H}_{0,k}$ selon les derniers passages aux niveaux $0, 1, 2, \dots, k-1$ fournit la factorisation non ambiguë

$$\mathcal{H}_{0,k} = \mathcal{H}_{0,0}^{[\geq 0]} a_0 \mathcal{H}_{1,1}^{[\geq 1]} a_1 \mathcal{H}_{2,2}^{[\geq 2]} \dots a_{k-1} \mathcal{H}_{k,k}^{[\geq k]};$$

soit en passant aux séries génératrices :

$$H_{0,k}(z) = \alpha_0 \alpha_1 \dots \alpha_{k-1} z^k H_{0,0}^{[\geq 0]} H_{1,1}^{[\geq 1]} \dots H_{k,k}^{[\geq k]}.$$

L'énoncé découle immédiatement du remplacement des $H_{j,j}^{[\geq j]}$ par leurs expressions données par la proposition précédente. $\square$

Des considérations élémentaires de symétrie montrent pour les $H_{k,0}(z)$ l'expression générale

$$H_{k,0}(z) = \frac{\sigma_1 \sigma_2 \dots \sigma_k}{\alpha_0 \alpha_1 \dots \alpha_{k-1}} H_{0,k}(z).$$

Ces résultats donnent la forme des séries génératrices des "passages à niveau".

Soit $N_{\omega,k,n}$ le nombre d'opérations de type ω effectuées à niveau k dans l'ensemble des histoires de longueur n , (du niveau 0 au niveau 0). Les séries génératrices correspondantes vérifient les égalités :

$$NA_k(z) = \sum_{n \geq 0} NA_{k,n} z^n = \alpha_k z H_{0,k}(z) H_{k+1,0}(z)$$

$$NS_k(z) = \sum_{n \geq 0} NS_{k,n} z^n = \sigma_k z H_{0,k}(z) H_{k-1,0}(z)$$

$$NQ_k(z) = \sum_{n \geq 0} NQ_{k,n} z^n = \kappa_k z H_{0,k}(z) H_{k,0}(z).$$

On pose enfin $N_{k,n} = \sum_{\omega} N_{\omega,k,n}$ et l'on a :

$$N_k(z) = \sum_{n \geq 0} N_{k,n} z^n = H_{0,k}(z) H_{k,0}(z).$$

2. LES STRUCTURES CLASSIQUES : DENOMBREMENTS.

Comme nous l'avons déjà dit dans l'introduction, le théorème d'équivalence très général de la section précédente ouvre aux histoires de fichiers l'accès à la théorie classique des fractions continues. Les sources les plus riches de développement en fractions continues de séries entières sont

- les équations fonctionnelles
- les formules d'addition correspondant au théorème de Stieltjes-Rogers.

Ainsi reprenant un exemple dû à [Polya ; 1937], si Q_n désigne le nombre d'isomères de structure sans carbone asymétrique de l'alcool $C_n H_{2n+1} OH$, la série génératrice $Q(z) = \sum Q_n z^n$ vérifie l'équation fonctionnelle

$$Q(x) = 1 + xQ(x) Q(x^2);$$

il s'ensuit la fraction continue

$$Q(x) = \frac{1}{1 - \frac{x}{1 - \frac{x^2}{1 - \frac{x^4}{\dots}}}}$$

Nous utilisons ici les deux méthodes -équations fonctionnelles, formules d'addition- pour retrouver par l'algèbre les dénombrements d'histoires de fichiers et relatifs aux structures classiques de dictionnaires, listes linéaires, files de priorité et tables de symboles obtenus initialement par Françon et Viennot au moyen de correspondances géométriques.

Nous effectuons ensuite le calcul des séries génératrices d'histoires de hauteur bornées dans le cas de ces mêmes structures informatiques : il s'agit d'identifier les polynômes qui apparaissent en numérateur et dénominateur, et l'on voit apparaître les polynômes classiques de l'analyse : polynômes de Laguerre, Hermite, Meixner, Poisson-Charlier Cette coïncidence n'est pas pour surprendre, car l'on sait que ces polynômes sont historiquement apparus dans les réduites de fractions simples et ce n'est qu'ultérieurement que les définitions algébriques (récurrence à trois termes) ont été remplacées par des définitions liées à l'expression analytique de relations d'orthogonalité. Ces relations d'orthogonalité,

sous leur aspect algébrique, sont utilisées à la section suivante pour le calcul des coûts intégrés. Nous terminons en donnant toujours dans le cas des structures informatiques classiques, l'expression des séries génératrices d'histoires relatives à une population bornée.

2.1. Séries génératrices.

Soit $\Sigma_H(z)$ la série génératrice des histoires de type Σ . Le théorème 1 appliqué aux 5 structures informatiques de base décrites plus haut, montre pour Σ = dictionnaire ; file de priorité ; table de symbole ; liste linéaire ; pile ; les égalités :

$$\text{DICT}_H(z) = \frac{1}{1 - 1z - \frac{1^2 z^2}{1 - 3z - \frac{2^2 z^2}{1 - 5z - \frac{3^2 z^2}{\dots}}}}$$

$$\text{FP}_H(z) = \frac{1}{1 - \frac{1z^2}{1 - \frac{2z^2}{1 - \frac{3z^2}{\dots}}}}$$

$$\text{TS}_H(z) = \frac{1}{1 - 0.z - \frac{1z^2}{1 - 1z - \frac{2z^2}{1 - 2z - \frac{3z^2}{\dots}}}}$$

$$\text{LL}_H(z) = \frac{1}{1 - \frac{1^2 z^2}{1 - \frac{2^2 z^2}{1 - \frac{3^2 z^2}{\dots}}}}$$

$$\text{PILE}_H(z) = \frac{1}{1 - \frac{z}{1 - \frac{z}{1 - \frac{z}{\dots}}}}$$

Les séries correspondant aux dictionnaires et files de priorité sont un cas particulier de la fraction continue de Gauss qui exprime le rapport de deux séries hypergéométriques contiguës.

Définissons la série hypergéométrique par

$$\Omega(a, b; z) = 1 + ab \frac{z}{1!} + a(a+1)b(b+1) \frac{z^2}{2!} + a(a+1)(a+2)b(b+1)(b+2) \frac{z^3}{3!} + \dots,$$

elle vérifie trivialement l'équation fonctionnelle

$$\Omega(a, b; z) = \Omega(a, b+1; z) - az\Omega(a+1, b+1; z).$$

Ainsi, en divisant par $\Omega(a, b+1; z)$, l'on met cette relation sous la forme

$$\frac{\Omega(a, b+1; z)}{\Omega(a, b; z)} = \frac{1}{1 - az \frac{\Omega(a+1, b+1; z)}{\Omega(a, b+1; z)}},$$

ce qui par itération conduit à la fraction continue

$$\frac{\Omega(a, b+1, z)}{\Omega(a, b, z)} = \frac{1}{1 - \frac{az}{1 - \frac{(b+1)z}{1 - \frac{(a+1)z}{1 - \frac{(b+2)z}{\dots}}}}}$$

La substitution $a \rightarrow \frac{1}{2}$; $b \rightarrow 0$; $z \rightarrow 2z^2$ conduit à la fraction continue des files de priorité :

$$FP_H(z) = \frac{1}{1 - \frac{1z^2}{1 - \frac{2z^2}{1 - \frac{3z^2}{\dots}}}} = \Omega(\frac{1}{2}, 1; 2z^2).$$

(On a effectué la simplification $\Omega(a, 0, z) = 1$).

On tire de là le dénombrement explicite $FP_{H_{2n}} = 1.3.5 \dots (2n-1)$.

La substitution $a \rightarrow 1$; $b \rightarrow 0$, conduit au développement en fraction continue

$$\Omega(1, 1; z) = \sum_{n \geq 0} n! z^n = \frac{1}{1 - \frac{1z}{1 - \frac{2z}{1 - \frac{2z}{1 - \frac{3z}{\dots}}}}};$$

la contraction de cette fraction continue (cf. [Perron ; 1954] p. 40) conduit à l'égalité

$$\Omega(1, 1, z) = \frac{1}{1 - 1z - \frac{1^2 z^2}{1 - 3z - \frac{2^2 z^2}{\dots}}},$$

où la fraction continue (déjà connue d'Euler) est identique à celle qui est relative aux dictionnaires. En résumé :

THEOREME 2^A. Les séries génératrices d'histoires de dictionnaire et file de priorité s'expriment sous forme hypergéométrique par

$$FP_H(z) = \Omega(\frac{1}{2}, 0; 2z^2) \quad ; \quad DICT_H(z) = \Omega(1, 1, z),$$

ce qui équivaut aux dénombrements

$$FP_{H_{2n}} = 1.3.5...(2n-1) ; DICT_{H_n} = n! \cdot \square$$

L'identification des fractions continues relatives aux listes linéaires et tables de symboles s'effectue grâce au théorème d'addition de Stieltjes-Rogers (cf. [Wall ; 1947] p.203).

THEOREME D'ADDITION (Stieltjes-Rogers).

Soit f une série qui admet la formule d'addition $f(x+y) = \sum_{k \geq 0} \omega_k f(x) f(y)$

où $f_k(x)$ est de la forme

$$\frac{x^k}{k!} + f_{k,k+1} \frac{x^{k+1}}{(k+1)!} + f_{k,k+2} \frac{x^{k+2}}{(k+2)!} + \dots,$$

alors, la transformée de Laplace-Borel formelle ⁺ de f possède un développement en fraction continue

$$\int_0^\infty e^{-tz} f(tz) dt = \frac{1}{1 - \kappa_0 z - \frac{\lambda_1 z^2}{1 - \kappa_1 z - \frac{\lambda_2 z^2}{\dots}}}$$

où pour tout $j \geq 0$: $\lambda_{j+1} = \frac{\omega_{j+1}}{\omega_j}$ et $\kappa_j = f_{j,j+1} - f_{j-1,j}$.

L'application de ce théorème à la fonction $\sec z = \frac{1}{\cos z}$ est immédiate. En effet

$$\begin{aligned} \sec(x+y) &= \frac{1}{\cos(x+y)} = \frac{1}{\cos x \cos y - \sin x \sin y} = \frac{1}{\cos x \cos y} \frac{1}{1 - \tan x \tan y} \\ &= \sum_{k \geq 0} (\sec x \cdot \tan^k x) (\sec y \cdot \tan^k y) = \sum_{k \geq 0} (k!)^2 (\sec x \frac{\tan^k x}{k!}) (\sec y \frac{\tan^k y}{k!}). \end{aligned}$$

Ainsi, en notant E_{2n} le coefficient de $\frac{z^{2n}}{(2n)!}$ dans le développement de Taylor de $\sec z$ (les nombres E_{2n} sont appelés nombres d'Euler ou nombres sécants) obtient-on l'égalité :

$$\sum_{n \geq 0} E_{2n} z^{2n} = \frac{1}{1 - \frac{1^2 z^2}{1 - \frac{2^2 z^2}{\dots}}},$$

fraction continue qui est celle des listes linéaires.

⁺ La transformée de Laplace-Borel formelle de la série $f(z) = \sum f_n \frac{z^n}{n!}$ est par définition $\hat{f}(z) = \sum f_n z^n$. Par commodité nous utilisons l'expression analytique pour dénoter la transformation formelle.

Considérons maintenant la série $f(z) = e^{z-1-z}$ dont le développement de Taylor définit les nombres de Bell 2-associés, β_n (cf. [Comtet ; 1970]) :

$f(z) = \sum \beta_n \frac{z^n}{n!}$. La série $f(z)$ vérifie la formule d'addition :

$$f(x+y) = e^{e^x-1-x} e^{e^y-1-y} e^{(e^x-1)(e^y-1)} ;$$

soit en développant en série la dernière exponentielle :

$$f(x+y) = \sum_{k \geq 0} k! e^{e^x-1-x} \frac{(e^x-1)^k}{k!} e^{e^y-1-y} \frac{(e^y-1)^k}{k!} .$$

Ceci conduit à la fraction continue :

$$\sum_{n \geq 0} \beta_n z^n = \frac{1}{1 - 0z - \frac{1z^2}{1 - 1z - \frac{2z^2}{1 - 2z - \frac{3z^2}{\dots}}}} ,$$

qui est celle des tables de symboles.

THEOREME 2^B. Les séries génératrices exponentielles d'histoires de liste linéaire et tables de symboles ont les expressions

$$\sum_{n \geq 0} LL_{H_{2n}} \frac{z^{2n}}{(2n)!} = \sec z ; \quad \sum_{n \geq 0} TS_{H_n} \frac{z^n}{n!} = e^{e^z-1-z} ;$$

soit pour les dénombrements correspondants :

$$LL_{H_{2n}} = E_{2n} \quad \text{et} \quad TS_{H_n} = \beta_n ,$$

où E_{2n} est le $2n$ -ième nombre d'Euler et β_n est le n -ième nombre de Bell 2-associé. □

Les dénombrements relatifs aux dictionnaires, files de priorité et listes linéaires ont été obtenus de manière très différente par [Françon, Viennot ; 1979] et [Françon ; 1978] .

Le cas des histoires de pile -qu'on sait d'ailleurs en bijection naturelle avec les arbres- est trivial : la fraction continue relative à cette structure :

$$C(z) = \frac{1}{1 - \frac{z^2}{1 - \frac{z^2}{\dots}}} ,$$

vérifie l'équation fonctionnelle $C(z) = \frac{1}{1-z^2 C(z)}$, ce qui donne la valeur

$$C(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z^2} ; \text{ soit, comme il est attendu, l'égalité } \text{PILE}_{H_{2n}} = \frac{1}{n+1} \binom{2n}{n}.$$

Remarque : Le théorème 2^A peut aussi s'obtenir comme conséquence des formules d'additions relatives aux fonctions $\frac{1}{1-z}$ et $e^{z^2/2}$.

2.2. Histoires de hauteur bornée.

A chaque type de structure correspond une fraction continue qui exprime la série génératrice $H(z)$ des histoires. Les réduites de la fraction continue font apparaître deux familles de polynômes numérateurs et dénominateurs. Dans les cas que nous considérons, les polynômes dénominateurs ont des expressions plus compactes que les polynômes numérateurs qui s'en déduisent par convolution.

L'interprétation combinatoire montre pour la série $H_{0,k}(z)$ la propriété :

$$\text{val}(H_{0,k}(z)) = k,$$

qui traduit la nécessité d'au moins k étapes pour atteindre la hauteur k .

Reprenant l'expression de $H_{0,k}(z)$ obtenue à la section précédente

$$H_{0,k}(z) = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_k z^k} (Q_{k-1}(z) H(z) - P_{k-1}(z)),$$

on voit qu'ainsi :

$$\text{val}[Q_{k-1}(z) H(z) - P_{k-1}(z)] = 2k.$$

Comme, d'autre part, le degré de P_{k-1} vérifie :

$$\text{deg } P_{k-1}(z) \leq k-1,$$

on en déduit que le polynôme $P_{k-1}(z)$ coïncide avec les k premiers termes de la série $Q_{k-1}(z) H(z)$. Ainsi, avec les opérateurs d'effacement E_m définis sur $C[[z]]$ par

$$E_m(z^n) = 0 \text{ si } n > m ; E_m(z^n) = z^n \text{ si } n \leq m$$

et étendus par linéarité, l'on a (en décalant les indices)

$$P_k(z) = E_{k-1}(Q_k(z) \cdot H(z)),$$

propriété qui permet commodément de calculer les P_k à partir des Q_k .

Nous nous limitons donc dans la suite de cette section à la détermination des polynômes dénominateurs Q_k . Dans chacun des cas classiques, la récurrence sur ces

polynômes conduit à une équation aux dérivées partielles sur la série génératrice exponentielle. L'équation s'avère à variables séparables et se résout par les méthodes classiques.

Pour chaque famille $\{Q_k\}$, on considère la série génératrice

$$K(z, t) = \sum_{k \geq 0} Q_{k-1}(z) \frac{t^k}{k!},$$

ainsi que les polynômes "réciproques" et leur série génératrice

$$\bar{Q}_{k-1}(z) = z^k Q_{k-1}\left(\frac{1}{z}\right) \\ \bar{K}(z, t) = \sum_{k \geq 0} \bar{Q}_{k-1}(z) \frac{t^k}{k!} = K\left(\frac{1}{z}; tz\right).$$

Nous renvoyons par exemple à [Szegö ; 1939] et [Chihara ; 1978] pour la présentation des familles classiques de polynômes orthogonaux.

Dictionnaires et polynômes de Laguerre.

Le calcul des premières réduites montre les valeurs

$$H^{[0]}(z) = \frac{1}{1-z}; \quad H^{[1]}(z) = \frac{1-3z}{1-4z+2z^2}; \quad H^{[2]}(z) = \frac{1-8z+11z^2}{1-9z+18z^2-6z^3}.$$

Les polynômes dénominateurs vérifient la récurrence

$$Q_k(z) = (1-(2k+1)z) Q_{k-1}(z) - k^2 z^2 Q_{k-2}(z); \quad Q_{-1}(z) = 1; \quad Q_0(z) = 1-z,$$

ce qui se traduit pour la série génératrice K par l'équation

$$\frac{\partial K(z, t)}{\partial t} (1+tz)^2 = K(z, t) (1-z-tz^2).$$

De là, en intégrant

$$K(z, t) = c(z) - \frac{1}{2}(1+tz)^{-1} - \ln(1+tz).$$

Les conditions initiales montrent que la constante d'intégration $c(z)$ vaut $\frac{1}{2}$, et finalement

$$\ln K(z, t) = \frac{t}{1+tz} - \ln(1+tz) \\ K(z, t) = \frac{1}{1+tz} \exp \frac{1}{1+tz} \\ \bar{K}(z, t) = \frac{1}{1+t} \exp z \frac{t}{1+t}.$$

Les polynômes $\bar{Q}_k$ sont donc identiques aux polynômes de Laguerre, et l'on a l'expression close

$$(-1)^m \bar{Q}_{m-1}(z) = \sum_{0 \leq k \leq m} (-1)^k \binom{m}{k} \frac{m!}{k!} z^k.$$

Files de priorité et polynômes d'Hermite.

Les premières réduites ont pour valeur

$$H^{[0]}(z) = 1 ; H^{[1]}(z) = \frac{1}{1-z^2} ; H^{[2]}(z) = \frac{1-2z^2}{1-3z^2} ; H^{[3]}(z) = \frac{1-5z^2}{1-6z^2+3z^4} ;$$

la récurrence de base et l'équation de K sont respectivement

$$Q_k(z) = Q_{k-1}(z) - kz^2 Q_{k-2}(z) \\ \frac{\partial K(z,t)}{\partial t} = K(z,t)(1-z^2t).$$

Soit finalement :

$$K(z,t) = e^{t-z^2t^2/2} \quad \text{et} \quad \bar{K}(z,t) = e^{zt-t^2/2}.$$

Les polynômes $\bar{Q}_k$ sont ainsi identiques aux polynômes d'Hermite, et possèdent

la forme explicite :

$$\bar{Q}_{m-1}(z) = \sum_{0 \leq k \leq \frac{m}{2}} (-1)^k \frac{m!}{2^k k! (m-2k)!} z^{m-2k}.$$

Tables de symboles et polynômes de Poisson-Charlier.

Les premières réduites sont

$$H^{[0]}(z) = 1 ; H^{[1]}(z) = \frac{1-z}{1-z-z^2} ; H^{[2]}(z) = \frac{1-3z}{1-3z+z^2+2z^3} ,$$

et la récurrence de base sur les polynômes Q_k s'écrit

$$Q_k(z) = (1-kz) Q_{k-1}(z) - kz^2 Q_{k-2}(z),$$

d'où l'équation :

$$\frac{\partial K(z,t)}{\partial t} = K(z,t)(1-z^2t) + zt \frac{\partial K(z,t)}{\partial t}.$$

On en déduit les expressions :

$$K(z,t) = e^{-zt(1+zt)}^{(1+\frac{1}{2})}$$

$$\bar{K}(z,t) = e^{-t(1+t)(1+z)} = (1+t)e^{-t} (1+t)^z.$$

Le développement de $\bar{K}(z-1,t)$ fournit pour les polynômes $\bar{Q}$ l'expression

$$(-1)^m \bar{Q}_{m-1}(z-1) = \sum_{0 \leq k \leq m} (-1)^k \binom{m}{k} z(z-1)(z-2) \dots (z-k+1).$$

Ainsi à un décalage de la variable près, les polynômes $\bar{Q}$ sont identiques aux polynômes de Poisson-Charlier.

