Université de Versailles-Saint-Quentin-en-Yvelines

Laboratoire PRiSM : Parallélisme, Réseaux et Systèmes de Modélisation

Thèse
pour obtenir le grade de Docteur
Spécialité : Informatique

Expressions Booléennes aléatoires :
probabilité, complexité et comparaison quantitative de logiques propositionnelles.

Antoine GENITRINI

Composition du jury :
M. Arnaud DURAND (rapporteur)
M. Hervé FOURNIER (co-directeur de thèse)
Mme Danièle GARDY (co-directrice de thèse)
M. Bernhard GITTENBERGER (président du jury)
M. Conrado MARTINEZ (rapporteur)
M. Christophe RAFFALLI

Versailles, le 9 Juillet 2009

Remerciements

Danièle et Hervé, je vous remercie de m'avoir guidé et accompagné tout au long de la préparation de ma thèse. Vos conseils, votre disponibilité, votre amabilité et la confiance que vous avez eue en moi m'ont permis d'accomplir ce travail dans les conditions les plus favorables. En outre, l'opportunité que vous m'avez offerte de participer à de nombreuses conférences et écoles, mais aussi de collaborer avec des collègues étrangers a également beaucoup contribué à ma formation.

Marek, Jakub, Grzegorz et Bernhard, je vous remercie de m'avoir fait partagé, avec patience, vos points de vue scientifiques par rapport à nos thèmes de recherche voisins. Nos collaborations m'ont été très enrichissantes.

Brigitte et Nicolas, les membres de l'équipe Alcaap et les doctorants du laboratoire, je vous remercie pour vos analyses critiques et pour la présentation de vos thèmes de recherche, via le séminaire Amap, par exemple. Vos regards extérieurs m'ont aidé à clarifier mes propos.

Arnaud et Conrado, je vous remercie d'avoir gentiment accepté de rapporter mon travail. Je remercie également Christophe et Bernhard qui ont été immédiatement d'accord pour participer au jury.

Claire, je te remercie d'avoir cru en mon projet, de m'avoir encouragé et soutenu tout au long de ma formation. Ton écoute, ta compréhension et l'intérêt que tu portes à mon travail ont participé à l'aboutissement de ce qui semblait un périple, il y a seulement trois ans.

Michèle et Bernard, Christine et Laurent, Marie et Pascal, je vous remercie en particulier pour le confort que vous m'avez apporté pendant la réalisation de ma thèse, mais aussi pour m'avoir aidé à vulgariser un peu mon travail.

Enfin, je remercie les nombreuses autres personnes qui m'ont fait partagé leur expérience afin de faire progresser ma recherche.

A ma compagne Claire

Table des matières

Première partie

Introduction

Chapitre 1

Présentation du contexte

Ecrivez une formule Booléenne de manière aléatoire, construite avec des ensembles fixés de connecteurs et de littéraux. Quelle est la fonction Booléenne "typique" qu'elle représente ? Quelle est la probabilité que la formule calcule la fonction *Vrai* ? Un littéral ? Ou une autre fonction donnée ? D'autres questions de nature un peu différente peuvent se poser : quelle est la complexité de la fonction Booléenne représentée par cette formule aléatoire (i.e. la taille de la plus petite formule la représentant) ? Quelle est la complexité moyenne d'une fonction Booléenne définie sur k variables ?

Shannon [Sha49] a démontré que si l'on considère une distribution uniforme sur l'ensemble des fonctions Booléennes à k variables, *la plupart des fonctions Booléennes ont une complexité en circuit valant* $O(2^k/k)$, ce qui est très proche de la complexité maximale $O(2^k)$ dans ce modèle. Si, à la place des circuits, on considère les formules construites avec les connecteurs binaires *Et* et *Ou* et les $2k$ littéraux positifs ou négatifs, *la plupart des fonctions Booléennes ont une complexité en formule valant* $O(2^k/\log k)$ [Weg87]. Ces résultats restent-ils valides si l'on ne considère plus la distribution uniforme sur les fonctions Booléennes, mais une autre distribution de probabilité ? Comment définir de façon naturelle une distribution sur ces fonctions ? A travers ma thèse, je vais répondre en partie à ces questions en prolongeant certains des travaux suivants.