Listes linéaires et polynômes de Meixner.

Les premières réduites ont ici pour valeur

$$H^{[0]} = 1 ; H^{[1]} = \frac{1}{1-z^2} ; H^{[2]} = \frac{1-4z^2}{1-5z^2} ; H^{[3]} = \frac{1-13z^2}{1-14z^2+9z^4} .$$

La récurrence de base et l'équation donnant K sont

$$Q_k(z) = Q_{k-1}(z) - k^2 z^2 Q_{k-2}(z) .$$

Cette récurrence fournit lorsqu'on en multiplie les termes par $\frac{t^k}{k!}$, après sommation, l'équation

$$\frac{\partial K(z,t)}{\partial t} = (1-z^2 t) K(z,t) - z^2 t^2 \frac{\partial K(z,t)}{\partial t} ,$$

d'où en intégrant

$$K(z,t) = \frac{1}{\sqrt{1+z^2 t^2}} \exp\left(\frac{1}{2} \arctan zt\right) ;$$

$$\bar{K}(z,t) = \frac{1}{\sqrt{1+t^2}} \exp(z \arctan t) .$$

Ces polynômes $\bar{Q}_k(z)$ sont connus sous le nom de polynômes de Meixner et sont reliés aux polynômes de Mittag-Leffler (cf. [Flajolet ; 1978]).

THEOREME 2^C. Pour les quatre structures classiques - dictionnaires, files de priorité, listes linéaires et tables de symboles, la série génératrice des histoires de hauteur bornée est rationnelle. Les polynômes dénominateurs réciproques $\bar{Q}_k$ ont une série génératrice exponentielle de la forme

$$\bar{K}(z,t) = a(t) \exp z b(t) ,$$

où les fonctions a et b sont données par la table

	a(t)	b(t)
DICT	$\frac{1}{1+t}$	$\frac{t}{1+t}$
FP	$e^{-t^2/2}$	t
LL	$\frac{1}{\sqrt{1+t^2}}$	arc tg t
TS	$(1+t)e^{-t}$	ln(1+t)

• □

Le type de donnée pile conduit, comme on le voit d'après le résultats du chapitre I, à l'apparition des polynômes de Fibonacci (Tchebycheff).

2.3. Histoires sur une population bornée.

J. Françon [1978] a également posé le problème de l'analyse de coûts de suites d'opérations, lorsque les clefs sont prises dans un ensemble de référence fini. Ces histoires sont dites histoires relatives à une population bornée. Par la suite le symbole N désigne la taille de cet ensemble de référence. Nous résolvons ici le problème de la détermination des nombres d'histoires (ou de leurs séries génératrices) dans le cas des dictionnaires, files de priorité et liste linéaire.

Pour un type de donnée Σ , on désigne par $\Sigma H_n^{N/}$ le nombre d'histoires de type Σ lorsque la population est limitée à une taille N fixée.

Files de priorité sur population bornée.

Ce cas correspond à une fonction de possibilité

$$\text{pos}(A; k) = N-k ; \text{pos}(S; k) = 1 ; \text{pos}(Q; k) = 0,$$

et les opérations sur un fichier de taille supérieure à N sont impossibles.

La fonction de possibilité traduit le fait qu'une adjonction dans un fichier de taille k , ne peut porter que sur les $N-k$ éléments de la population de référence non compris dans le fichier. Ainsi, dans ce cas, a-t-on pour la série génératrice l'expression :

$$H^{N/}(z) = \frac{1}{1 - \frac{Nz^2}{1 - \frac{(N-1)z^2}{1 - \frac{(N-2)z^2}{\dots \frac{1}{1-z^2}}}}}$$

Le calcul des premières valeurs montre

$$H^{0/}(z) = 1 ; H^{1/}(z) = \frac{1}{1-z^2} ; H^{2/}(z) = \frac{1-z^2}{1-3z^2} ; H^{3/}(z) = \frac{1-3z^2}{1-6z^2+3z^4} .$$

Cette fraction rationnelle est obtenue par "renversement" de la N -ième réduite

de la fraction continue des files de priorité (sur ensemble de clefs infini).

Les renversées s'expriment, de manière générale, au moyen des mêmes familles de $\{P_k\}$, $\{Q_k\}$ que les fractions d'origine. On le vérifie ici en remarquant que

$$H^{N+1}/(z) = \frac{1}{1 - (N+1)z^2 H^N/(z)},$$

d'où, pour tout N , l'expression

$$H^N/(z) = \frac{Q_{N-1}(z)}{Q_N(z)} \quad \text{où} \quad Q_N(z) = Q_{N-1}(z) - Nz^2 Q_{N-2}(z).$$

Ainsi les polynômes qui apparaissent sont identiques aux variantes élémentaires des polynômes d'Hermite des files de priorité ordinaire ; leur série génératrice vaut

$$\sum_{N \geq 0} Q_{N-1}(z) \frac{t^N}{N!} = e^{t - z^2 t^2 / 2}.$$

Listes linéaires sur une population bornée.

Ici la fonction de possibilité vaut

$$\text{pos}(A; k) = N - k ; \text{pos}(S; k) = k ; \text{pos}(Q; k) = 0,$$

d'où l'expression

$$H^N/(z) = \frac{1}{1 - \frac{N z^2}{1 - \frac{(N-1)2z^2}{\dots}}}$$

$$1 - \frac{2(N-1)z^2}{1 - 1Nz^2}$$

Cette fraction est reliée à certaines fractions continues calculées par Stieltjes et Rogers. Le point de départ est la formule d'addition de $(\text{ch } x)^N$:

$$\begin{aligned} \text{ch}^N(x+y) &= (\text{ch } x \text{ch } y + \text{sh } x \text{sh } y)^N = \text{ch}^N x \text{ch}^N y \sum_k \binom{N}{k} \text{th}^k x \text{th}^{N-k} y \\ &= \sum_{0 \leq k \leq N} (1.2 \dots k) N(N-1) \dots (N-k+1) \text{ch}^N x \frac{\text{th}^k x}{k!} \text{ch}^N y \frac{\text{th}^{N-k} y}{(N-k)!}. \end{aligned}$$

La fraction correspondante est exactement celle des listes linéaires sur population bornée, de sorte qu'avec le théorème de Stieltjes-Rogers :

$$H^N/(z) = \int_0^\infty e^{-t} (\text{ch } t z)^N dt.$$

La série génératrice exponentielle des histoires vaut ainsi $\hat{H}^N/(z) = \text{ch}^N z$.

Le calcul de la transformée de Laplace conduit à la décomposition en éléments

simples de $H^{N/}(z)$. En effet en développant $\text{ch } x = \frac{e^x + e^{-x}}{2}$;

$$H^{N/}(z) = \frac{1}{2^N} \int_0^\infty \left(\sum_j \binom{N}{j} e^{zt(N-2j)} \right) e^{-t} dt = \frac{1}{2^N} \sum_j \frac{\binom{N}{j}}{1 - (N-2j)z}$$

puisque'on a l'égalité

$$\int_0^\infty e^{azt} e^{-t} dt = \frac{1}{1-az}.$$

Ceci montre pour le nombre d'histoires la valeur

$$H_n^{N/} = \frac{1}{2^N} \sum_j \binom{N}{j} (N-2j)^n.$$

Dictionnaires sur population bornée.

La fonction de possibilité correspondante est

$$\text{pos}(A; k) = N - k ; \text{pos}(S; k) = k ; \text{pos}(Q^+; k) = k ; \text{pos}(\bar{Q}; k) = (N - k).$$

La série génératrice des histoires admet donc le développement en fraction continue

$$H^{N/}(z) = \frac{1}{1 - Nz - \frac{N \cdot 1z^2}{1 - Nz - \frac{(N-1)2z^2}{\dots \frac{1 \cdot Nz^2}{1 - Nz}}}}.$$

Par rapport aux listes linéaires, cette fraction est obtenue en décalant de la constante N les coefficients correspondant aux possibilités d'interrogation.

Ces cas se traitent par le lemme suivant, dit lemme d'insertion des paliers.

LEMME. Etant donnés deux systèmes Σ et T dont les fonctions de possibilités Σ_{pos} et T_{pos} vérifient les relations

$$T_{\text{pos}}(A; k) = \Sigma_{\text{pos}}(A; k) ; T_{\text{pos}}(S; k) = \Sigma_{\text{pos}}(S; k) ; T_{\text{pos}}(Q; k) = \Sigma_{\text{pos}}(S; k) + \lambda$$

pour une certaine constante entière λ , alors les séries génératrices d'histoires vérifient la relation

$$T_H(z) = \frac{1}{1-\lambda z} \Sigma_H\left(\frac{z}{1-\lambda z}\right) ;$$

les séries génératrices exponentielles correspondantes $T_f(z)$ et $\Sigma_f(z)$ vérifient

$$T_f(z) = e^{\lambda z} \cdot \Sigma_f(z).$$

Preuve : Ces deux égalités sont la traduction au niveau des séries génératrices de la relation

$$T_{H_n} = \sum_{0 \leq k \leq n} \binom{n}{k} \sum H_k \lambda^{n-k},$$

laquelle représente la décomposition des histoires de type T selon le nombre de paliers de type λ . $\square$

L'application de ce lemme à la fraction des histoires de dictionnaire sur population bornée est immédiate à partir du résultat précédent ; on en déduit l'expression

$$H^{N/}(z) = \frac{2^{-N}}{1-Nz} \sum_j \frac{\binom{N}{j}}{1-(N-2j)\frac{z}{1-Nz}} = 2^{-N} \sum_j \frac{\binom{N}{j}}{1-2jz},$$

d'où : $H_n^{N/} = 2^{n-N} \sum_j \binom{N}{j} j^n.$

Notons également la valeur de la série génératrice exponentielle $\hat{H}^{N/}(z) = \left(\frac{e^{2z}+1}{2}\right)^N.$

THEOREME 2^D. Les séries génératrices des histoires relatives à une population de taille N ont pour expression

(i) dans le cas des files de priorité

$H^{N/}(z) = \frac{Q_{N-1}(z)}{Q_N(z)}$ où les polynômes Q sont les variantes élémentaires des polynômes d'Hermite de série génératrice $\sum_{N \geq 0} Q_{N-1}(z) \frac{t^N}{N!} = e^{t-z^2 t^2/2}.$

(ii) dans le cas des listes linéaires :

$H^{N/}(z) = \frac{1}{2^N} \sum_j \frac{\binom{N}{j}}{1-(N-2j)z}$, ce qui correspond à la série génératrice exponentielle $\sum H_n^{N/} \frac{z^n}{n!} = (\text{ch } z)^N$

(iii) dans le cas des dictionnaires :

$H^{N/}(z) = \frac{1}{2^N} \sum_j \frac{\binom{N}{j}}{1-2jx}$, ce qui correspond à la série génératrice exponentielle $\left(\frac{e^{2x}+1}{2}\right)^N.$ $\square$

Remarque : On peut également dans les deux derniers cas opérer directement sur la formule d'addition sous-jacente à la fraction continue en observant que si une fonction $f(x)$ vérifie la formule d'addition

$$f(x+y) = \sum_{k \geq 0} \omega_k f_k(x) f_k(y),$$

alors la fonction $e^{\lambda x} f(x)$ vérifie trivialement

$$e^{\lambda(x+y)} f_{(x+y)} = \sum_{k \geq 0} \omega_k (e^{\lambda x} f_k(x)) (e^{\lambda y} f_k(y)).$$

Nous verrons, au chapitre suivant, que les fonctions f_k qui apparaissent dans les formules d'addition ne sont autres que les séries génératrices exponentielles des histoires terminant à hauteur k (ce à un facteur multiplicatif $\alpha_0 \alpha_1 \dots \alpha_{k-1}$ près). De la sorte on obtient les expressions suivantes de ces séries génératrices :

- dans le cas des listes linéaires :

$$\sum_{n \geq 0} H_{0,k,n} \frac{z^n}{n!} = \binom{N}{k} \text{ch}^N x \text{th}^k x$$

- dans le cas des dictionnaires :

$$\sum_{n \geq 0} H_{0,k,n} \frac{z^n}{n!} = \binom{N}{k} \left(\frac{e^{2x} + 1}{2} \right)^N \text{th}^k x.$$

3. TRANSFORMATIONS INTEGRALES ASSOCIEES AUX STRUCTURES DE DONNEES.

On sait que se trouve associé à chaque série entière un produit scalaire étroitement lié au développement en fraction continue de la série. On montre notamment que les polynômes dénominateurs des réduites de la fraction sont orthogonaux vis-à-vis de ce produit.

Ce produit scalaire est défini algébriquement à partir des coefficients de la série de départ, mais l'on peut -c'est le problème des moments- en rechercher une représentation analytique sous la forme

$$\langle f, g \rangle = \int_I f(x) g(x) d\alpha(x), \text{ pour un certain intervalle } I$$

et une certaine mesure $d\alpha$.

Nous n'utilisons ici que les propriétés algébriques des relations d'orthogonalité.

Ces relations sont appliquées aux séries et aux polynômes associés aux types de données classiques - dictionnaire, file de priorité ... -. Dans chaque cas les polynômes ont une série génératrice de forme particulière

$$a(t) \exp z b(t).$$

Les polynômes orthogonaux de ce type constituent la classification de Meixner des polynômes orthogonaux.

On montre pour chaque structure Σ : dictionnaire, file de priorité..., le résultat suivant : il existe une transformation intégrale $\Sigma \mathcal{L}$ qui transforme la série génératrice $c(u)$ des coûts unitaires d'un algorithme, en la série exponentielle $\hat{K}(z)$ des coûts intégrés du même algorithme ; c'est-à-dire qu'on a l'égalité

$$\hat{K}(z) = \Sigma \mathcal{L}(c(u); z).$$

De surcroît la nature relativement simple des transformées permet le calcul explicite des coûts intégrés de bon nombre d'algorithmes pratiquement utilisés.

3.1. Relations d'orthogonalité ; classification de Meixner des polynômes orthogonaux.

Etant donnée la fraction continue de type J :

$$1 - x_0 z - \frac{1}{\alpha_0 \sigma_1 z^2} \frac{1}{1 - x_1 z - \frac{\alpha_1 \sigma_2 z^2}{1 - x_2 z - \frac{\alpha_2 \sigma_3 z^2}{\dots}}}$$

associée à un type de donnée[†]. Il lui correspond les deux familles de polynômes $\{P_k\}$ et $\{Q_k\}$, ainsi que le développement en série entière

$$H(z) = \sum_{n \geq 0} H_n z^n.$$

On définit sur l'ensemble des polynômes à coefficients complexes, une forme linéaire associée à la suite des nombres H_n (donc à la série, ou encore à la fraction), notée $\langle \cdot \rangle_H$ ou simplement $\langle \cdot \rangle$:

$$\langle x^n \rangle = H_n ; \langle ap(x) + bq(x) \rangle = a \langle p(x) \rangle + b \langle q(x) \rangle \text{ pour } a, b \in \mathbb{C} \text{ et } p, q \in \mathbb{C}[x].$$

Cette forme linéaire permet à son tour de définir une forme bilinéaire notée $\langle \cdot | \cdot \rangle_H$ ou simplement $\langle \cdot | \cdot \rangle$:

$$\langle p(x) | q(x) \rangle = \langle p(x) \cdot q(x) \rangle ,$$

forme qui est la composée de la forme linéaire $\langle \cdot \rangle$ et du produit d'algèbre de $\mathbb{C}[x]$.

PROPOSITION. Soient $\bar{Q}_k$ les polynômes réciproques des polynômes Q_k , c'est-à-dire $\bar{Q}_k(x) = x^{k+1} Q_k(\frac{1}{x})$. Les polynômes $\bar{Q}_k$ vérifient les relations d'orthogonalité :

$$\begin{aligned} \langle x^\ell | \bar{Q}_{k-1}(x) \rangle &= 0 \quad \text{pour } 0 \leq \ell < k \\ \langle x^k | \bar{Q}_{k-1}(x) \rangle &= \alpha_0 \alpha_1 \dots \alpha_{k-1} \sigma_0 \sigma_1 \dots \sigma_k. \end{aligned}$$

Preuve : Rappelons que le polynôme $\bar{Q}_{k-1}$ a degré k exactement.

On a vu à la section précédente l'égalité

[†] Les développements qui suivent valent clairement pour toute J-fraction.

$$H_{0,k}(z) = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_k z^k} [Q_{k-1}(z) H(z) - P_{k-1}(z)] .$$

D'après l'origine combinatoire du problème, on sait que

$$\text{val}(H_{0,k}) = k,$$

égalité déjà rencontrée qui exprime que toute histoire de 0 à k a longueur $\geq k$ (le résultat s'obtient aussi simplement par le calcul).

Posons $Q_{k-1}(x) = \sum_j Q_{k-1,j} x^j$ et $\bar{Q}_{k-1}(x) = \sum_j \bar{Q}_{k-1,j} x^j$. Le calcul du produit scalaire $\langle \bar{Q}_{k-1}(x) | x^k \rangle$ montre les égalités :

$$\begin{aligned} \langle \bar{Q}_{k-1}(x) | x^k \rangle &= \sum_j H_{k,j} \bar{Q}_{k-1,j} = \sum_j H_{k,j} Q_{k-1,k-j} = \text{coeff}(x^{k+l}; Q_{k-1}(x)H(x)) \\ &= \sigma_1 \sigma_2 \dots \sigma_k \text{coeff}(x^k; H_{0,k}(x)). \end{aligned}$$

En vertu de la remarque ci-dessus concernant $\text{val}(H_{0,k}(x))$, et de la valeur $H_{0,k,k} = \alpha_0 \alpha_1 \dots \alpha_{k-1}$, on déduit immédiatement l'énoncé de la proposition. $\square$
Le raisonnement montre encore :

PROPOSITION. *Le nombre d'histoires finissant à hauteur k et ayant longueur n vaut*

$$H_{0,k,n} = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_k} \langle x^n | \bar{Q}_{k-1}(x) \rangle . \quad \square$$

De manière analogue :

PROPOSITION. *Le nombre d'histoires commençant à hauteur k, finissant à hauteur l, et ayant longueur n vaut*

$$H_{k,l,n} = \frac{1}{\alpha_0 \alpha_1 \dots \alpha_{k-1} \sigma_1 \sigma_2 \dots \sigma_l} \langle \bar{Q}_{k-1}(x) \bar{Q}_{l-1}(x) | x^n \rangle . \quad \square$$

3.2. Classification de Meixner des polynômes orthogonaux ; types de données classiques.

Meixner a montré ([Meixner ; 1934], [Chihara ; 1978]) qu'il existe seulement cinq familles de polynômes orthogonaux dont la série génératrice exponentielle est de la forme $a(t) \exp x b(t)$. Nous avons rencontré jusqu'ici les familles de Laguerre, Hermite, Charlier et Meixner (dits de premier type). La combinaison des relations d'orthogonalité générales et des propriétés particulières des poly-

nômes de Meixner permet de démontrer :

THEOREME 3^A. Pour des suites de références $\omega_{k,\ell}$ dépendant du type de donnée, les séries génératrices triples d'histoires :

$$\sum H_{k,\ell,n} \omega_{k,\ell} u^k v^\ell \frac{z^n}{n!} = \Xi(u,v,z)$$

ont des expressions données par le tableau suivant :

	$\Xi(u,v,z)$	$\omega_{k,\ell}$
DICT	$\frac{1}{1-z(1+u)(1+v)-uv}$	1
FP	$e^{z^2/2} e^{zu+uv+vz}$	$\frac{1}{\ell!}$
LL	$\frac{1}{(1-uv)\cos z - (u+v)\sin z}$	1
TS	$e^{e^z(1+u)(1+v)-z-u-v-1}$	$\frac{1}{\ell!}$