Paris *et al.* [PVW94], suivis par Lefman et Savický [LS97] ont défini une distribution de probabilité sur les fonctions Booléennes à k variables, induite par les formules Booléennes, construites avec les deux connecteurs *Et/Ou*. Leurs travaux ont été repris et approchés différemment par Chauvin *et al.* [CFGG04] qui ont défini une autre manière, plus facile à manipuler, de considérer les ensembles de formules pour en déduire la distribution sur les fonctions – identique à la distribution de Lefman et Savický. Ainsi pour une fonction fixée, on détermine les proportions successives de l'ensemble des formules de taille n représentant la fonction par rapport à l'ensemble de toutes les formules de taille n. En laissant la taille des formules tendre vers l'infini, la suite de proportions induit une distribution de probabilité sur les fonctions Booléennes. On observe que cette distribution n'est pas uniforme. Par ailleurs, Chauvin *et al.* ont défini une seconde distribution basée sur les formules construites à l'aide d'un processus de branchement. La distribution induite sur les fonctions Booléennes n'est pas uniforme non plus. Par la suite, nous appellerons ces deux modèles respectivement le modèle des *grands arbres* et le modèle des *processus de branchement*.

Pour ces deux modèles de distributions, des bornes sur les probabilités des fonctions ont été prouvées [LS97], [CFGG04], [Woo05] et [Koz08]. Ces bornes font directement interve-

nir la complexité (en formules *Et/Ou*) des fonctions considérées. Cependant elles ne sont pas suffisantes pour infirmer (ou confirmer) l'effet Shannon sur ces modèles de distribution. Villa Monteiro a développé durant son stage de Master [Mon04] un programme construisant aléatoirement de grandes formules suivant le modèle des grands arbres. En étudiant les résultats, on a l'impression que la distribution de probabilité est biaisée vers les fonctions de petite complexité. En parallèle, certains travaux ([DFLS04], [PSS08]) ont développé la génération de Boltzmann qui consiste, dans le cas des arbres, à construire un arbre de taille donnée de manière uniforme parmi l'ensemble des arbres de cette taille.

Nous avons la possibilité de calculer de façon numérique les distributions pour un nombre de variables très petit – car le nombre de fonctions croit de manière doublement exponentielle par rapport au nombre de variables. L'article de synthèse de Gardy [Gar06] présente les outils nécessaires et l'applique au cas des formules *Et/Ou*. Les formules sont désormais vues comme des arbres binaires complets dont les noeuds internes sont étiquetés par les connecteurs et les feuilles par les littéraux. Les théorèmes de Drmota [Drm97], Lalley [Lal93] et Woods [Woo97] nous permettent de justifier aisément l'existence d'une distribution de probabilité sur les fonctions Booléennes pour de nombreux systèmes logiques (suivant les ensembles de connecteurs et de littéraux utilisés).

Le système avec les deux connecteurs *Et/Ou* et les $2k$ littéraux positifs et négatifs a suscité beaucoup d'intérêt du fait qu'il s'agit d'un système complet – i.e. toutes les 2^{2^k} fonctions Booléennes y sont expressibles. Toutefois, d'autres systèmes sont étudiés, notamment par l'école polonaise autour de Zaionc, dont le but est de quantifier les différences entre plusieurs logiques. Elle s'intéresse en particulier aux tautologies (formules toujours *vraie*) qui correspondent à ce qui est prouvable dans une logique, et permettent donc d'exprimer la puissance d'une logique par rapport à une autre. Dans cette optique, Moczurad *et al.* [MTZ00] ont étudié le système composé du connecteur *Implication* et de k littéraux positifs. L'intérêt porté à ce modèle de l'implication provient de l'isomorphisme de Curry-Howard qui relie une partie des tautologies (les tautologies intuitionnistes) aux programmes du *lambda-calcul*. L'équipe autour de Zaionc a obtenu des bornes sur la probabilité de *Vrai* – ou proportion asymptotique (par rapport à la taille) des expressions calculant *Vrai* parmi toutes les expressions. En se restreignant au système à une seule variable, ces chercheurs ont donné la valeur exacte de cette même probabilité. Cependant, leur but était de comparer de manière quantitative les logiques classique et intuitionniste – qui est un sous-ensemble de la logique classique. Ils ont conjecturé que lorsque le nombre de variables devient grand, la plupart des tautologies classiques sont intuitionnistes. En 2005, Zaionc [Zai05] s'est intéressé au modèle de l'implication et la négation à une variable. Il a obtenu la probabilité exacte de *Vrai* dans ce modèle. Par ailleurs, Matecki [Mat05] a étudié le système basé sur l'unique connecteur *Equivalence*. Pour ce modèle, il n'y a pas de distribution induite par l'ensemble des formules sur les fonctions Booléennes. Cependant, si l'on ne considère que les formules de taille ayant la même parité, il est tout de même possible de définir une distribution sur les fonctions Booléennes.