Preuve : On considère les quatre structures classiques de l'énoncé pour lesquelles les polynômes $\bar{Q}_k$ ont une série génératrice de la forme

$$K(x,t) = \sum \bar{Q}_{k-1}(x) \frac{t^k}{k!} = a(t) \exp(x b(t)).$$

On calcule d'abord les séries génératrices

$$\Psi(v,z) = \Xi(0,v,z) = \sum H_{0,k,n} \omega_{0,k} u^k \frac{z^n}{n!}.$$

Pour cela on évalue de deux manières différentes le produit scalaire

$$\langle K(x,s) | K(x,t) \rangle \quad \text{produit portant sur la variable } x.$$

Formellement, il vient :

(i) en utilisant les relations d'orthogonalité :

$$\begin{aligned} \langle K(x,s) | K(x,t) \rangle &= \left\langle \sum_i \bar{Q}_{i-1}(x) \frac{s^i}{i!} \mid \sum_j \bar{Q}_{j-1}(x) \frac{t^j}{j!} \right\rangle = \sum_k \langle \bar{Q}_{k-1}(x) | \bar{Q}_{k-1}(x) \rangle \frac{s^k t^k}{(k!)^2} \\ &= \sum_k \frac{\alpha_0 \alpha_1 \dots \alpha_{k-1} \sigma_1 \sigma_2 \dots \sigma_k}{(k!)^2} s^k t^k = \Omega(st) \end{aligned}$$

$$\text{avec } \Omega(y) = \sum \frac{\alpha_0 \dots \alpha_{k-1} \sigma_1 \dots \sigma_k}{(k!)^2} y^k.$$

(ii) en utilisant la forme de type Meixner de $K(x,t)$:

$$\begin{aligned} \langle K(x,s) | K(x,t) \rangle &= \langle K(x,s) | a(t) \exp(x b(t)) \rangle = \langle K(x,s) | a(t) \sum_n \frac{x^n b(t)^n}{n!} \rangle \\ &= a(t) \langle \sum_k Q_{k-1}(x) \frac{s^k}{k!} | \sum_n \frac{x^n b(t)^n}{n!} \rangle \\ &= a(t) \sum_{k,n} \langle Q_{k-1}(x) | x^n \rangle \frac{s^k}{k!} \frac{b(t)^n}{n!} \\ &= a(t) \sum_{k,n} H_{0,k,n} \frac{1 \cdots \sigma_k}{k!} s^k \frac{b(t)^n}{n!} . \end{aligned}$$

En comparant ces deux expressions, l'on obtient une certaine série génératrice des $H_{0,k,n}$ sous la forme :

$$a(t) \sum_{k,n} H_{0,k,n} \frac{1 \cdots \sigma_k}{k!} s^k \frac{b(t)^n}{n!} = \sum_n \frac{\alpha_0 \alpha_1 \cdots \alpha_{k-1} \sigma_1 \cdots \sigma_k}{(k!)^2} s^k t^k .$$

Dans le cas de la classification de Meixner, la valuation de la série $b(t)$ est égale à 1 ; de la sorte, on peut définir un inverse de b vis-à-vis de la composition, soit β , tel que

$$\beta(b(t)) = t ,$$

et l'on a, en posant $b(t) = z$, $t = \beta(z)$:

$$\sum_{k,n} H_{0,k,n} \frac{1 \cdots \sigma_k}{k!} s^k \frac{z^n}{n!} = \frac{1}{a(\beta(z))} \Omega(s \beta(z)) .$$

Dans les cas que nous considérons, $\Omega(y)$ vaut soit $\frac{1}{1-y}$ (dictionnaires, listes linéaires) soit e^y (files de priorité, tables de symbole). Par ailleurs, la série $\beta(z)$ possède dans chaque cas une expression explicite, donnée par la table suivante

	DICT	FP	LL	TS
$\beta(z)$	$\frac{z}{1-z}$	z	$\operatorname{tg} z$	$e^z - 1$

Ainsi :

PROPOSITION. Les séries génératrices des histoires finissant à niveau k et de longueur n

$$\Xi(o, v, z) = \sum_{\ell, n \geq 0} H_{o, \ell, n} \omega_{o, \ell} \frac{v^\ell}{\ell!} \frac{z^n}{n!}$$

ont des expressions données par la table suivante

	$\Xi(0,v,z)$
DICT	$\frac{1}{1-z-vz}$
FP	$e^{z^2/2+zv}$
LL	$\frac{1}{\cos z - v \sin z}$
TS	$e^{e^z(1+v)-z-v-1}$

. □

Ce calcul se généralise à la détermination de la série triple $\Xi(u,v,z)$. On part de la relation

$$H_{k,\ell,n} = \frac{1}{\alpha_0 \dots \alpha_{k-1} \sigma_1 \dots \sigma_k} \langle \bar{Q}_{k-1}(x) \bar{Q}_{\ell-1}(x) x^n \rangle,$$

et l'on exprime de deux manières différentes la forme triple

$$\langle K(x,t) K(x,u) K(x,v) \rangle.$$

En développant la forme de Meixner de $K(x,t)$, et en utilisant la définition de $K(x,u)$ et $K(x,v)$, on obtient :

$$\langle K(x,t) K(x,u) K(x,v) \rangle = a(t) \sum \langle Q_{k-1}(x) Q_{\ell-1}(x) x^n \rangle \frac{u^k}{k!} \frac{v^\ell}{\ell!} \frac{b(t)^n}{n!}.$$

En développant par contre les formes de Meixner de $K(x,t)$, $K(x,u)$ et $K(x,v)$:

$$\langle K(x,t) K(x,u) K(x,v) \rangle = a(u) a(v) a(w) \langle e^{x(b(t)+b(u)+b(v))} \rangle.$$

Ainsi Ξ possède-t-il la forme

$$\Xi(u,v,z) = \frac{a(u) a(v)}{a(R(z+b(u)+b(v)))};$$

ceci conduit après calcul simple à l'énoncé du théorème. □

3.3. Formes intégrales des séries génératrices de coûts intégrés.

Les résultats de cette section sont résumés par les transformations intégrales qui relient séries génératrices de coûts unitaires[†] et séries génératrices de coûts intégrés.

†

Nous nous plaçons pour l'énoncé du résultat dans l'hypothèse simplificatrice où les coûts d'opérations sont indépendants de la nature des opérations. Nous verrons par la suite qu'on peut soit se ramener à ce cas, soit disposer, pour chaque opération, d'une transformation intégrale particulière.

THEOREME 3^B. Soit $\{C_k\}_{k \geq 0}$ une famille de coûts unitaires d'opérations : C_k représente le coût d'une opération effectuée à niveau k . Soit $C(u)$ la série génératrice des coûts unitaires $C(u) = \sum_k C_k \omega_k u^k$ où $\{\omega_k\}$ est une suite de référence qui dépend de la structure ($\omega_k = 1$ ou $\omega_k = \frac{1}{k!}$). La série exponentielle des coûts intégrés $\hat{K}(z) = \sum_n K_n \frac{z^{n+1}}{(n+1)!}$ est reliée à la série des coûts unitaires $C(u)$ par une transformation intégrale

$$\hat{K}(z) = \mathcal{L}(C(u) ; z),$$

où la transformation $\mathcal{L}$ est donnée par la table suivante

	$\mathcal{L}$	ω_k
DICT	$\frac{2}{2-z} \int_0^{\left(\frac{z}{2-z}\right)^2} \frac{C(u) du}{\sqrt{(1-u)\left(\left(\frac{z}{2-z}\right)^2 - u\right)}}$	1
FP	$2 e^{z^2/2} \int_0^{z^2/4} e^{-u} C(u) \sqrt{z^2 - 4u} du$	$\frac{1}{k!}$
LL	$\frac{2}{\cos z} \int_0^{\operatorname{tg} \frac{z}{2}} C(u) \frac{du}{\sqrt{(1-u)^2 \operatorname{tg}^2 z - 4u}}$	1
TS	$2 e^{e^z - z - 1} \int_0^{(e^{z/2} - 1)^2} e^{-u} C(u) \frac{du}{\sqrt{(e^z + 1 - u)^2 - 4uz}}$	$\frac{1}{k!}$

Preuve : On part de la définition des coûts intégrés,

$$K_n = \sum_{k \geq 0} C_k N_{k,n} = \sum_{k,j \geq 0} C_k H_{0,k,j} H_{k,0,n-j}.$$

Au niveau des séries génératrices :

$$\begin{aligned} K(z) &= \sum_{n \geq 0} K_n z^n = \sum_{k \geq 0} C_k H_{0,k}(z) H_{k,0}(z) \\ &= \sum_{k \geq 0} C_k \frac{\sigma_1 \sigma_2 \dots \sigma_k}{\alpha_0 \alpha_1 \dots \alpha_{k-1}} (H_{0,k}(z))^2. \end{aligned}$$

On a vu plus haut qu'une certaine série génératrice exponentielle par rapport à n des $H_{0,k,n}$ s'exprimait sous la forme

$$\sum_{k,n \geq 0} H_{0,k,n} \omega_{0,k} s^k \frac{z^n}{n!} = \frac{1}{a(\beta(z))} \Omega(s \beta(z)).$$

Ainsi la dépendance en s est elle très simple, et en utilisant la définition de Ω , l'on a :

$$\hat{H}_{0,k}(z) = \sum_{n \geq 0} H_{0,k,n} \frac{z^n}{n!} = \frac{\alpha_0 \alpha_1 \dots \alpha_{k-1}}{k!} \frac{\beta(z)^k}{a(\beta(z))}.$$

La série génératrice exponentielle $\hat{H}_{0,k}$ possède donc une forme particulière, alors que la série ordinaire (de rayon de convergence nul) ne se ramène à aucune fonction classique. On recherche donc une expression pour la série génératrice exponentielle des K_n , en l'occurrence la série

$$\hat{K}(z) = \sum K_n \frac{z^{n+1}}{(n+1)!}.$$

La transformation de Laplace-Borel fait passer des séries génératrices exponentielles aux séries génératrices ordinaires ; l'on applique ainsi une transformation de Laplace inverse à la définition de $K(z)$, en utilisant les expressions des $\hat{H}_{0,k}(z)$ (l'usage que nous faisons de la transformation de Laplace est purement formel, et ce n'est que par commodité que nous utilisons la notation intégrale usuelle).

LEMME. Soient $\hat{A}(x) = \sum_{n \geq 0} a_n \frac{x^n}{n!}$ et $\hat{B}(x) = \sum_{n \geq 0} b_n \frac{x^n}{n!}$ deux séries génératrices exponentielles ; le produit de convolution formel

$$\hat{P}(x) = (\hat{A} * \hat{B})(x) \stackrel{\text{def}}{=} \sum_{n \geq 0} \left(\sum_i a_i b_{n-i} \right) \frac{x^{n+1}}{(n+1)!}$$

s'exprime par l'intégrale

$$\hat{P}(x) = \int_0^x A(x-\tau) B(\tau) d\tau.$$

Preuve : Il s'agit du lemme classique qui exprime que la transformée de Laplace d'un produit de convolution est le produit de Cauchy des transformées des facteurs. La propriété est algébrique, comme le montre la preuve. On part de l'intégrale :

$$I = \int_0^x \hat{A}(x-\tau) B(\tau) d\tau = \sum_{n,m \geq 0} \int_0^x \frac{a_n b_m}{n! m!} (x-\tau)^n \tau^m d\tau$$

$$I = \sum_{n,m,k \geq 0} \int_0^x (-1)^k \frac{a_n b_m}{n!m!} \binom{n}{k} x^{n-k} \tau^{m+k} d\tau = \sum_{n,m \geq 0} \frac{a_n b_m}{n!m!} x^{n+m+1} \sum_k \frac{(-1)^k}{m+k+1} \binom{n}{k}.$$

Or le calcul des différences de la fonction $\frac{1}{x+1}$ montre l'identité

$$\Delta^n \frac{1}{1+x} = \sum_k (-1)^{n-k} \frac{\binom{n}{k}}{x+k+1} = \frac{(-1)^n n!}{(x+1)(x+2)\dots(x+n+1)}.$$

Le remplacement de x par m dans l'égalité donnant I conduit au résultat :

$$I = \sum_{n,m \geq 0} a_n b_m \frac{t^{n+m+1}}{n+m+1!} = (\hat{A} * \hat{B})(x). \quad \square$$

L'application de ce lemme à $\hat{K}(z)$ qui vaut

$$\hat{K}(z) = \sum_{k \geq 0} c_k \frac{\sigma_1 \sigma_2 \dots \sigma_k}{\alpha_0 \alpha_1 \dots \alpha_{k-1}} (\hat{H}_{0,k} * \hat{H}_{0,k})(z)$$

conduit, étant donnée l'expression des $\hat{H}_{0,k}$ à la suite d'égalités :

$$\begin{aligned} \hat{K}(z) &= \sum_{k \geq 0} c_k \frac{\alpha_0 \dots \alpha_{k-1} \sigma_1 \dots \sigma_k}{(k!)^2} \left(\frac{\beta(z)^k}{a(\beta(z))} * \frac{\beta(z)^k}{a(\beta(z))} \right) \\ &= \sum_{k \geq 0} c_k \omega_k \int_0^z \frac{\beta(\tau)^k}{a(\beta(\tau))} \cdot \frac{\beta(z-\tau)}{a(\beta(z-\tau))} d\tau = \int_0^z C[\beta(\tau)\beta(z-\tau)] \frac{d\tau}{a(\beta(\tau)) \cdot a(\beta(z-\tau))} \end{aligned}$$

L'étape suivante consiste, pour chaque structure, à transformer cette intégrale en une forme plus facilement calculable. Ceci s'obtient par le changement de variable

$$u = \beta(\tau) \beta(z-\tau)$$

dans l'intégrale donnant $\hat{K}(z)$ écrite sous la forme

$$\hat{K}(z) = 2 \int_0^{z/2} C(\beta(\tau) \beta(z-\tau)) \frac{d\tau}{a(\beta(\tau)) a(\beta(z-\tau))}.$$

Le calcul utilise les expressions de a et β du paragraphe précédent. Nous commençons par le cas le plus simple des files de priorité.

a. Files de priorité : On part de l'expression

$$\hat{K}(z) = 2 \int_0^{z/2} C(\tau(z-\tau)) e^{-\tau^2/2} e^{-(z-\tau)^2/2} d\tau;$$

le changement de variable $u = \tau(z-\tau)$ équivaut à

$$\tau = \frac{z}{2} - \frac{\sqrt{z^2 - 4u}}{2} \quad d\tau = \frac{\sqrt{z^2 - 4u}}{2} du,$$

d'où finalement l'expression

$$\hat{K}(z) = 2e^{z^2/2} \int_0^{z^2/4} e^{-u} C(u) \sqrt{z^2 - 4u} du.$$

b. Dictionnaires : On part encore de la forme intégrale de convolution :

$$\hat{K}(z) = 2 \int_0^{z/2} C\left(\frac{\tau(z-\tau)}{(1-\tau)(1-z+\tau)}\right) \frac{d\tau}{(1-\tau)(1-z+\tau)} ;$$

$$\frac{\tau(z-\tau)}{(1-\tau)(1-z+\tau)} = u \leftrightarrow (1-u)\tau^2 = z(1-u)\tau + u(1-z) = 0.$$

Les calculs se conduisent simplement si l'on remarque que les deux racines de cette équation sont τ et $z-\tau$: les quantités intervenant dans le changement de variable s'expriment comme fonctions symétriques élémentaires de ces racines.

On obtient :

$$(1-\tau)(1-z+\tau) = \frac{1-z}{1-u}, \quad \frac{du}{d\tau} = 2 \frac{1-u}{1-z} \sqrt{\Delta}$$

avec $\Delta = z^2 - u(2-z)^2$. Soit toutes réductions faites

$$\hat{K}(z) = \frac{2}{2-z} \int_0^t \frac{C(u) du}{\sqrt{(1-u)(t-u)}} \text{ avec } t = \left(\frac{z}{2-z}\right)^2.$$

c. Listes linéaires :

$$\hat{K}(z) = 2 \int_0^{z/2} C(\operatorname{tg} \tau \operatorname{tg}(z-\tau)) \frac{d\tau}{\cos \tau \cos(z-\tau)} ;$$

dans le changement de variable $u = \operatorname{tg} \tau \operatorname{tg}(z-\tau)$, $\operatorname{tg} \tau$ et $\operatorname{tg}(z-\tau)$ sont les racines en θ de l'équation

$$\theta^2 - \theta \operatorname{tg} z(1-u) + u = 0.$$

Avec $\Delta = (1-u)^2 \operatorname{tg}^2 z - 4u$ discriminant, il vient

$$\frac{du}{d\tau} = \sqrt{\Delta}(1-u) \text{ et } \frac{1}{\cos \tau \cos(z-\tau)} = \frac{1-u}{\cos z}.$$

Ainsi :

$$\hat{K}(z) = \frac{2}{\cos z} \int_0^{\operatorname{tg}^2 \frac{z}{2}} C(u) \frac{du}{\sqrt{(1-u)^2 \operatorname{tg}^2 z - 4u}}.$$

d. Tables de symboles.

$$\hat{K}(z) = 2 \int_0^{z/2} e^{e^\tau - \tau - 1} e^{e^{z-\tau} - z + \tau - 1} C((e^\tau - 1)(e^{z-\tau} - 1)) d\tau.$$

Posons d'abord $v = e^\tau + e^{z-\tau}$; e^τ et $e^{z-\tau}$ sont les racines en y de l'équation

$$y^2 - vy + e^\tau = 0 \rightarrow e^\tau = \frac{v - \sqrt{v^2 - 4e^\tau}}{2}.$$

et l'on a $\frac{dv}{d\tau} = (e^\tau - e^{z-\tau}) = -\frac{1}{2} \sqrt{v^2 - 4e^\tau},$

$$\hat{K}(z) = -2 e^{-z-2} \int_{1+e^z}^{2e^{z/2}} e^v C(e^z - v + 1) \frac{dv}{\sqrt{v^2 - 4e^z}}.$$

On effectue la translation $u = e^z - v + 1$ et il vient finalement

$$\hat{K}(z) = 2 e^{z-z-1} \int_0^{(e^{z/2}-1)^2} C(u) \frac{e^{-u} du}{\sqrt{(e^z+1-u)^2 - 4e^z}}. \quad \square$$

3.4. Prolongements.

(i) Certaines des intégrales apparues dans l'énoncé du théorème sont en fait des intégrales de convolution :

$${}^{\text{DICT}}\hat{K}(z) = \frac{2}{2-z} \left| \frac{C(s)}{\sqrt{1-s}} * \frac{1}{\sqrt{(\frac{z}{2-z})^2 - s}} \right|_{s=(\frac{z}{2-z})^2};$$

$${}^{\text{FP}}\hat{K}(z) = 4e^{z^2/4} \left| C(s) * e^s \sqrt{s} \right|_{s=\frac{z^2}{4}}.$$

Dans le cas des files de priorité, l'expression se prête à l'évaluation par une nouvelle transformation de Laplace-Borel :

$${}^{\text{FP}}\hat{K}(z) = \left| 4e^s (C(s) * e^s \sqrt{s}) \right|_{s=\frac{z^2}{4}}.$$

On pose donc $s = \frac{z^2}{4}$ et l'on applique la transformation

$$\mathcal{B}[f(s); t] = \int_0^\infty e^{-st} f(st) dt.$$

$\hat{K}(z)$ vaut $\sum_{n \geq 0} K_{2n} s^{\frac{n+1}{2}} \frac{2^{2n+1}}{(2n+1)!}$ et se transforme par $\mathcal{B}$ en $2\Gamma(\frac{1}{2}) t^{\frac{1}{2}} \sum_{n \geq 0} K_{2n} \frac{t^n}{n!}$.

D'autre part, le produit de convolution se transforme en produit de Cauchy, et l'on obtient avec $\bar{C}$ série génératrice ordinaire des coûts unitaires,

$$\bar{C}(x) = \sum C_k x^k :$$

$$\sum K_{2n} \frac{t^n}{n!} = \frac{1}{\sqrt{1-2t}} \frac{1}{1-t} \bar{C}\left(\frac{t}{1-t}\right).$$

THEOREME 3^c. La série génératrice des coûts unitaires $\bar{C}(x)$ et la série génératrice des coûts intégrés définie par $\bar{K}(t) = \sum K_{2n} \frac{t^n}{n!}$ sont liées par la relation

$$\bar{K}(t) = \frac{1}{\sqrt{1-2t}} \frac{1}{1-t} \bar{C}\left(\frac{t}{1-t}\right).$$

(ii) Les calculs du paragraphe 3.3 reposent sur le fait que les séries exponentielles $\sum_{n \geq 0} H_{0,k,n} \frac{z^n}{n!}$ sont proportionnelles aux puissances d'une même fonction :

$$\sum_{n \geq 0} H_{0,k,n} \frac{z^n}{n!} = \lambda_k \alpha(z) [\beta(z)]^k \quad \lambda_k \in \mathbb{C}.$$

C'est encore le cas des dictionnaires et des listes linéaires sur population bornée, ce bien que les polynômes associés ne fassent pas partie de la classification de Meixner. Les pires vérifient une propriété analogue, et l'on a :

THEOREME 3^D. (i) Pour le type de donnée pile, la série génératrice des

coûts unitaires $C(u) = \sum_{k \geq 0} c_k u^k$ et la série génératrice des coûts intégrés $K(z) = \sum_{n \geq 0} K_n z^{2n}$ vérifient la relation

$$K(z) = C(z B^2(z)) \quad \text{avec} \quad B(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

(ii) Pour les types de données liste linéaire et dictionnaires sur popula-

tion de taille N , la série binômiale des coûts unitaires $C(u) = \sum_{k \geq 0} c_k \binom{N}{k} u^k$, et la série exponentielle des coûts intégrés $\hat{K}(z) = \sum_{n \geq 0} K_n \frac{z^{n+1}}{n+1}$ sont liées par la relation :

$$\begin{aligned} \hat{K}(z) &= ch^N z \mathcal{L}^{N/} [C(u); z] \quad \text{pour les listes linéaires} \\ \hat{K}(z) &= \left(\frac{e^{2x} + 1}{2} \right)^N \mathcal{L}^{N/} [C(u); z] \quad \text{pour les dictionnaires,} \end{aligned}$$

où la transformation $\mathcal{L}^{N/}$ est définie par

$$\mathcal{L}^{N/} [C(u); z] = 2 \int_0^{\text{th}^2 z / 2} \frac{C(u)}{(1+u)^N (1-u)} \frac{du}{\sqrt{(1+u)^2 \text{th}^2 z - 4u}} \quad \square$$

(iii) Il existe pour chaque type d'opération (sur structure non bornée)

$\omega \in \{A, S, Q\}$, une transformation intégrale $\mathcal{L}^\omega$ qui fait passer des séries génératrices de coûts unitaires de l'opération ω aux coûts intégrés correspondants.

Prenons par exemple $\omega = A$ (adjonction) ; avec CA_k coût d'une adjonction à niveau k , et KA_n le coût intégré des adjonctions dans l'ensemble des histoires de longueur n :

$$KA_{n+1} = \sum_{k, j \geq 0} CA_k \alpha_k H_{0,k,j} H_{k,0,n-j},$$

soit au niveau des séries exponentielles :

$$KA(z) = \sum_{n+1} KA_{n+1} \frac{z^{n+1}}{(n+1)!} = \sum_{k \geq 0} CA_k \alpha_k \hat{H}_{0,k}(z) * \hat{H}_{k+1,0}(z).$$

Les séries $\hat{H}_{0,k}$ et $\hat{H}_{k+1,0}$ s'expriment encore comme puissances d'une même série, et il s'introduit ainsi une expression sous forme d'intégrale de type convolution très analogue aux précédentes

$$\begin{aligned} KA(z) &= \sum_{k \geq 0} CA_k \frac{\alpha_0 \dots \alpha_k}{k!} \frac{\sigma_1 \dots \sigma_{k+1}}{k+1!} \left(\frac{\beta(z)^k}{a(\beta(z))} \frac{\beta(z)^{k+1}}{a(\beta(z))} \right) \\ &= \int_0^z CA(\beta(\tau)) \cdot \beta(z-\tau) \frac{\beta(\tau)}{a(\beta(\tau)) \cdot a(\beta(z-\tau))} d\tau. \end{aligned}$$

Les mêmes changements de variable s'appliquent encore, et conduisent dans chaque cas à une transformation particulière à l'opération considérée qui n'est qu'une variante simple de la transformation associée à l'ensemble des opérations.

(iii) On peut calculer les coûts intégrés d'algorithmes, en évaluant, comme il est dit en (ii), les coûts intégrés de chacune des opérations. On peut aussi utiliser les relations simples entre $\{N_{k,n}\}$ et les $\{N_{\omega,k,n}\}$. Ainsi pour une file de priorité, l'on a :

$$NS_{k,n} = NA_{k+1,n} \quad NS_{k,n} + NA_{k,n} = N_{k,n},$$

relations dont l'interprétation est immédiate. $\square$

(v) La même méthode d'analyse développée dans ce chapitre s'applique à tous les types de structures à fonction de possibilités de la forme

$$\begin{aligned} \text{pos}(Q;k) &= \alpha k + \beta \\ \text{pos}(A;k-1) \cdot \text{pos}(S;k) &= k(\gamma k + \delta) \end{aligned} \quad \alpha, \beta, \gamma, \delta \in \mathbb{R};$$

pour celles-ci en effet, les polynômes associés se ramènent à la classification de Meixner.

C'est ainsi essentiellement à la linéarité des fonctions de possibilités rencontrées dans les applications informatiques, qu'est due l'existence de transformation intégrales simples exprimant les séries génératrices de coûts intégrés.

4. APPLICATIONS

Il est clair à ce point qu'on dispose d'un mode de calcul effectif des coûts intégrés de toutes les structures stationnaires dont les coûts unitaires possèdent des séries génératrices de forme simple. Or, à l'exception des files binomiales, les coûts unitaires des structures usuelles sont des combinaisons linéaires des fonctions de k suivantes :

$$1, k, k^2; \frac{1}{k}; \frac{1}{k+1}; H_k = 1 + \frac{1}{2} + \dots + \frac{1}{k},$$

dont les séries génératrices ont les formes simples :

$$\frac{1}{1-x}, \frac{1}{(1-x)^2}, \frac{2-x}{(1-x)^3}; \ln \frac{1}{1-x}; \frac{1}{x} \ln \frac{1}{1-x} - x; \frac{1}{1-x} \ln \frac{1}{1-x}.$$

Les calculs, parfois fastidieux, se ramènent essentiellement à des déterminations de primitives de fonctions élémentaires -produits de logarithmes et de fonctions algébriques- puis aux développements en séries de Taylor. Pour ne pas alourdir la présentation ; nous nous contenterons de reproduire les tables tirées de [Flajolet, Françon, Vuillemin ; 1979 a et b].

Par ailleurs combinant ces méthodes avec des techniques voisines de celles de Delange et du chapitre III de ce travail, L. Chéno [1979] a réalisé l'analyse des files binomiales pour lesquelles il a obtenu l'expression des coûts intégrés

$$\frac{K_{2n}}{1.3.5 \dots (2n-1)} = n \cdot \log_2 n + n \cdot I(n) + o(n),$$

où I est une fonction oscillante de n qui vérifie $I(n) = I(2n)$.

DATA TYPE	DATA ORGANISATION	CA_k	CS_k	CQ_k^+	CQ_k^-
Dictionary	Sorted list	$(k+2)/2$	$(k+1)/2$	$(k+1)/2$	$(k+2)/2$
	Unsorted list	0	$(k+1)/2$	$(k+1)/2$	k
	Binary search tree	$2(H_{k+1}-1)$	$2(1+\frac{1}{k})H_k-3$	$2(1+\frac{1}{k})H_k-3$	$2(H_{k+1}-1)$
Priority Queue	Sorted list	$(k+2)/2$	0		
	Binary search tree	$2(H_{k+1}-1)$	0		
	Binary Tournament	$H_{k+1}-1/2$	$2(H_k-2+1/k)$		
	Pagoda	$2(1-1/(k+1))$	$2(H_k-2+1/k)$		
	Binomial queues	$1+v(k)-v(k+1)$	$\sigma(k)+v(k)-1-v(k-1)$		
Linear Lists	List	$(k+2)/2$	$(k+1)/2$		
	Position Tournament	$2(H_{k+1}-1)$	$2(1+\frac{1}{k})H_k-3$		

Table 1 : Coût évalué en nombre de comparaisons des principales structures stationnaires ; dans le cas

des files binomiales, $k = \sum b_i 2^i$, $v(k) = \sum b_i$ et $\sigma(k) = \frac{1}{k} \sum i b_i 2^i$.

	ULL	SLL	BST
number of comparison	$\frac{1}{10} n^2 + \frac{3n}{10} - \frac{7}{15}$	$\frac{n}{12} - \frac{3n}{4} + \frac{1}{6}$	$2nH_n - \frac{5}{2} n + O(\log^2 n)$
MIX time	$\frac{3}{4} n^2 + \frac{167}{12} n + \frac{13}{4} + O(\frac{1}{n})$	$\frac{23}{40} n^2 + \frac{613}{40} n + \frac{293}{40} + O(\frac{1}{n})$	$\frac{49}{3} n H_{n+1} - \frac{739}{18} n + O(\log^2 n)$

Table 2 : Coûts intégrés des structures de dictionnaires.

Structure	Integrated Cost
sorted list	$\frac{n(n+5)}{6}$
binary search tree	$\frac{2n!}{n?} \left\{ \sum_{1 \leq i < n} \frac{i?}{i!} \left[\frac{i(2^{n-i}-i)}{2^{i-1}} H_{n-i} + \frac{2^{n-i}-1}{n-i} \right] + \frac{2^n-1}{n} \right\} - 2n$ $= n \ln n + O(n)$
binary tournament	$\frac{n!}{n?} \left\{ \sum_{1 \leq i < n} \frac{i?}{i!} \left[\frac{3i(2^{n-i}-i)}{2^{i-1}} H_{n-i} + 5 \frac{2^{n-i}-1}{n-i} \right] + 5 \frac{2^n-1}{n} \right\} - \frac{9n}{2}$ $= \frac{3}{2} n \ln n + O(n)$
pagoda	identical to binary search trees
binomial queue	$= n \log_2 n + n I(n) + O(n)$

Table 3 : Coûts intégrés des files de priorité ; par convention $n? = 1.3.5...2n-1$.

CHAPITRE V

QUELQUES ASPECTS DE LA COMBINATOIRE DES FRACTIONS CONTINUES.

Nous avons montré, au chapitre précédent, un théorème simple d'équivalence entre séries génératrices de chemins plans valués apparus alors comme histoires de structures de données et fractions continues, et nous avons montré comment l'algèbre sous-jacente permettait de traiter simplement les dénombrements relatifs aux histoires de structures de données.

Nous revenons ici sur ce résultat, en montrant qu'il permet de donner des preuves combinatoires simples de diverses identités de fractions continues comprenant :

- d'une part des résultats généraux de développement en séries entières de fractions continues.
- d'autre part des preuves combinatoires de développements en fractions continues de séries entières divergentes particulières.

Ce chapitre montre au plan des généralités que divers travaux de Stieltjes, Rogers, Touchard et Carlitz où interviennent les fractions continues sont susceptibles d'un traitement unifié, la combinatoire naturelle des fractions continues étant celle des chemins plans valués. Il montre ensuite que les développements en fractions continues de séries entières divergentes à coefficients entiers et dont le prototype est la fraction continue d'Euler

$$\sum_n n! z^n = \frac{1}{1-1z - \frac{1^2 z^2}{1-3z - \frac{2^2 z^2}{1-5z - \frac{3^2 z^2}{\dots}}}}$$

s'obtiennent, au moyen de correspondances dues pour l'essentiel à [Françon, Viennot ; 1979], par la combinatoire des permutations et des partitions d'ensemble.

Une partie de ce chapitre est ainsi une réinterprétation et une synthèse de résultats connus. Cette tentative d'expliquer les théorèmes classiques de l'algèbre par des correspondances géométriques entre objets combinatoires a son origine dans les travaux de Foata et Schützenberger, qui écrivent[†] :

"Plus important nous semble la démonstration du fait que toutes les identités classiques... sont seulement la traduction de propriétés très simples de morphismes d'ensemble ordonnés."

Ce chapitre contient aussi des résultats nouveaux, parmi lesquels :

- une interprétation très simple des coefficients de Taylor des fonctions elliptiques de Jacobi *en* et *dn* différente de celle de [Viennot ; 1978].
- les interprétations des coefficients de Taylor des inverses des polynômes orthogonaux classiques : polynômes de Laguerre, Hermite, Meixner (Mittag-Leffler) et Poisson-Charlier.
- les développements en fractions continues des séries génératrices de nombres classiques : nombres de Stirling de première et seconde espèce, nombres de Bell, nombres de Bell associés, ainsi que certaines de leurs q -généralisations.

Ce chapitre montre certains prolongements combinatoires d'idées apparues au chapitre précédent à propos d'histoire de structures de données. Le traitement que nous en donnons en est cependant largement indépendant. Pour ne pas alourdir la présentation nous nous sommes parfois contentés d'esquisser certains développements possibles, renvoyant à [Flajolet ; 1979] pour une description plus détaillée de certains aspects.

[†] Voir [Foata, Schützenberger ; 1970]

1. COMBINATOIRE DES FRACTIONS CONTINUES DE JACOBI-STIELTJES.

L'un des problèmes classiques en algèbre des fractions continues est la relation entre une fraction continue du type

$$1 - \kappa_0 z - \frac{1}{1 - \kappa_1 z - \frac{\lambda_1 z^2}{1 - \kappa_2 z - \frac{\lambda_2 z^2}{1 - \kappa_3 z - \frac{\lambda_3 z^2}{\dots}}}} \quad \kappa_j, \lambda_j \in \mathbb{C},$$

et la série entière qui lui correspond par développement formel :

$$H(z) = H_0 + H_1 z + H_2 z^2 + H_3 z^3 \dots$$

Nous nous proposons de montrer ici que bon nombre de résultats élémentaires connus s'obtiennent sans calcul comme conséquences directes du théorème d'équivalence donné au chapitre IV.

1.1. L'équivalence fractions continues/chemins et les polynômes de Rogers.

Nous partons ici d'une fraction continue de type J

$$J(\bar{\kappa}, \bar{\lambda}; z) = \frac{1}{1 - \kappa_0 z - \frac{\lambda_1 z^2}{1 - \kappa_1 z - \frac{\lambda_2 z^2}{\dots}}} \quad (1)$$

que nous considérons pour l'instant comme une série formelle en les indéterminées $\bar{\kappa} = (\kappa_0, \kappa_1, \dots)$ $\bar{\lambda} = (\lambda_1, \lambda_2, \dots)$ et z . Soit $J^{[h]}(\bar{\kappa}, \bar{\lambda}; z)$ la h -ième réduite qui vaut

$$J^{[h]}(\bar{\kappa}, \bar{\lambda}; z) = \frac{1}{1 - \kappa_0 z - \frac{\lambda_1 z^2}{\dots \frac{\lambda_{h-1} z^2}{1 - \kappa_{h-1} z}}} \quad (1')$$

Chacune de ces fractions représente une série entière qu'on peut ordonner selon les puissances croissantes de z :

$$J(\bar{\kappa}, \bar{\lambda}; z) = \sum_{j \geq 0} H_j z^j; \quad J^{[h]}(\bar{\kappa}, \bar{\lambda}; z) = \sum_{j \geq 0} H_j^{[h]} z^j \quad (2)$$

où les H_n et $H_n^{[h]}$ sont des polynômes en $\bar{\kappa}; \bar{\lambda}$.

Le théorème d'équivalence entre fractions continues et séries caractéristiques de chemin du chapitre IV peut s'exprimer sous la forme

PROPOSITION 1^A. Le n -ième coefficient de Taylor de la fraction continue de Jacobi (1) est le polynôme caractéristique des chemins de longueur n dont les étapes montantes sont valuées[†] par 1, les paliers horizontaux à niveau k par κ_k et les descentes du niveau k par λ_k . $\square$

Ce résultat permet de former facilement les monômes qui interviennent dans la constitution des $H_n(\bar{\kappa}, \bar{\lambda})$

$H_0 = 1$: correspondant au chemin vide ;

$H_1 = \kappa_0$: correspondant au palier horizontal ;

$H_2 = \kappa_0 \kappa_0 + \lambda_2$: correspondant aux deux chemins $--, \wedge$;

$H_3 = \kappa_0 \kappa_0 \kappa_0 + \kappa_0 \lambda_1 + \lambda_1 \kappa_0 + \kappa_1 \lambda_1$: correspondant à $---, \wedge, \wedge, \wedge$;

et de même

$$H_4 = \kappa_0^4 + 2\kappa_0 \kappa_1 \lambda_1 + \kappa_1^2 \lambda_1 + \lambda_1^2.$$

A noter que si l'on pose $\lambda_j = \alpha_{j-1} \sigma_j$ les polynômes H_n deviennent homogènes en les indéterminées $\bar{\kappa}$, $\bar{\alpha} = (\alpha_0, \alpha_1, \dots)$ et $\bar{\sigma} = (\sigma_1, \sigma_2, \dots)$; l'on revient de la sorte à l'énoncé du théorème 1^B du chapitre IV.

Rogers dans son article de 1907 semble être le premier à avoir considéré explicitement ces polynômes que nous appellerons polynômes de Jacobi-Rogers (pour une fraction de type J). Dans le cas d'une fraction de Stieltjes où tous les coefficients κ_j sont nuls, on obtient les polynômes de Stieltjes-Rogers définis par

$$H_n'(\bar{\lambda}) = H_n(\bar{0}, \bar{\lambda}) \quad \text{où} \quad \bar{a} = (a, a, a, \dots)$$

(avec une définition identique pour les $H_n^{[h]}$).

Les polynômes $H_n'(\bar{\lambda})$ sont formés comme les H_n sans l'introduction de paliers ;

[†] Les chemins sont valués multiplicativement par les λ_k et κ_k . Ils sont donc équivalents aux histoires de structures de données relatives à une fonction de possibilité $\text{pos}(A; k) = 1$; $\text{pos}(S; k) = \lambda_k$; $\text{pos}(Q; k) = \kappa_k$ (les κ_k, λ_k apparaissant ici comme paramètres). Pour le traitement de cette section, il n'est nécessaire de valuer que les montées ou les descentes et nous avons choisi arbitrairement de valuer les descentes.

on a notamment l'égalité $H'_{2j+1} = 0$ et les premières valeurs des H'_{2j} sont

$$H'_0 = 1 ; H'_2 = \lambda_1 ; H'_4 = \lambda_1^2 + \lambda_1 \lambda_2 ; H'_6 = \lambda_1^3 + 2\lambda_1^2 \lambda_2 + \lambda_1 \lambda_2^2 + \lambda_1 \lambda_2 \lambda_3 ,$$

où par exemple H'_6 correspond aux cinq chemins



La proposition 1^A se simplifie donc en :

PROPOSITION 1^B. La fraction continue de Stieltjes $J'(\bar{\lambda}; z) = J(\bar{0}, \bar{\lambda}; z)$ se développe en série entière en z ; le coefficient H'_n est le polynôme caractéristique des chemins de longueur n sans étapes horizontales dont les étapes montantes sont étiquetées par 1 et les étapes descendantes du niveau k par λ_k . $\square$

Ces propositions s'appliquent aux réduites et les polynômes $H_n^{[h]}$ et $H'_n^{[h]}$ sont polynômes caractéristiques de chemins de hauteur bornée par h .

Ces résultats sont voisins de remarques dues à [Touchard ; 1952] et de résultats analogues indépendants de [Jackson ; 1978] et [Read ; 1979].

Il se pose naturellement la question de la nature des coefficients des polynômes et il est immédiat, par l'algèbre ou la combinatoire, que

H_n dépend seulement de $\lambda_1, \lambda_2, \dots, \lambda_j$ et $\kappa_0, \kappa_1, \dots, \kappa_{j-1}$

H_{n+1} dépend seulement de $\lambda_1, \lambda_2, \dots, \lambda_j$ et $\kappa_0, \kappa_1, \dots, \kappa_{j-1}, \kappa_j$.

La somme des coefficients de H_n vaut $H_n(\bar{1}, \bar{1})$ qui est le n -ième coefficient de la série associée à la fraction continue

$$\frac{1}{1 - 1z - \frac{1z^2}{1 - 1z - \frac{1z^2}{\dots}}},$$

et qui d'après l'interprétation combinatoire vaut le nombre de chemins de longueur n , c.à.d. le n -ième nombre de Motzkin M_n . On dispose d'une propriété analogue pour les H'_n et $H'_{2n}(\bar{1}) = H_{2n}(\bar{0}, \bar{1})$ est le $2n$ -ième coefficient de la série associée à la fraction continue

$$\frac{1}{1 - \frac{z^2}{1 - \frac{z^2}{\dots}}},$$

lequel vaut le nombre de chemins sans paliers de longueur $2n$, soit le n -ième nombre de Catalan $\frac{1}{n+1} \binom{2n}{n}$.[†] Quant aux coefficients eux-mêmes, ils sont susceptibles d'une forme explicite comme produit de coefficients binomiaux.

PROPOSITION 2^A. Les polynômes de Jacobi-Rogers ont l'expression explicite

$$H_n(\bar{\kappa}, \bar{\lambda}) = \sum_{\substack{h, n_1, \dots, n_h \\ m_0, \dots, m_h}} \rho(n_1, \dots, n_h; m_0, \dots, m_h) \lambda_1^{n_1} \lambda_2^{n_2} \dots \lambda_h^{n_h} \kappa_0^{m_0} \kappa_1^{m_1} \dots \kappa_h^{m_h}$$

à la somme est prise sur l'ensemble des $h \geq 0$ et des séquences $(n_1, \dots, n_h; m_0, \dots, m_h)$ telles que $2n_1 + 2n_2 + \dots + 2n_h + m_0 + m_1 + \dots + m_h = n$, ρ étant donné par

$$\begin{aligned} \rho(n_1, \dots, n_h; m_0, \dots, m_h) &= \binom{n}{m_0} \binom{n_1 + n_2 - 1}{n_1 - 1} \binom{n - m_0 - 2n_1}{m_1} \binom{n_1 + n_2 - 1}{n_1 - 1} \dots \\ &\dots \binom{n_{h-1} + n_h - 1}{n_{h-1} - 1} \binom{n - m_0 - \dots - m_{h-1} - 2n_1 - \dots - 2n_{h-1}}{m_h} . \quad \square \end{aligned}$$

Conventionnellement $\binom{p}{-1} = \delta_{p,-1}$ où δ est le symbole de Kronecker.

Dans le cas des fractions de Stieltjes, on dispose d'une expression explicite plus simple encore :

PROPOSITION 2^B. Les polynômes de Stieltjes-Rogers possèdent l'expression explicite :

$$H'_{2n}(\bar{\lambda}) = \sum_{\substack{h \geq 0 \\ n_0 + \dots + n_h = n}} \binom{n_1 + n_2 - 1}{n_1 - 1} \binom{n_2 + n_3 - 1}{n_2 - 1} \dots \binom{n_{h-1} + n_h - 1}{n_{h-1} - 1} \lambda_1^{n_1} \lambda_2^{n_2} \dots \lambda_h^{n_h} .$$

Preuve ($2^A, 2^B$). Les coefficients binomiaux $\binom{n}{m_0}$, $\binom{n - m_0 - 2n_1}{m_1}$... comptent le nombre de façons d'insérer des paliers horizontaux à hauteur $0, 1, \dots$. Les coefficients binomiaux $\binom{n_{r-1} + n_r - 1}{n_{r-1}}$ comptent le nombre de façons d'associer n_{r-1} points de niveau $r=1$, à n_r points de niveau r de manière compatible avec la structure de chemin positif. $\square$

Il est à noter que l'identité de la proposition 2^B correspond à deux modes

[†]

Ce développement a déjà été rencontré aux chapitres I, IV.

différents de génération des chemins : l'un (la fraction continue) à une génération de la gauche vers la droite ; l'autre (la série formelle) à une génération par niveaux successifs. Un résultat analogue vaut pour les formes quadratiques continues évoquées au chapitre I. Ce résultat généralise certains développements calculatoires de [Carlitz ; 1964] repris par [Riordam ; 1968] et issus d'un problème de tri. Il permet de retrouver directement les résultats de Carlitz (cf. [Flajolet ; 1979]).

1.2. L'algorithme de développement en fractions continues de Stieltjes.

On a vu qu'une fraction continue de type J est série caractéristique de chemins valués (proposition 1). Soit à chercher le développement en série entière de la fraction $J(\bar{\kappa}, \bar{\lambda} ; z)$; le problème équivaut à la détermination des nombres de chemins valués par $\bar{\kappa}, \bar{\lambda}$. Or, si l'on introduit les quantités

$H_{k,n}$ = nombre de chemins valués du niveau 0 au niveau k, de longueur n, les $H_{k,n}$ vérifient de simples récurrences linéaires tandis que les $H_{0,n} = H_n$ sont les coefficients du développement en série de $J(\bar{\kappa}, \bar{\lambda} ; z)$.

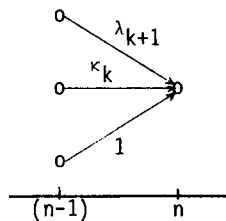
Ces récurrences sont en effet

$$H_{k,n} = H_{k-1,n-1} + \kappa_k H_{k,n-1} + \lambda_k H_{k+1,n-1} \quad (*)$$

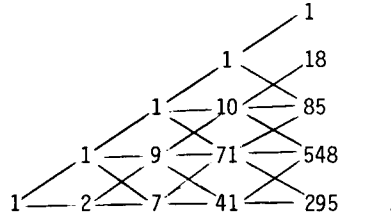
qui exprime qu'on arrive au niveau k en n étapes :

- soit par (n-1) étapes conduisant au niveau k-1, suivies d'une montée de coefficient 1 ;
- soit par (n-1) étapes conduisant au niveau k-1 suivies d'un palier horizontal de coefficient κ_k ;
- soit par (n-1) étapes conduisant au niveau k+1 suivies d'une descente de coefficient λ_{k+1} .

Ceci est résumé simplement par le diagramme suivant :



Donc si l'on dispose les $H_{k,n}$ en un tableau (le second indice étant indice de colonne), le calcul s'organise simplement de la gauche vers la droite, colonne par colonne. Par exemple le tableau correspondant à $\bar{\kappa} = (2,7,1,8...)$ et $\bar{\lambda} = (3,1,4,1,5...)$ commence comme suit



En arrêtant de la sorte à un niveau convenable, on s'aperçoit que le calcul de $H_0, H_1, \dots, H_{2n}$ nécessite le calcul de n^2 éléments du tableau, soit approximativement $4n^2$ opérations arithmétiques.

Réciproquement, le calcul des coefficients du développement en fraction continue d'une série entière

$$H(x) = \sum_{n=0}^{\infty} H_n x^n \quad \text{avec } f_0 = 1,$$

s'effectue simplement par la construction inductive de la matrice des $H_{k,n}$ par diagonales successives en partant de la gauche, la détermination des κ_j et λ_j progressant en même temps que la construction de la matrice. A l'étape $k+1$ ayant déjà construit : $\lambda_1, \lambda_2, \dots, \lambda_k$; $\kappa_0, \kappa_1, \dots, \kappa_k$; $\{H_{j,n} \mid j+n \leq 2k+1\}$, on calcule de proche en proche

$$A - \underline{\underline{H_{2k+2,0} = H_{2k+2}}}$$

$$H_{2k+1,1} = \frac{H_{2k+2,0} - \kappa_0 H_{2k+1,0}}{\lambda_1}$$

$$H_{2k,2} = \frac{H_{2k+1,1} - \kappa_1 H_{2k,1} - H_{2k,0}}{\lambda_2}$$

$$H_{2k-1,3} = \frac{H_{2k+2,2} - \kappa_2 H_{2k-1,2} - H_{2k-1,1}}{\lambda_3}$$

...

$$H_{k+1,k+1} = 1 \quad \Rightarrow \quad \boxed{\lambda_{k+1} = H_{k+2,k} - \kappa_k H_{k+1,k} - H_{k+1,k-1}}$$

$$\begin{aligned}
 B - \underline{H_{2k+3,0}} &= \underline{H_{2k+3}} \\
 H_{2k+2,1} &= \frac{H_{2k+3,0}^{-\kappa_0} H_{2k+2,0}}{\lambda_1} \\
 \dots \\
 H_{k+2,k+1} &= \frac{H_{k+3,k}^{-\kappa_k} H_{k+2,k}^{-H_{k+2,k-1}}}{\lambda_{k+1}} \Rightarrow \boxed{\kappa_{k+1} = H_{k+2,k+1} - H_{k+1,k}}
 \end{aligned}$$

Il s'agit là exactement de l'algorithme donné par Stieltjes qui calcule les $2n$ premiers coefficients du développement en fraction continue d'une série entière en temps $O(n^2)$.

DEFINITION : Etant donné une fraction continue $J(\bar{\kappa}, \bar{\lambda}; z)$; la matrice infinie des chemins valués $\{H_{k,n}\}_{k,n \geq 0}$ est appelée matrice de Stieltjes associée à J .

1.3. Propriétés de convolution de Stieltjes-Rogers.

Soit la fraction continue $J(\bar{\kappa}, \bar{\lambda}; z)$ correspondant à la série entière $H(z) = \sum_{n \geq 0} H_n z^n$ et soit $S = \{H_{k,n}\}_{k,n \geq 0}$ la matrice de Stieltjes associée. Il est commode de définir les quantités $\tilde{H}_{k,n}$ représentant les chemins valués partant du niveau k et finissant au niveau 0 . Les propriétés de symétrie évidentes entraînent les égalités :

$$\tilde{H}_{k,n} = \lambda_1 \lambda_2 \dots \lambda_k H_{k,n}.$$

On a, d'autre part, les propriétés de convolution remarquées par Françon dans le cadre des histoires de fichier :

$$H_{p+q} = \sum_{k \geq 0} H_{k,p} \tilde{H}_{k,q} \quad p, q \geq 0.$$

Cette propriété traduit le fait que pour aller du niveau 0 en $p+q$ étapes, on passe à l'étape p par un certain niveau $k > 0$.

On a ainsi entre $H(z)$ et S les relations :

$$H_{p+q} = \sum_{k \geq 0} \lambda_1 \lambda_2 \dots \lambda_k H_{k,p} H_{k,q} \quad \text{pour } p, q \geq 0.$$

Soient $\{X_p\}$ et $\{Y_q\}$ deux ensembles de variables ; si l'on multiplie cette égalité par $X_p Y_q$ et qu'on somme en p et q , on obtient la décomposition de la forme

quadratique $\sum_{p,q} H_{p+q} x_p y_q$:

$$\sum_{p,q \geq 0} H_{p+q} x_p y_q = \left(\sum_n H_{0,n} x_n \right) \left(\sum_n H_{0,n} y_n \right) + \lambda_1 \left(\sum_n H_{1,n} x_n \right) \left(\sum_n H_{1,n} y_n \right) \\ + \lambda_1 \lambda_2 \left(\sum_n f_{2,n} x_n \right) \left(\sum_n f_{2,n} y_n \right) \dots$$

Soient x et y deux variables ; si l'on multiplie l'égalité par $\frac{x^p}{p!} \frac{y^q}{q!}$ et qu'on somme en p et q , on obtient la formule d'addition relative à la série exponentielle

$$\hat{H}(x) = \sum_n H_n \frac{x^n}{n!} \\ \hat{H}(x+y) = \sum_{k \geq 0} \lambda_1 \lambda_2 \dots \lambda_k \hat{H}_k(x) \hat{H}_k(y)$$

où l'on a posé $\hat{H}_k(x) = \sum_{n \geq 0} H_{k,n} \frac{x^n}{n!}$. A noter que $\hat{H}_k(x) = \frac{x^k}{k!} + O(x^{k+1})$. On a ainsi démontré directement

PROPOSITION 3. Soit une fraction continue $J(\bar{\kappa}, \bar{\lambda}; z)$; et soit $S = \{H_{k,n}\}$ la matrice de Stieltjes associée.

(i) Définissant pour tout k la forme linéaire $\phi_k(\bar{X}) = \sum_{n \geq 0} H_{k,n} x_n$, la forme bilinéaire $\psi(\bar{X}, \bar{Y}) = \sum_{p,q \geq 0} H_{p+q} x_p y_q$ se décompose en :

$$\psi(\bar{X}, \bar{Y}) = \sum_{k \geq 0} \lambda_1 \lambda_2 \dots \lambda_k \phi_k(\bar{X}) \phi_k(\bar{Y}).$$

(ii) Définissant pour tout k la série exponentielle $\hat{H}_k(x) = \sum_{n \geq 0} H_{k,n} \frac{x^n}{n!}$, la série $\hat{H}(x) = \sum_{n \geq 0} H_n \frac{x^n}{n!}$ possède la formule d'addition :

$$\hat{H}(x+y) = \sum_{k \geq 0} \lambda_1 \lambda_2 \dots \lambda_k \hat{H}_k(x) \hat{H}_k(y).$$

Ces résultats expriment dans deux formulations différentes les relations entre une série entière et son développement en fraction continue. Le premier est dû à Stieltjes, le second est dû à Rogers. Ils sont chacun susceptibles d'une réciproque (cf. [Wall ; 1948] p. 203, [Perron ; 1953] p. 131).

Remarques : On peut interpréter de même sans calcul par la combinatoire des chemins, les opérations de contraction des fractions continues ou les formules de dérivations données par [Touchard ; 1954].

2. FRACTIONS CONTINUES, PARTITIONS ET PERMUTATIONS.

L'équivalence entre série caractéristique de chemins et fractions continues conduit, comme nous l'avons vu à la section 1 aux développements en fractions continues des séries génératrices des nombres de Catalan et Motzkin.

L'utilisation de correspondances "géométriques" -bijections combinatoires- entre systèmes de chemins valués et partitions ou permutations conduit à diverses expressions sous forme de fractions continues de séries génératrices ordinaires -divergentes- de quantités comptant des partitions d'ensembles ou des permutations. On obtient de la sorte les développements relatifs aux :

- nombre de Bell, nombres de Stirling de seconde espèce, nombres d'involution avec ou sans point fixe ;
- nombres factoriels ; nombres eulériens, nombres de Stirling de première espèce, nombres d'Euler et leurs généralisés.

Certaines de ces fractions continues ont été obtenues, par le calcul, par Euler, Stieltjes et Rogers, et nous en donnons des formes plus générales. Les correspondances chemins-permutations que nous utilisons en conjonction avec notre théorème d'équivalence sont dues à [Françon, Viennot ; 1978] .

Enfin, nous déduisons de ces développements une interprétation nouvelle des coefficients des fonctions elliptiques de Jacobi cn et dn .

2.1. Chemins plans.

Soit encore M_n le nombre de chemins positifs de 0 à 0 avec paliers ayant longueur n , et soit B_n le nombre⁺ de ces chemins de longueur $2n$ sans paliers horizontaux. La suite $\{M_n\} = \{1, 1, 2, 4, 9, 21, \dots\}$ est la suite des nombres de Motzkin de série génératrice $M(z) = \frac{1+z-\sqrt{1-2z+3z^2}}{2z}$; les B_n sont les nombres de Catalan de série génératrice $\frac{1-\sqrt{1-4z}}{2z}$ et l'on a vu qu'on avait comme conséquence directe du théorème d'équivalence :

⁺

B_n représente également le nombre d'arbres binaires formés de n sommets internes.

PROPOSITION 1^A. Les séries génératrices des nombres de Motzkin et Catalan admettent les développements

$$M(z) = \sum_{n \geq 0} M_n z^n = \frac{1}{1-z - \frac{z^2}{1-z - \frac{z^2}{\dots}}}$$

$$C(z) = \sum_{n \geq 0} C_n z^{2n} = \frac{1}{1 - \frac{z^2}{1 - \frac{z^2}{1 - \frac{z^2}{\dots}}}}$$

L'application de divers types de morphismes conduit à d'autres développements en fractions continues. Par exemple l'utilisation du morphisme

$$\mu(a_j) = \mu(s_j) = z \quad ; \quad \mu(q_j) = zu \quad \text{pour} \quad j \geq 0,$$

lequel marque par u les paliers horizontaux conduit à l'identité :

PROPOSITION 1^B. Soit $M_{n,k}$ le nombre de chemins de longueur n ayant k paliers horizontaux ; on a alors :

$$\sum_{n,k \geq 0} M_{n,k} u^k z^n = \frac{1}{1 - zu - \frac{z^2}{1 - zu - \frac{z^2}{\dots}}} \quad . \quad \square$$

On obtient de manière identique les développements des séries génératrices d'arbres binaires et d'arbres généraux[†] partagés selon leur nombre de feuilles. Ainsi, soit $G_{n,r}$ le nombre d'arbres généraux à n sommets ayant r feuilles (cf. chapitre I) :

$$\sum_{n,r \geq 0} G_{n,r} u^r z^n = \frac{zu}{1 - z(1+u) - \frac{z^2 u}{1 - z(1+u) - \frac{z^2 u}{\dots}}} \quad .$$

Enfin, introduisons l'aire sous un chemin comme la somme des hauteurs des points sur le chemin, l'aire $\alpha(u)$ d'un chemin u se définit inductivement par

$$\alpha(\epsilon) = 0 \quad ; \quad \alpha(a_j) = \alpha(s_j) = b \quad \text{pour} \quad j \geq 0 \quad ; \quad \alpha(x.y) = \alpha(x) + \alpha(y).$$

[†] On utilise pour cela les correspondances classiques arbres — chemins.

Soit $A(n,k)$ le nombre de chemins de longueur n ayant une aire égale à k .

Le morphisme μ défini par $\mu(a_j) = \mu(b_j) = zq^j$ pour $j \geq 0$ vérifie pour tout chemin u :

$$\mu(u) = z^{|u|} q^{\alpha(u)}.$$

L'application de ce morphisme conduit à une fraction continue due à [Carlitz ; 1974] .

PROPOSITION 1^C. *La série génératrice des chemins partitionnés selon le paramètre d'aire possède l'expression*

$$A(z,q) = \sum_{n,k \geq 0} A_{n,k} z^n q^k = \frac{1}{1 - \frac{z^2 q^1}{1 - \frac{z^2 q^3}{1 - \frac{z^2 q^5}{\dots}}}} \quad . \quad \square$$

2.2. Partitions d'ensembles.

Les partitions d'ensembles sont en correspondance naturelle avec certains systèmes de chemins valués. Cette correspondance qui étend une correspondance due à Françon et Viennot fournit diverses fractions continues relatives à des énumérations de partitions.

PROPOSITION 2. *Les partitions d'ensembles de taille n sont en bijection avec les chemins valués relatifs à la fonction de possibilité :*

$$\text{pos}(A; j) = 1 ; \text{pos}(S; j) = j+1 ; \text{pos}(Q; j) = j+1 \quad \text{pour tout } j \geq 0.$$

Preuve. La preuve est constructive. Soit π une partition relative à un ensemble à n éléments qu'on peut toujours supposer être l'intervalle $[1..n]$. Nous allons construire un couple (s,v) constitué d'un schéma s et d'une valuation v . Etant donné π les éléments de $[1..n]$ se répartissent en trois types :

- les éléments initiaux : ce sont les éléments appartenant à une classe de cardinalité ≥ 2 , et qui sont minimaux dans leur classe vis à vis de l'ordre usuel sur $[1..n]$;
- les éléments finaux : ce sont les éléments appartenant à une classe de cardinalité ≥ 2 et qui sont maximaux dans leur classe ;

- les éléments transitoires : ceux qui ne sont ni initiaux ni finaux ; il s'agit soit des éléments intermédiaires des classes de cardinalité ≥ 2 , soit des éléments des classes singletons.

Le schéma $s = s_1 s_2 \dots s_n$ est donné par

$$\begin{aligned} s_j &= A & \text{si } j \text{ est élément initial dans } \pi \\ s_j &= S & \text{si } j \text{ est élément final dans } \pi \\ s_j &= Q & \text{si } j \text{ est élément transitoire dans } \pi . \end{aligned}$$

La séquence des valuations $v_1 v_2 \dots v_n$ est définie de proche en proche comme suit :

- si j est élément initial ($s_j = A$), alors $v_j = 0$.

Etant donné un élément j et une classe $\alpha = \{\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_s\}$ de π , on dit que α recouvre j ssi $\alpha_1 < j \leq \alpha_s$; alors :

- si j est élément final ou élément transitoire d'une classe de cardinalité ≥ 2 , on considère les classes recouvrant j , et on les arrange selon l'ordre de leurs premiers éléments. Soient ainsi

$$\{\alpha_{01} \leq \alpha_{02} \leq \dots \leq \alpha_{0s_0}\}, \{\alpha_{11} < \alpha_{12} < \dots < \alpha_{1s_1}\} \dots$$

ces classes ; si j appartient à la v -ième classe $\{\alpha_{v1} < \alpha_{v2} \dots\}$, on pose $v_j = v$.

- si j appartient à une classe singleton, alors v_j est le nombre de classes recouvrant j .

On vérifie alors que la correspondance conduit à un couple (s, v) qui est cohérent avec la fonction de possibilité :

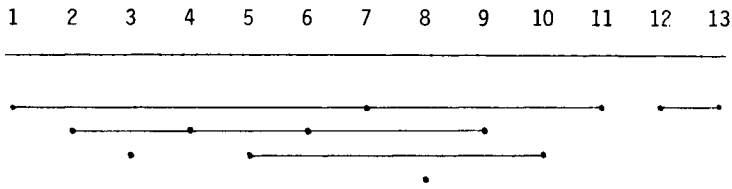
$$\text{pos}(A ; j) = 1 ; \text{pos}(S ; j) = j+1 ; \text{pos}(Q ; j) = j+1.$$

Cette correspondance est d'autre part clairement réversible. $\square$

Il est utile d'illustrer cette correspondance par un exemple. Soit ainsi $n=13$, et considérons la partition

$$\pi = \{1, 7, 11\} \{2, 4, 6, 9\} \{3\} \{5, 10\} \{8\} \{12, 13\} .$$

Cette partition se représente graphiquement de manière simple, des éléments d'une même classe étant joints par des segments horizontaux :



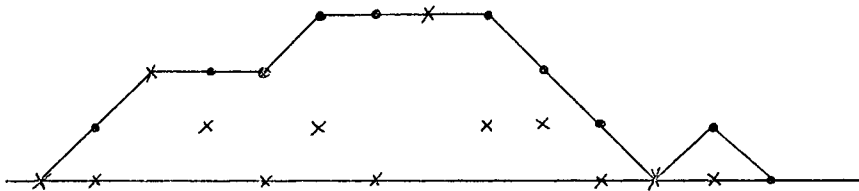
On voit clairement sur ce graphe que les classes recouvrant l'élément 6 sont $\{1,7,11\}$, $\{2,4,6,9\}$ et $\{5,10\}$. Le schéma associé à π est $s = s_1 s_2 \dots s_{13}$ où $s_1 = A$ car 1 est élément initial ; $s_2 = A$ car 2 est également initial ; $s_3 = Q$ car 3 est transitoire en tant qu'élément d'une classe singleton.... ; enfin $s_{13} = S$ car 13 est élément final de la classe $\{12,13\}$. Ainsi le schéma associé à π est-il

AAQQAQQSSAS

On détermine de même la valuation $v = v_1 v_2 \dots v_{13}$: 1 et 2 sont éléments initiaux de leur classe et donc $v_1 = v_2 = 0$; le point 3 est recouvert par les deux classes ouvertes précédemment $\{1, \dots\}$ et $\{2, \dots\}$ donc $v_3 = 2$; l'élément 4 est transitoire dans la seconde classe et donc $v_4 = 1$ (le rang des classes commence à zéro)... . Ainsi

$v = 002101311000$

et la représentation planaire du diagramme (s, v) est :



L'utilisation de cette correspondance conduit au résultat :

THEOREME 1. Soit $\beta_{n_1, n_2, m}$ le nombre de partitions d'ensembles comprenant n_1 classes singleton, n_2 classes de cardinalité ≥ 2 et m éléments non-transitoires ; la série génératrice

$$\beta(u_1, u_2, t, z) = \sum_{n_1, n_2, m \geq 0} \beta_{n_1, n_2, m} u_1^{n_1} u_2^{n_2} t^m z^{m+n_1+2n_2}$$

possède le développement

$$\beta(u_1, u_2, t, z) = \frac{1}{1 - u_1 z - \frac{1u_2 z^2}{1 - (u_1 + 1t)z - \frac{2u_2 z^2}{1 - (u_1 + 2t)z - \frac{3u_2 z^2}{\dots}}}}$$

En particulier :

$$(ia) \quad \sum \beta_n z^n = \frac{1}{1 - 1z - \frac{1z^2}{1 - 2z - \frac{2z^2}{\dots}}}}$$

$$(ib) \quad \sum S_{n,k} u^k z^n = \frac{1}{1 - uz - \frac{1uz^2}{1 - (1+u)z - \frac{2uz^2}{\dots}}}}$$

$$(iia) \quad \sum I_n z^n = \frac{1}{1 - z - \frac{1z^2}{1 - z - \frac{2z^2}{\dots}}}}$$

$$(iib) \quad \sum J_n z^n = \frac{1}{1 - \frac{1z^2}{1 - \frac{2z^2}{1 - \frac{3z^2}{\dots}}}}}$$

où les β_n sont les nombres de Bell (ou nombres exponentiels) ; les $S_{n,k}$ sont les nombres de Stirling de seconde espèce ; I_n compte les involutions de n et J_n compte les involutions sans point fixe de n , i.e. $J_{2n+1} = 0$ et $J_{2n} = 1.3.5 \dots 2n-1$.

Preuve. La première partie s'obtient en utilisant un morphisme μ

$$\mu(a_j) = u_2 z \quad \mu(b_j) = j z \quad \mu(c_j) = (u_1 + (j-1)t)z \quad \text{pour } j \geq 0.$$

Ce morphisme marque les classes de cardinalité ≥ 2 par leurs éléments initiaux étiquetés par u_2 ; pour les éléments transitoires correspondant à une hauteur de schéma de j , les $(j-1)$ premières possibilités sont associées à des éléments intermédiaires de classes de cardinalité ≥ 3 et la j -ième possibilité correspond à un élément singleton. $\square$

Des expressions analogues valent pour les nombres de Bell (2-) associés[†]

[†] Ces nombres sont déjà apparus dans un autre contexte au chapitre IV.

(cf. [Comtet ; 1970]) : soit $\beta_n^{(2)}$ le nombre de partitions d'un ensemble à n élément ne comprenant pas de classe singleton et soit $S_{n,k}^{(2)}$ le nombre de ces partitions constituées d'exactly k classes :

$$\sum_{n \geq 0} \beta_n^{(2)} z^n = \frac{1}{1 - \frac{1z^2}{1 - 1.z - \frac{2z^2}{1 - 2.z - \frac{3z^2}{\dots}}}} \quad \text{et} \quad \sum_{n,k \geq 0} S_{n,k}^{(2)} z^n = \frac{1}{1 - \frac{1uz^2}{1 - 1.z - \frac{2uz^2}{\dots}}}.$$

Il est à noter que toutes ces quantités ont des séries génératrices exponentielles de forme simple calculables directement ;

$$\beta(u_1, u_2, t, z) = \sum \beta_{n_1, n_2, m} u_1^{n_1} u_2^{n_2} t^m \frac{z^{n_1+m+2n_2}}{(n_1+m+2n_2)!}$$

est donné par :

$$\beta(u_1, u_2, t, z) = \exp(u_1 z + u_2 (\frac{z^2}{2!} + \frac{z^3 t}{3!} + \frac{z^4 t^2}{4!} \dots)).$$

Il est en particulier bien connu que : $\sum \beta_n \frac{z^n}{n!} = \exp(e^z - 1)$;

$$\sum S(n, k) u^k \frac{z^n}{n!} = \exp u(e^z - 1) ; \quad \sum I_n \frac{z^n}{n!} = e^{z+z^2/2} ; \quad \sum J_n \frac{z^n}{n!} = e^{z^2/2} ;$$

$$\sum \beta_n^{(2)} \frac{z^n}{n!} = \exp(e^z - z - 1) ; \quad \text{et} \quad \sum S_{n,k}^{(2)} u^k \frac{z^n}{n!} = \exp u(e^z - z - 1).$$

On pourrait ainsi réexprimer le théorème 2 comme développement en fraction continue de transformées de Laplace de séries exponentielles.

2.3. Permutations.

Nous rappelons ici la bijection fondamentale entre permutations et chemins due à Françon et Viennot. Nous l'utilisons pour déduire des fractions continues relatives aux nombres factoriels et à certains de leurs partages. L'utilisation du théorème d'équivalence en liaison avec cette bijection et un développement classique, nous permet de déduire une interprétation combinatoire nouvelle des coefficients de Taylor de fonctions elliptiques de Jacobi.

La construction d'une bijection entre permutations et chemins valués, passe par la représentation en arbres des permutations décrite au chapitre 1.

THEOREME F-V. (Correspondance entre permutations et chemins valués).

Les permutations de $[n+1]$ sont en bijection avec les chemins valués de longueur n correspondant au système de possibilités :

$$\text{pos}(A,j) = j+1 ; \quad \text{pos}(S,j) = j ; \quad \text{pos}(Q,j) = 2j+2 \quad \text{pour tout } j \geq 0.$$

Preuve : Nous ne faisons que rappeler ici le principe de la preuve, renvoyant à l'article de Françon-Viennot pour une description complète de la correspondance. Il est commode, comme dans le cas des histoires de fichiers, de considérer que les paliers horizontaux sont de deux types différents notés Q' et Q'' , le système correspondant alors à la fonction de possibilité

$$\text{pos}(A,j) = j+1 ; \quad \text{pos}(S,j) = j ; \quad \text{pos}(Q'_j) = \text{pos}(Q''_j) = j+1 \quad \text{pour } j \geq 0.$$

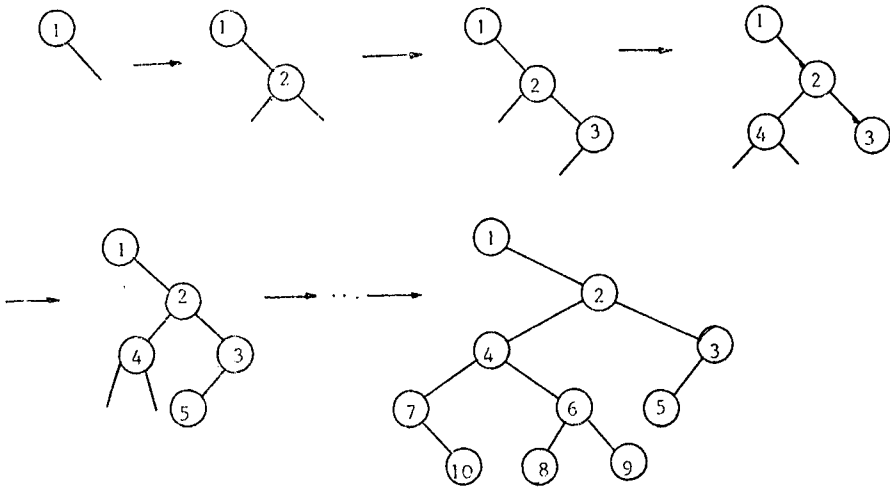
Partant d'un chemin valué (t,v) de longueur n , on construit un arbre tournoi sur $[n+1]$; la correspondance est algorithmique.

L'algorithme procède par insertions successives de sommets $1, 2, 3, \dots$ en partant à l'étape 0 de l'arbre vide. A l'étape j , pour $1 \leq j \leq n$, j positions sont possibles pour insérer le sommet j . Si la lettre t_j du schéma est un A , alors le sommet j est un point double ; si t_j est un S , c'est une feuille ; si t_j est un Q' , j est un point double à gauche ; enfin si t_j est un Q'' , j est un point double à droite.

A chaque étape j , la hauteur dans le schéma étant h_j , le nombre d'emplacements vacants, est $1+h_j$ avant que j soit inséré. Si l'élément correspondant de la valuation est v_j , le sommet j est placé à (v_j+1) -ième position vacante en partant de la gauche.

La construction se termine par le placement du sommet $(n+1)$ dans la dernière position vacante. □

Exemple. Soit le diagramme (t,v) où $t = Q'' A Q' A S A Q'' S S$ et $v = 001021011$; l'algorithme engendre la suite d'arbres partiels suivante :



□

Comme précédemment, l'intérêt principal de cette correspondance est dans son application aux problèmes d'énumération.

Soit $\sigma = \sigma_1 \sigma_2 \dots \sigma_n$ une permutation ; l'élément σ_j est un pic (ou maximum) ssi $\sigma_{j-1} < \sigma_j > \sigma_{j+1}$; un creux (ou minimum) ssi $\sigma_{j-1} > \sigma_j < \sigma_{j+1}$; une double montée ssi $\sigma_{j-1} < \sigma_j < \sigma_{j+1}$; une double descente ssi $\sigma_{j-1} > \sigma_j > \sigma_{j+1}$ (par convention $\sigma_0 = \sigma_{n+1} = 0$). Un pic d'une permutation est une valeur σ_j telle que $\sigma_{j-1} < \sigma_j$; c'est donc soit un creux, soit une double montée. Il est clair que le nombre de maxima d'une permutation est égal au nombre de ses minima augmenté d'une unité.

THEOREME 2^A. Soit $P_{k,\ell,m}$ le nombre de permutations comportant k minima (donc $k+1$ maxima), ℓ doubles montées et m doubles descentes. La série

$P(u,v,w,z) = \sum_{k,\ell,m} u^k v^\ell w^m z^{2k+\ell+m+1}$ possède l'expression

$$P(u,v,w,z) = \frac{1}{1-1(v+w)z - \frac{1.2uz^2}{1-2(v+w)z - \frac{2.3uz^2}{\dots}}}$$

En particulier :

$$\begin{aligned}
 \text{(i)} \quad \sum_{n \geq 0} (n+1)! z^n &= \frac{1}{1 - 2z - \frac{1.2z^2}{1 - 4z - \frac{2.3z^2}{\dots}}} ; \\
 \text{(ii)} \quad \sum_{\substack{n \geq 0 \\ k \geq 0}} A_{n+1, k+1} u^k z^n &= \frac{1}{1 - 1(1+u)z - \frac{1.2uz^2}{1 - 2(1+u) - \frac{2.3uz^2}{\dots}}} ; \\
 \text{(iii)} \quad \sum_{n \geq 0} E_{2n+1} z^{2n+1} &= \frac{1}{1 - \frac{1.2z^2}{1 - \frac{2.3z^2}{1 - \frac{3.4z^2}{\dots}}}} ;
 \end{aligned}$$

où $A_{n,k}$ est le nombre Eulérien comptant les permutations de $[1..n]$ à k pics ;
 E_{2n+1} est le nombre d'Euler impair comptant les permutations alternantes de $[2n+1]$.

Preuve : La preuve résulte directement de l'équivalence fondamentale et de la remarque que les pics, creux, doubles-montées et doubles descentes sont respectivement associées aux points doubles, feuilles, points simples à droite et points simples à gauche de l'arbre tournoi.

Il suffit donc d'appliquer au théorème d'équivalence le morphisme

$$\mu(a_j) = (j+1)uz \quad \mu(s_j) = (j+1)z \quad \mu(q'_j) = (j+1)vz \quad \mu(q''_j) = (j+1)wz.$$

Les autres cas s'obtiennent par particularisation du résultat général. $\square$

Il existe une modification importante de la correspondance de Françon-Viennot qu'ils ont également considérée dans leur article. Soit S'_{n+1} l'ensemble des permutations de $[1..n+1]$ se terminant par $n+1$:

$$S'_{n+1} = \{\sigma_1 \sigma_2 \dots \sigma_{n+1} \mid \sigma_{n+1} = n+1\}.$$

L'ensemble S'_{n+1} est évidemment isomorphe à l'ensemble S_n de toutes les permutations sur $[1..n]$. Les permutations de S'_{n+1} correspondent aux arbres ayant $(n+1)$ en bas de la branche droite issue de la racine. Ils sont donc tels que tous les arbres partiels de la construction chemin-arbres présentent une position libre sur la branche droite issue de la racine. C'est ainsi que ne doivent pas apparaître en

cette position des points simples à gauches ou des feuilles.

Ceci se traduit sur les chemins valués par l'interdiction d'une position dans le cas où apparaît dans le schéma **un** S ou **un** Q". Ainsi S'_{n+1} correspond-il au système de possibilités

$$\text{pos}(A, j) = j+1 \quad \text{pos}(S, j) = j \quad \text{pos}(Q', j) = j+1 \quad \text{pos}(Q'', j) = j.$$

THEOREME 2^B. On a les développements suivants

$$\sum_{n \geq 0} n! z^n = \frac{1}{1 - z - \frac{1^2 z^2}{1 - 3z - \frac{2^2 z^2}{\dots}}};$$

$$\sum_{n, k \geq 0} A_{n, k} u^k z^n = \frac{1}{1 - uz - \frac{1^2 uz^2}{1 - (1+2u)z - \frac{2^2 z^2}{\dots}}};$$

$$\sum E_{2n} z^{2n} = \frac{1}{1 - \frac{1^2 z^2}{1 - \frac{2^2 z^2}{1 - \frac{3^2 z^2}{\dots}}}};$$

$$\sum s_{n, k} u^k z^n = \frac{1}{1 - uz - \frac{uz^2}{1 - (2+u)z - \frac{(1+u)2z^2}{\dots}}};$$

où les $A_{n, k}$ sont les nombres Eulériens ; E_{2n} est le $2n$ -ième nombre d'Euler (ou nombre sécant) comptant les permutations alternantes de $[2n]$; $s_{n, k}$ est le nombre de Stirling de première espèce comptant les permutations de $[1..n]$ ayant k minima relatifs[†]

Preuve : cf. supra ; par exemple les minima relatifs correspondent aux positions de type A ou Q" et de valuation maximum. $\square$

[†] Un minimum relatif est un minimum dans l'ordre de gauche à droite.

THEOREME 2^C. Soit $C_{n_1, n_2, n}$ le nombre de permutations de S_n ayant n_1 cycles de longueur 1 et n_2 cycles de longueur 2. La série génératrice

$$C(u_1, u_2, z) = \sum_{n_1, n_2, n \geq 0} C_{n_1, n_2, n} u_1^{n_1} u_2^{n_2} z^n$$

possède le développement

$$C(u_1, u_2, z) = \frac{1}{1 - u_1 z - \frac{u_2 z^2}{1 - (2 + u_1)z - \frac{2(1 + u_2)z^2}{\dots}}}$$

En particulier, pour le nombre D_n de permutations sans points fixes (dérangements) :

$$\sum_{n \geq 0} D_n z^n = \frac{1}{1 - \frac{z^2 z^2}{1 - 2z - \frac{2^2 z^2}{\dots}}}$$

Preuve. La preuve est conséquence directe de la bijection décrite dans [Foata, Schützenberger ; 1970] p. 13 qui envoie S_n sur lui-même en échangeant les minima de droite à gauche avec les plus petits éléments de cycles. Les cycles singletons correspondent aux minima de droite à gauche qui sont en même temps double montées ; les éléments minimaux de cycles non singletons correspondent aux minima de droite à gauche qui sont aussi des creux. On ramène ainsi les comptages de cycles au marquage d'éléments repérables directement sur le chemin valué par la correspondance de Françon-Viennot. $\square$

Remarque : Les quantités qui apparaissent dans le théorème 2 ont des séries génératrices exponentielles de forme simple. En particulier

$$\begin{aligned} \sum_{n, k \geq 0} A_{n, k} u^k \frac{z^n}{n!} &= \frac{1-u}{1-ue^z(1-u)} \\ \sum_{n \geq 0} E_{2n+1} \frac{z^{2n+1}}{(2n+1)!} &= \operatorname{tg} z ; \quad \sum_{n \geq 0} E_{2n} \frac{z^{2n}}{(2n)!} = \sec z = \frac{1}{\cos z} ; \\ \sum_{n, k \geq 0} s_{n, k} u^k \frac{z^n}{n!} &= \exp(-u \cdot \ln(1-t)) = \frac{1}{(1-t)^u} ; \\ \sum_{n_1, n_2, n \geq 0} C_{n_1, n_2, n} u_1^{n_1} u_2^{n_2} \frac{z^n}{n!} &= \frac{\exp(u_1 - u_2)z}{(1-z)^{u_2}} ; \quad \sum_{n \geq 0} D_n \frac{z^n}{n!} = \frac{e^{-z}}{(1-z)} . \end{aligned}$$

Le théorème 2 pourrait donc également s'exprimer comme développement en fractions

continues de transformées de Laplace de séries exponentielles.

Certains de ces développements ont été obtenus préalablement par le calcul : les développements des séries $\sum n! z^n$ et $\sum (n+1)! z^n$ sont dûs à Euler qui les a utilisés pour sommer des séries divergentes solutions formelles d'équations différentielles simples. La manière la plus simple de les établir consiste à passer par la fraction continue de Gauss (cf. chapitre IV), à laquelle se réduit également le développement relatif aux nombres de Stirling de première espèce. Les développements relatifs aux nombres d'Euler et nombres Eulériens sont dûs à Stieltjes (et ont été partiellement retrouvés par Rogers) ; ils s'établissent au moyen du théorème d'addition de Stieltjes-Rogers.

On peut étendre la construction qui a permis de passer au théorème 2^A au théorème 2^B de la manière suivante. Considérons les r -forêts d'arbres tournois telles que $(n+j)$ occurre au bas de la branche droite de la j -ième composante de la forêt pour tout $j \in [1..n]$. De telles forêts sont bijectivement associées à la classe $S_{n+r}^{(r)}$ des permutations de S'_{n+r} telles que les valeurs $n+1, n+2, \dots, n+r$ apparaissent comme sous-séquence dans cet ordre :

$$S_{n+r}^{(r)} = \{ \sigma_1 \sigma_2 \dots \sigma_{n+r} \mid \sigma_{j_1} = n+1, \dots, \sigma_{j_r} = n+r \Rightarrow j_1 < j_2 < \dots < j_r = n+r \}.$$

Il est clair que $S_{n+1}^{(1)}$ coïncide avec S'_{n+1} . On peut modifier la correspondance de Françon-Viennot et vérifier qu'alors les r -forêts de ce type correspondent à la fonction de possibilité :

$$\text{pos}(A, j) = j+r ; \quad \text{pos}(S, j) = j ; \quad \text{pos}(Q', j) = j ; \quad \text{pos}(Q'', j) = j+r \quad \text{pour } j \geq 0.$$

La cardinalité de $S_{n+r}^{(r)}$ est donnée par les factorielles montantes

$$\text{card } S_{n+r}^{(r)} = (r)_n = r(r+1) \dots (r+n-1).$$

Les paramètres des arbres tournois s'étendent additivement aux forêts, ce qui conduit à de nouveaux paramètres définis sur $S_{n+r}^{(r)}$. Ainsi par exemple, si l'on part des points doubles et points simples à droite des arbres tournois, le paramètre correspondant des forêts est le nombre total de points doubles ou points simples à droite ; pour une r -forêt de taille $n+r$, il correspond au nombre de montées dans la permutation associée de $S_{n+r}^{(r)}$, en omettant de compter les montées possibles

$(n+j, n+j+1)$. Soit $A_{n,k}^{(r)}$ le nombre de permutations de $S_{n+r}^{(r)}$ présentant r montées de ce type ; les $A_{n,k}^{(r)}$ sont les nombres Eulériens d'ordre r (cf. [Foata, Schützenberger ; 1970], [Riordan ; 1968]). Soit de même $E_{2n}^{(r)}$ le nombre de permutations de $S_{2n+r}^{(r)}$ telles que les éléments $1, 2, 3, \dots, 2n$ sont soit pic soit creux. Les $E_{2n}^{(r)}$ sont les nombres d'Euler d'ordre r . On montre alors par l'utilisation de morphismes adéquats :

THEOREME 2^D. Les séries génératrices des factorielles montantes, des nombres d'Euler et nombres Eulériens d'ordre r admettent les développements en fractions continues

$$\sum_{n \geq 0} (r)_n z^n = \frac{1}{1 - rz - \frac{1 \cdot rz^2}{1 - (r+2)z - \frac{2(r+1)z^2}{\dots}}} ;$$

$$\sum_{n, k \geq 0} A_{n,k}^{(r)} u^k z^n = \frac{1}{1 - ruz - \frac{1ruz^2}{1 - ((r+1)u+1)z - \frac{2(r+1)uz^2}{\dots}}} ;$$

$$\sum_{n \geq 0} E_{2n}^{(r)} z^{2n} = \frac{1}{1 - \frac{1rz^2}{1 - \frac{2(r+1)z^2}{\dots}}} \quad \cdot \quad \square$$

Remarques : L'extension additive des paramètres de S_n à $S_{n+r}^{(r)}$ correspond à l'élévation à puissance r -ième des séries génératrices exponentielles associées :

$$\sum_{n \geq 0} (r)_n \frac{z^n}{n!} = \frac{1}{(1-z)^r} ; \quad \sum_{n \geq 0} E_{2n}^{(r)} \frac{z^{2n}}{2n!} = \sec^r z ;$$

$$\sum_{n, k \geq 0} A_{n,k}^{(r)} u^k \frac{z^n}{n!} = \left(\frac{1-u}{1-u \exp(z(1-u))} \right)^r .$$

Ces développements ont été également donnés par Stieltjes et Rogers. Les ${}^r A_{n,k}$ sont reliées aux énumérations de permutations avec positions interdites et l'on a :

$$A_{n,k}^{(r)} = r \cdot {}^r a_{n+r, n+1-k} ,$$

où les ${}^r a_{n,p}$ sont les nombres considérés dans [Foata, Schützenberger ; 1970] p. 45.

Fonctions elliptiques.

Il est possible à l'inverse de déduire de développements connus des interprétations en termes de chemins valués qu'on peut chercher à traduire par l'une des correspondances précédentes. C'est ce que nous allons faire ici pour les fonctions elliptiques de Jacobi $\text{cn}(u, \alpha)$ et $\text{dn}(u, \alpha)$.

Ces fonctions sont classiquement définies[†] par

$$\text{cn}(u, \alpha) = \cos \text{am}(u, \alpha) ; \text{dn}(u, \alpha) = \sqrt{1 - \alpha^2 \sin^2 \text{am}(u, \alpha)},$$

où $\text{am}(u, \alpha)$ est l'inverse d'une intégrale elliptique : par définition

$$\text{am}(u, \alpha) = \phi \quad \text{ssi} \quad u = \int_0^\phi \frac{dt}{\sqrt{1 - \alpha^2 \sin^2 t}}.$$

Les fonctions $\text{cn}(u, \alpha)$ et $\text{dn}(u, \alpha)$ sont développables en série de Taylor en u :

$$\text{cn}(u, \alpha) = \sum_{n \geq 0} (-1)^n \frac{u^{2n}}{(2n)!} C_n(\alpha^2) ; \quad \text{dn}(u, \alpha) = \sum_{j \geq 0} (-1)^j \frac{u^{2j}}{(2j)!} D_j(\alpha^2),$$

où C_n et D_n sont des polynômes de degré $(n-1)$, D_n étant le polynôme réciproque de C_n . La propriété remarquable des coefficients de C_n est d'être entiers : de plus, leur somme vaut le nombre d'Euler E_{2n} , ce qui traduit la propriété bien connue des fonctions elliptiques de se réduire aux fonctions hyperboliques lorsque le "module" α vaut 1. Comme E_{2n} compte les permutations alternantes de $[1..2n]$ il se pose naturellement le problème de l'existence d'un paramètre naturel de partitionnement des permutations alternantes, dont les dénombrements conduiraient aux coefficients des C_n .

THEOREME 3. *Le coefficient $C_{n,r}$ du développement en série de la fonction elliptique*

$$\text{cn}(u, \alpha) = \sum_{n, k \geq 0} (-1)^n C_{n,k} \alpha^{2k} \frac{u^{2n}}{2n!}$$

compte le nombre de permutations alternantes sur $[1..2n]$ ayant k creux à valeur paire.

Preuve : Les théorèmes classiques d'addition des fonctions elliptiques conduisent au développement

[†] On pourra se reporter à [Whittaker, Watson ; 1902].

$$\sum_{n,k \geq 0} (-1)^n c_{n,k} \alpha^{2k} z^{2n} = \frac{1}{1 + \frac{1^2 z^2}{1 + \frac{2^2 \alpha^2 z^2}{1 + \frac{3^2 z^2}{1 + \frac{4^2 \alpha^2 z^2}{\dots}}}}}$$

où l'on voit que $c_{n,r}$ compte les chemins valués de longueur $2n$ relatifs à la fonction de possibilité

$$\text{pos}(A,j) = j+1 ; \text{pos}(S,j) = j ; \text{pos}(Q',j) = \text{pos}(Q'',j) = 0,$$

dont le schéma comprend $2r$ étapes dans $\{a_1, a_3, a_5, \dots\} \cup \{b_2, b_4, b_6, \dots\}$. De manière équivalente $c_{n,r}$ compte les chemins valués comprenant r étapes dans $\{a_1, a_3, a_5, \dots\}$ puisqu'à toute étape de type a_{2j-1} correspond une étape b_{2j} . Il suffit de remarquer que dans un schéma sans palier horizontal étiqueté par hauteur comme par exemple

$$a_0 \underline{a_1} b_2 \underline{a_1} a_2 \underline{a_3} b_4 s_3 s_2 s_1$$

les $a_1, a_3, a_5, \dots$ (soulignés dans le schéma) apparaissent aux positions paires en partant de la gauche, et sont ainsi associées dans la correspondance de Françon-Viennot aux creux (étapes de type A) de valeur paire. $\square$

Cette interprétation est distincte de la première interprétation des coefficients des fonctions elliptiques, donnée par [Viennot ; 1978].

Exemple. Soit $\mathcal{C}_{n,r}$ l'ensemble des permutations alternantes de $2n$ comprenant r creux de valeur paire :

$$c_{0,0} = \{0\} ; c_{1,0} = \{21\} ; c_{2,0} = \{2143\} ; c_{2,1} = \{4231, 3142, 3241, 4132\},$$

ce qui est cohérent avec les premières valeurs des coefficients $c_{n,r}$.

3. INTERPRETATIONS COMBINATOIRES LIEES AUX POLYNOMES CLASSIQUES.

Nous avons vu au chapitre IV que les polynômes numérateurs et dénominateurs des réduites des fractions sont liés de diverses manières aux énumérations de chemins valués.

L'application de ces résultats aux fractions comptant partitions et permutations conduit, en liaison avec les bijections chemins-partitions et chemins-permutations à diverses interprétations de ces polynômes comprenant notamment les interprétations des coefficients de Taylor de leurs inverses. Les polynômes ainsi interprétés appartiennent aux familles classiques de Tchebycheff, Hermite, Laguerre, Meixner et Poisson-Charlier.

3.1. Rappels.

Nous nous contentons ici de reprendre les propositions établies au chapitre précédent et relative à une J-fraction :

$$J(z) = J(\bar{\kappa}, \bar{\lambda}; z) = \frac{1}{1 - \kappa_0 z - \frac{\lambda_1 z^2}{1 - \kappa_1 z - \frac{\lambda_2 z^2}{\dots}}}$$

Cette J-fraction énumère les chemins valués correspondant à une fonction de possibilité tout à fait générale :

$$\text{pos}(A; k) = \alpha_k; \quad \text{pos}(S; k) = \sigma_k; \quad \text{pos}(Q; k) = \kappa_k$$

où $\lambda_k = \alpha_{k-1} \sigma_k$.

La h-ième réduite notée $J^{[h]}(z)$ s'exprime comme quotient $\frac{P_h(z)}{Q_h(z)}$, où les P_h et Q_h vérifient la récurrence

$$P_{-1}(z) = 0 \quad P_0(z) = 1 \quad P_h(z) = (1 - \kappa_h z) P_{h-1}(z) - \lambda_h z^2 P_{h-2}(z)$$

$$Q_{-1}(z) = 1 \quad Q_0(z) = 1 - \kappa_0 z \quad Q_h(z) = (1 - \kappa_h z) Q_{h-1}(z) - \lambda_h z^2 Q_{h-2}(z).$$

Les polynômes "réciproques"

$$\bar{P}_{h-1}(z) = z^{h-1} \bar{P}_h(z) \quad \text{et} \quad \bar{Q}_{h-1}(z) = z^h \bar{Q}_h(z),$$

vérifient des récurrences trivialement déduites des précédentes.

PROPOSITION 1. La série génératrice des chemins valués de 0 à 0 ayant hauteur $\leq h$ vaut

$$J^{[h]}(z) = \frac{P_h(z)}{Q_h(z)}.$$

PROPOSITION 2. La série génératrice des chemins valués du niveau 0 au niveau k vaut

$$H_{0k}(z) = \sum_{n \geq 0} H_{0,k,n} z^n = \frac{1}{\sigma_1 \sigma_2 \dots \sigma_k z^k} [Q_{k-1}(z) J(z) - P_{k-1}(z)].$$

Remarque. Les séries $\frac{1}{\alpha_0 \dots \alpha_{k-1}} H_{0,k}(z)$ sont les séries génératrices ordinaires des colonnes de la matrice de Stieltjes associées à la fraction $J(z)$.

De la propriété $\text{val}(H_{0,k}(z)) = k$, on déduit que $P_{k-1}(z)$ coïncide avec les termes de degré inférieur à k de $Q_{k-1}(z) J(z)$. Ceci permet d'exprimer les coefficients des P_{k-1} comme convolution de coefficients des Q_{k-1} avec les coefficients de $J(z)$.

PROPOSITION 3. Les polynômes $\bar{Q}_{k-1}$ vérifient les relations d'orthogonalité formelle

$$\begin{aligned} \langle x^n | \bar{Q}_{k-1} \rangle &= 0 && \text{pour } 0 \leq n < k \\ \langle x^k | \bar{Q}_{k-1} \rangle &= \alpha_0 \alpha_1 \dots \alpha_{k-1} \sigma_1 \sigma_2 \dots \sigma_k \end{aligned}$$

pour le produit scalaire canoniquement associée à $J(z)$.

D'où en introduisant les coefficients des $\bar{q}_{h,r}$ polynômes $\bar{Q}_h$ définis par

$$\bar{Q}_h(z) = \sum_{n \geq 0} \bar{q}_{h,r} z^r :$$

PROPOSITION 4. La matrice de Stieltjes $S = \{H_{k,n}\}_{k,n \geq 0}$ et la matrice des coefficients des polynômes dénominateurs $K = \{\bar{q}_{h,r}\}$ sont inverses l'une de l'autre :

$$S.K = I \quad \text{où} \quad I = [\delta_{ij}]_{i,j \geq 0}.$$

Les premiers éléments des matrices S et K sont donnés ci-dessous lorsque les coefficients κ_j sont tous nuls.

$$S_{5 \times 5} = \begin{pmatrix} 1 & & & & \\ 0 & 1 & & & \\ \lambda_0 & 0 & 1 & & \\ 0 & \lambda_0 + \lambda_1 & 0 & 1 & \\ \lambda_0^2 + \lambda_0 \lambda_1 & 0 & \lambda_0 + \lambda_1 + \lambda_2 & 0 & 1 \end{pmatrix} \quad K_{5 \times 5} = \begin{pmatrix} 1 & & & & \\ 0 & 1 & & & \\ -\lambda_0 & 0 & 1 & & \\ 0 & -\lambda_0 - \lambda_1 & 0 & 1 & \\ \lambda_0 \lambda_0 & 0 & -\lambda_0 - \lambda_1 - \lambda_2 & 0 & 1 \end{pmatrix}$$

Il est à noter que les éléments de la matrice K représentent certains types de chemins dans un réseau. Dans le cas où $\kappa_j = 0$ pour tout j , les $\bar{q}_{h,r}$ vérifient les récurrences

$$\bar{q}_{h,r} = \bar{q}_{h-1,r-1} - \lambda_h \bar{q}_{h-2,r},$$

ce qui montre qu'au signe près les $\bar{q}_{h,r}$ comptent des chemins valués dont les étapes dans le plan $x-y$ sont formées des vecteurs $\alpha|_1^1$ et $\beta|_2^{0+}$. G. Viennot (communication à l'auteur) a montré l'existence d'une preuve de ce résultat fondée sur des transformations de chemins.

PROPOSITION 5. Les chemins de 0 à h ayant hauteur $< h$ ont pour série génératrice

$$J_h^{[h]}(z) = \frac{\alpha_0 \alpha_1 \dots \alpha_{h-1} z^h}{Q_{h-1}(z)}.$$

Preuve : Il suffit de remplacer $J(z)$ par $\frac{P_h(z)}{Q_h(z)}$ dans l'expression de la proposition 2 et d'utiliser l'"identité des déterminants" :

$$Q_{h-1}(z) P_h(z) - P_{h-1}(z) Q_h(z) = \lambda_1 \lambda_2 \dots \lambda_h z^{2h}. \quad \square$$

3.2. Partitions ; polynômes de Poisson-Charlier et Hermite.

Chaque système de chemins valués est associé à une classe de polynômes orthogonaux. L'identification des polynômes associés aux fractions de la section 2, s'effectue commodément par l'intermédiaire des équations différentielles traduisant les relations de récurrence (cf. chapitre IV). Les correspondances géométriques permettent alors d'interpréter en termes de permutations ou partitions les fractions apparues en 3.1 qui s'expriment au moyen de ces polynômes.

[†] C'est l'interprétation d'Euler-Mindig des polynômes des réduites.

Etant donnée une partition π de $[1..n]$, on définit la largeur de π en x , notée $\omega(\pi; x)$ comme le nombre de classes c recouvrant x

$$\omega(\pi; x) = \text{card} \{c \mid c \in \pi; \exists y_1, y_2 \in c \quad y_1 \leq x < y_2\}.$$

La largeur de π est simplement

$$\omega(\pi) = \max_{x \in [1..n]} \omega(\pi; x).$$

Par exemple pour $\pi = \{1,7,11\}, \{2,4,6,9\} \{3\} \{5,10\} \{8\} \{12,13\}$, la largeur de π en 6 vaut 3 car il existe trois classes recouvrant 6 : $\{1,7,11\}, \{2,4,6,9\}$ et $\{5,10\}$; de même $\omega(\pi; 3) = 2$ et $\omega(\pi) = 3$.

La largeur d'une partition est une mesure du chevauchement de ses classes.

THEOREME 4^A. Soit $B_n^{[h]}$ le nombre de partitions de $[1..n]$ de largeur $\leq h$; soit $\hat{B}_n^{[h]}$ le nombre de partitions de $[1..n+h]$ telles que les éléments $1, 2, \dots, h$ appartiennent à des classes distinctes non singleton, alors

$$\sum_{n \geq 0} B_n^{[h]} z^n = \frac{P_{h-1}(z)}{Q_{h-1}(z)} \quad \text{et} \quad \sum_{m \geq 0} \hat{B}_m^{[h]} z^m = \frac{h! z^h}{Q_{h-1}(z)}$$

où $Q_{h-1}(z)$ est le h -ième polynôme de Charlier réciproque ; P_{h-1} est déterminé à partir de Q_{h-1} par les règles de convolution de la section 3.1.

Preuve : Les polynômes de Charlier sont définis par

$$C_h(z) = \sum_{0 \leq k \leq h} (-1)^{h-k} \binom{h}{k} x(x-1)\dots(x-k+1)$$

et ont pour série exponentielle

$$\sum_{h \geq 0} C_h(z) \frac{u^h}{h!} = e^{-u} (1+u)^z$$

On vérifie que les polynômes réciproques donnés par $Q_{h-1}(z) = z^h C_h(\frac{1}{z})$ satisfont aux récurrences des polynômes de réduites associés à la fraction continue des nombres de Bell.

On vérifie d'autre part que la correspondance chemins partitions transforme la hauteur d'un chemin en la largeur de la partition. $\square$

La notion de largeur s'applique aux involutions qui sont des partitions particulières. La largeur d'une involution représente ainsi une mesure du chevauchement de ses cycles.

THEOREME 4^B. Soit I_n^h le nombre d'involutions de n de largeur h ;
soit $\hat{I}_{n+h}^h$ le nombre d'involutions de $n+h$ de largeur h , telles que $1, 2, \dots, h$
appartiennent à des cycles de longueur 2 différents. Soient I_n' et $\hat{I}_{n+h}'$ les
quantités homologues relatives aux involutions sans points fixes. Alors

$$\sum_{n \geq 0} I_n^{[h]} z^n = \frac{P_{h-1}(z)}{Q_{h-1}(z)} \quad \text{et} \quad \sum_{m \geq 0} \hat{I}_m^{[h]} z^m = \frac{h! z^h}{Q_{h-1}(z)}$$

$$\sum_{n \geq 0} I_n'^{[h]} z^n = \frac{P'_{h-1}(z)}{Q'_{h-1}(z)} \quad \text{et} \quad \sum_{m \geq 0} \hat{I}_m'^{[h]} z^m = \frac{h! z^h}{Q'_{h-1}(z)}$$

où $Q'_{h-1}(z) = z^h H_h(\frac{1}{z})$ est le h -ième polynôme d'Hermite réciproque et
 $Q_{h-1}(z) = z^h H_h(\frac{1}{z} - 1)$. Les polynômes P_{h-1} et P'_{h-1} sont donnés par les règles de
convolution.

3.3. Permutations : polynômes de Laguerre et Meixner.

Etant donnée une permutation $\sigma \in S_n$, et une valeur $x \in [0..n]$, on considère
le mot $w(x) \in \{+, -\}^n$ appelé signature de x dans σ et défini par

$$(w(x))_i = \begin{cases} + & \text{si } \sigma_i > x \\ - & \text{si } \sigma_i \leq x \end{cases}.$$

Ainsi pour $\sigma = 6 \ 4 \ 9 \ 3 \ 1 \ 5 \ 8 \ 2 \ 7$ les signatures $w(0), w(1), \dots$ sont

$$\begin{aligned} w(0) &= + + + + + + + + \\ w(1) &= + + + + - + + + \\ w(2) &= + + + + - + - + \\ w(3) &= + + + - - + - + \\ w(4) &= + - + - - + - + \\ w(5) &= + - + - - + - + \\ w(6) &= - - + - - + - + \\ w(7) &= - - + - - + - - \\ w(8) &= - - + - - - - - \\ w(9) &= - - - - - - - - \end{aligned}$$

Un groupe dans un mot $w \in \{+, -\}^n$ est un facteur maximal de w formé de symboles $+$
seulement. Le groupement de σ en x noté $gr(\sigma; x)$ est le nombre de groupes
de $w(x)$. Le groupement de σ est défini comme

$$\text{gr}(\sigma) = \max_{x \in [0..n]} \text{gr}(\sigma; x).$$

Dans l'exemple précédent, le groupement de 2 vaut 3, ce qui correspond aux trois groupes +++, ++ et + dans $w(2)$.

THEOREME. Soit $F_n^{[h]}$ le nombre de permutations de S_n de groupement $< h$; soit $E_{2n}^{[h]}$ le nombre de permutations alternantes de S_{2n} de groupement $< h$; alors :

$$\sum_{n \geq 0} F_n^{[h]} z^n = \frac{P_h(z)}{Q_h(z)} \quad \text{et} \quad \sum_{n \geq 0} E_{2n}^{[h]} z^{2n} = \frac{K_h'(z)}{N_h(z)}$$

où Q_h est le polynôme réciproque du h -ième polynôme de Laguerre d'ordre 1 et N_h est le h -ième polynôme de Meixner réciproque ; P_h et K_h sont déterminés par les règles de convolution.

Preuve. Rappelons que le m -ième polynôme de Laguerre d'ordre 1 est défini par

$$L_m^{(1)}(z) = \sum_{0 \leq k \leq m} \binom{n+1}{k+1} \frac{n!}{k!} (-1)^{n-k} z^k,$$

ce qui correspond à la série exponentielle

$$\sum_{m \geq 0} L_m^{(1)}(z) \frac{u^m}{m!} = \frac{1}{(1+u)^2} \exp \frac{zu}{1+u}.$$

Les polynômes de Meixner M_h ont pour série génératrice

$$\sum_{h \geq 0} M_h \frac{t^h}{h!} = (1+t^2)^{-1/2} \exp(z \arctan t).$$

Les Q_h et N_h de l'énoncé sont ainsi définis par

$$Q_h(z) = z^h L_h^{(1)}\left(\frac{1}{z}\right) \quad \text{et} \quad N_h(z) = z^h M_h\left(\frac{1}{z}\right).$$

On vérifie facilement que dans la correspondance de Françon-Viennot, le groupement d'une permutation diffère de 2 de la hauteur du chemin valué associé. $\square$

On peut comme dans le cas des partitions donner une interprétation des inverses des polynômes de Laguerre et Meixner.(cf. [Flajolet ; 1979]).

4. GENERALISATIONS.

Ces développements en fractions continues donnent lieu à différents types de généralisation lorsqu'on introduit dans les dénombrements de chemins valués des paramètres liés aux hauteurs des points du schéma, ou aux valeurs de la valuation.

Soit donné un système de possibilités :

$$(\sum) \text{ pos}(A; j) = \alpha_j ; \quad \text{pos}(S; j) = \sigma_j ; \quad \text{pos}(Q; j) = \kappa_j , \quad j \geq 0,$$

et un chemin valué $c = (u, v)$ de schéma $u = u_1 u_2 \dots u_n$ de valuation $v = v_1 v_2 \dots v_n$; soit $h_1 h_2 \dots h_n$ la suite des hauteurs des sommets dans le schéma. On a étudié précédemment la notion de hauteur (maximum) définie par

$$h.\max(c) = \max_{1 \leq i \leq n} h_i.$$

On définit de même la valuation maximum de c

$$v.\max(c) = \max_{1 \leq i \leq n} v_i.$$

Il est naturel d'introduire le paramètre de hauteur cumulée :

$$h.\text{cum} = \sum_{1 \leq i \leq n} h_i,$$

lequel représente l'aire comprise sous le schéma du chemin valué. On considère de même le paramètre de valuation cumulée :

$$v.\text{cum} = \sum_{1 \leq i \leq n} v_i.$$

Il se pose alors le problème des dénombrements relatifs à chacun de ces paramètres.

1. Le paramètre de hauteur se traduit au niveau des énumérations par le passage de la fraction continue à ses réduites (qui est rationnelle).
2. Le paramètre de valuation maximum conduit en général à des séries algébriques. Soit C_n le nombre de chemins de type $\sum$ et soit $v.\max_{C_n}^{[m]}$ le nombre de celles qui ont valuation maximum m . La série génératrice $v.\max_{C_n}^{[m]}(z) = \sum_{n \geq 0} v.\max_{C_n}^{[m]} z^n$ est susceptible du développement en fraction continue

$$\sum v.\max C_n^{[m]} z^n = \frac{1}{1 - \kappa'_0 z - \frac{\alpha'_0 \sigma'_1 z^2}{1 - \kappa'_1 z - \frac{\alpha'_1 \sigma'_2 z^2}{\dots}}}$$

où $\kappa'_j = \min(m, \kappa_j)$; $\alpha'_j = \min(m, \alpha_j)$; $\sigma'_j = \min(m, \sigma_j)$.

En particulier si les α_j , σ_j , κ_j sont croissants, les coefficients de la fraction continue se stabilisent à m à partir d'un certain niveau et pour j assez grand la j -ième troncature de la fraction continue vaut

$$\frac{m^2 z^2}{1 - mz - \frac{m^2 z^2}{1 - mz - \frac{m^2 z^2}{\dots}}}$$

soit la quantité $m^2 z^2 M(mz)$ où M est la série génératrice des nombres de Motzkin.

Ainsi $v.\max C_n^{[m]}$ est-il donné par une expression du type

$$\frac{A_m(z) M(mz) + B_m(z)}{C_m(z) M(mz) + D_m(z)}.$$

Dans les cas simples considérés précédemment, les A_m , B_m , C_m et D_m ne sont autres (à des facteurs de normalisation près) que les polynômes numérateurs et dénominateurs des réduites.

On pourra par exemple énumérer de la sorte différentes classes de permutations de Catalan (cf. [Knuth, 1968]) p. 234).

3. Le paramètre de hauteur cumulée conduit à une première classe de q -généralisations. Soit $h.\text{cum} C_{n,r}$ le nombre de chemins de type $\sum$ dont la hauteur cumulée vaut r ; on considère la série génératrice double

$$h.\text{cum} C(z; q) = \sum_{n,r \geq 0} h.\text{cum} C_{n,r} z^n q^r.$$

Celle-ci possède le développement en fractions continues

$$h.\text{cum}_C(z; q) = \frac{1}{1 - \kappa_0 q^0 z - \frac{\alpha_0 \sigma_1 q^1 z^2}{1 - \kappa_1 q^1 z - \frac{\alpha_1 \sigma_2 q^2 z^3}{1 - \kappa_2 q^2 z - \frac{\alpha_2 \sigma_3 q^3 z^4}{\dots}}}},$$

comme on le voit par l'utilisation du morphisme

$$\mu(A; j) = \alpha_j q^j; \quad \mu(S; j) = \sigma_j q^j; \quad \mu(Q; j) = \kappa_j q^j.$$

Les coefficients $h.\text{cum}_{C_{n,r}}$ vérifient les propriétés

- (i) $h.\text{cum}_{C_{n,r}} = 0$ si $r \geq n^2$
- (ii) $\sum_r h.\text{cum}_{C_{n,r}} = C_n$,

et constituent une "q-généralisation" des nombres C_n .

A l'exception du cas simple d'une fonction de possibilité constante, ces q-généralisations sont non classiques.

4. Le paramètre de valuation cumulée conduit à un second type de q-généralisations. Soit $v.\text{cum}_{C_{n,r}}$ le nombre de chemins de type Σ dont la valuation cumulée vaut r ; on considère maintenant la série double

$$v.\text{cum}_C(z; q) = \sum_{n,r \geq 0} v.\text{cum}_{C_{n,r}} z^n q^r.$$

Si l'on note pour tout nombre entier m

$$[m] = [m]_q = 1 + q + q^2 + \dots + q^{m-1} = \frac{1-q^m}{1-q},$$

on dispose pour $v.\text{cum}_C$ d'une expression en fraction continue

$$v.\text{cum}_C(z; q) = \frac{1}{1 - [\kappa_0] z - \frac{[\alpha_0][\sigma_1] z^2}{1 - [\kappa_1] z - \frac{[\alpha_1][\sigma_2] z^2}{\dots}}}.$$

Ces q-généralisations semblent se réduire aux q-généralisations classiques (en ne cumulant éventuellement que les valuations associées à certains types de paliers)

Interprétations combinatoires.

Nous avons donné, dans le cas des partitions et permutations, des interprétations combinatoires de la hauteur d'où il est immédiat de tirer des interprétations de la hauteur cumulée.

L'interprétation des valuations repose sur la traduction en termes des objets combinatoires, de la valeur des valuations dans un chemin valué et peut s'exprimer en termes de "largeur gauche" pour les partitions, ou en termes de "groupement gauche" pour les permutations.

BIBLIOGRAPHIE

- [Eatcher ; 1968] : "Sorting networks and their applications" ; in Proceedings AFIPS Spring Joint Comp. Conf., Montvale (1968), pp. 307-314.
- [Bentley ; 1978] : "Decomposable Searching Problems" ; Carnegie-Mellon University Report N° CMU-CS-78-145 (1978).
- [Berstel ; 1978] : "Séries formelles en variables non-commutatives et applications" ; J. Berstel Editeur, LITP et ENSTA, Paris (1978).
- [de Bruijn ; 1975] Communication personnelle à J. Vuillemin.
- [de Bruijn, Knuth, Rice ; 1972] : "The average height of planted plane trees" ; in "Graph Theory and Computing" ; R.C. Read Editor, Academic Press, New-York (1972), pp. 15-22.
- [Camion, Flajolet, Monier ; 1979] : "On Pollard's factorization algorithm", en préparation.
- [Carlitz ; 1964] : "A binomial identity arising from a sorting problem" ; in SIAM Rev. 6 (1964), pp. 20-30.
- [Carlitz ; 1974] : "q-analog of the Lagrange expansion" ; in Eulerian Series and Applications, Pennsylvania State Univ. (1974).
- [Chihara ; 1978] : "An Introduction to Orthogonal Polynomials", Gordon and Breach, New-York (1978).
- [Chomsky, Schützenberger ; 1963] : "The algebraic theory of context-free languages" in "Computer Programming and Formal Systems" ; North Holland P.C. (1963).
- [Chung, Luccio, Wong ; 1979] : "On the complexity of sorting in magnetic bubble memory systems", manuscript.
- [Comtet ; 1970] : "Analyse Combinatoire" ; 2 vol. P.U.F., Paris (1970).
- [Delange ; 1975] : "Sur la fonction sommatoire de la fonction somme des chiffres" ; in Enseignement Math. 21 (1975), pp. 31-47.
- [Edwards ; 1974] : "Riemann's Zeta Functions" ; Academic Press, New-York (1974).
- [Eilenberg ; 1974] : "Automata, Languages and Machines" ; Vol A ; Academic Press, New-York (1974).
- [Ershov ; 1958] : "On programming of arithmetic operations", CACM 1 (1958), 8 pp. 3-6.
- [Flajolet ; 1978] : "Analyse d'Algorithmes de manipulation de fichiers" ; Rapport IRIA, Rocquencourt (1978).

- [Flajolet ; 1978] : "Analyse de la détection d'arbres" in "3^{ème} Colloque de Lille sur les Arbres en Algèbres et en Programmation" ; Lille (1979).
- [Flajolet ; 1979] : "Combinatorial Aspects of Continued Fractions", soumis à Discrete Math.
- [Flajolet, Françon, Vuillemin ; 1979a] : "Computing Integrated costs of sequences of operations with application to dictionaries" ; in 11th ACM-SIGACT Conf., Atlanta (1979).
- [Flajolet, Françon, Vuillemin ; 1979b] : "Towards analyzing sequences of operations for dynamic data structures" ; in 20th IEEE-FOCS Conf., Porto-Rico (1979).
- [Flajolet, Ramshaw ; 1979] : "A note on Gray-Code and Odd-Even Merge" ; SIAM Journal on Comp. (à paraître).
- [Flajolet, Raoult, Vuillemin ; 1977 et 1979] : "The number of registers required for evaluating arithmetic expressions" ; version préliminaire dans "18th IEEE Symp. FOCS" (1977) ; Theoret. Comp. Sc. 9 (1979), pp. 99-125.
- [Flajolet, Steyaert ; 1979] : "On the analysis of tree matching algorithms", en préparation.
- [Foata ; 1971] : "La série génératrice exponentielle dans les problèmes d'énumération" ; Presses de l'Université de Montréal (1971).
- [Foata, Schützenberger ; 1970] : "Théorie géométrique des polynômes Eulériens" ; Lecture Notes in Mathematics, N° 138, Springer Verlag, Berlin (1970).
- [Françon ; 1977] : "Sur la fonction nombre de registres", manuscript.
- [Françon ; 1978] : "Histoires de Fichiers" ; in RAIRO Inf. Th., vol. 12 (1978), pp. 49-67.
- [Françon ; 1979] : "Combinatoire des Structures de données" ; Thèse, Faculté des Sciences de Strasbourg (1979).
- [Françon, Viennot ; 1979] : "Permutations selon leurs pics, creux, doubles montées et doubles descentes ; nombres d'Euler et de Genocchi" ; in Discrete Math. (1979), à paraître.
- [Françon, Viennot, Vuillemin ; 1978] : "Description and analysis of an efficient priority queue representation" ; in "19th I.E.E.E.F.O.C.S.", Ann Harbor (1978).
- [Françon, Viennot, Vuillemin ; 1978] : "Description et analyse d'une représentation performante des files de priorité" ; Rapport Informatique Université Paris-Sud (1978).
- [Henrici ; 1978] : "Applied and Computation Complex Analysis", vol.2 ; J. Wiley, New-York (1978).
- [Jackson ; 1978] : "Some results on product-weighted lead-codes" ; in J. Comb. Th., ser. A, 25 (1978), pp. 181-187.

- [Kemp ; 1977] : "The average number of registers to evaluate a binary tree optimally", Saarbrücken University Report (1977) et Acta Informatica (1979).
- [Knuth ; 1968] : "The Art of Computer Programming : Fundamental Algorithms" ; Addison Wesley, Reading (1968).
- [Knuth ; 1969] : "The Art of Computer Programming : Seminumerical Algorithms" ; Addison-Wesley, Reading (1969).
- [Knuth ; 1973] : "The Art of Computer Programming : Sorting and Searching" ; Addison-Wesley, Reading (1973).
- [Knuth, Schönhage ; 1978] : "The expected linearity of a simple equivalence algorithm" ; in Stanford Univ. Report CS-77-599 (1977).
- [Kreweras ; 1970] : "Sur les éventails de segments" ; Cahiers du B.U.R.O., n° 15 (1970), pp. 1-41.
- [Lucas ; 1891] : "Théorie des Nombres" ; Gauthier-Villard, Paris (1891).
- [Mac Ilroy ; 1974] : "The number of ones in binary integers : bounds and extremal properties" ; in SIAM J. on Comp. 3, n°4 (1974), pp. 255-261.
- [Mc Mahon ; 1975] : "The mechanical desing of trees" ; in Scient. Am. 233 (1975), 1 pp. 92-102.
- [Meixner ; 1934] : "Orthogonale Polynomsysteme mit einem besonderen Gestalt der erzeugenden funktion" ; J. Lond. Math. Soc. 9 (1934), pp. 6-13.
- [Odlyzko ; 1979] : "Periodic oscillations of coefficients of power series that satisfy functional equations" ; Bell Lab., Murray Hill (1979).
- [Perron ; 1954] : "Die Lehre von den Kettenbrüchen", 2 vol. Teubner, Stuttgart (1954).
- [Pollard ; 1975] : "A Monte-Carlo method for factorization" ; in BIT. 15 (1975), pp. 331-334.
- [Raney ; 1960] : "Functional composition patterns and power series reversion" ; Trans. A.M.S. 94 (1960), pp. 441-451.
- [Read ; 1979] : "The chord intersection problem" ; in Annals of N.Y. Ac. of Sc., 319 (1979), pp. 444-454.
- [Riordan ; 1968] : "Combinatorial Identities", John-Wiley and Sons, New-York (1968).
- [Rogers ; 1907] : "On the representation of certain asymptotic series as continued fractions" ; Proc. Lond. Math. Soc., 2 (1907), pp. 72-89.
- [Rota ; 1975] : "Finite Operator Calculus" ; Academic Press, New-York (1975).
- [Salomaa, Soittola ; 1978] : "Automata-theoretic Aspects of Formal Power Series" ; Springer Verlag, New-York (1978).
- [Sedgwick ; 1978] : "Data Movement in Odd-Even Merge" ; SIAM Journal on Comp. 7 (1978), pp. 239-272.

- [Sethi, Ullman ; 1970] : "The generation of optimal code for arithmetic expressions" ; in JACM 17 (1970), 4 pp. 715-728.
- [Shreve ; 1966] : "Statistical law of stream numbers" ; in Geology, 74 (1966), pp. 17-37.
- [Stieltjes ; 1889] : "Sur la réduction en fraction continue d'une série procédant suivant les puissances descendantes d'une variable" ; Ann. Fac. Sc. Toulouse, 3 (1889), pp. 1-17.
- [Szegő ; 1939] : "Orthogonal Polynomials" ; A.M.S. Colloq. Pub., Providence (1939).
- [Touchard ; 1952] : "Sur un problème de configurations et sur les fractions continues" ; Can. J. of Math. 4 (1952), pp. 2-25.
- [Viennot ; 1978] : "Une interprétation combinatoire des développements en série entière des fonctions elliptiques de Jacobi" ; soumis à Discrete Math. (1979).
- [Vuillemin ; 1978] : "A Data Structure for Manipulating Priority Queues" ; in CACM, 21 (1978), pp. 309-315.
- [Wall ; 1967] : "Analytic Theory of Continued Fractions" ; Chelsea Pub. Co., New-York (1967) réed.
- [Whittaker, Watson ; 1902] : "A Course on Modern Analysis" ; Cambridge Univ. Press (1902).
- [Widder ; 1971] : "An Introduction to Transform Theory" ; Academic Press, New-York (1971).