Une autre approche considérant les formules Booléennes a été celle de Valiant [Val84] en 1984. Son but était de construire, avec un unique connecteur, des petites formules monotones calculant la fonction *Majorité* sur k variables, lorsque k devient grand. Les travaux de Ajtai *et al.* [AKS83] venaient alors de prouver qu'il existait de telles formules de taille polynomiale, mais dont l'exposant était énorme. L'approche probabiliste de Valiant a permis d'obtenir, avec une grande probabilité, de telles formules de taille $O(k^{5.3})$. Les résultats sont obtenus en construisant des formules dont les arbres les représentant sont équilibrés – toutes

les feuilles sont au même niveau. La construction a été plus tard améliorée par Gupta et Mahajan [GM97]. Puis cette construction de formules équilibrées a été utilisée pour obtenir de petites formules représentant les fonctions *seuil* [Bop85], des fonctions *read-once* [DZ97] ou les fonctions *seuil linéaire* [Ser04]. La méthode, nommée *amplification probabiliste* par Boppana, a par la suite été reprise afin de définir des distributions de probabilité sur les fonctions Booléennes induites par de grandes formules équilibrées. Ainsi, Savický [Sav90] et Brodsky et Pippenger [BP05] ont étudié le procédé de construction des formules équilibrées et ont classifié les distributions sur les fonctions Booléennes selon les propriétés de l'unique connecteur utilisé.

Ma thèse est construite autour des différentes démarches qui viennent d'être présentées : distributions de probabilité induites sur les fonctions Booléennes, comparaison quantitatives de logiques propositionnelles. Plus précisément, mon travail porte sur les expressions Booléennes de différents types et sur les probabilités qu'elles induisent sur les fonctions Booléennes. Je vais démontrer un lien entre la probabilité d'une fonction et sa complexité – la taille de la plus petite expression la représentant. Par ailleurs, je vais être amené à détailler la structure d'une partie des expressions représentant une fonction et je vais étudier en détail celle des tautologies qui me permet d'établir une comparaison quantitative des logiques intuitionniste et classique dans différents systèmes.

Après avoir décrit formellement les deux modèles d'arbres que je prendrai en considération, je commence par proposer une étude du système composé d'un unique connecteur *Implication* et de k littéraux positifs. Pour un nombre de variables k inférieur ou égal à 4, nous sommes en mesure de calculer numériquement de façon exhaustive les probabilités de chaque fonction Booléenne. Ensuite, je me tourne vers une approche théorique. La définition de trois familles de formules dont une qui ne contient que des tautologies me permet d'obtenir des bornes inférieure et supérieure sur la probabilité de la fonction *Vrai* telles que leur terme principal coïncide lorsque k devient grand. Puis, pour une fonction Booléenne fixée, je donne le terme principal (quand k tend vers l'infini) de la probabilité de cette fonction suivant les deux modèles d'arbres – grands arbres et processus de branchement. Ce résultat est obtenu en prouvant que la plupart des formules calculant une fonction ont une structure simple, basée sur un des plus petits arbres représentant la fonction auquel (dans le cas des grands arbres) on "greffe" un sous-arbre qui ne modifie pas la fonction représentée. Ceci a été présenté lors de la conférence MFCS'08 [FGGG08]. Un article étudiant les tautologies dans le système de l'implication avec les littéraux positifs et négatifs est soumis au journal Mathematical Logic Quarterly. Par ailleurs, un article concernant le système de l'implication avec k variables est en préparation et sera soumis au journal Random Structures and Algorithms.

Dans une seconde partie, j'étudie les logiques intuitionniste et classique dans des modèles divers. Cette étude prend en considération la structure "typique" des tautologies lorsque le nombre de variables tend vers l'infini. Après avoir exhibé des familles de tautologies ayant une forme simple, je démontre que la taille de l'ensemble des tautologies résultantes est négligeable face à la taille de l'ensemble des tautologies simples. Je démontre qu'un seul connecteur (dans le système propositionnel complet) engendre une différence significative entre les deux logiques lorsque le nombre de variables tend vers l'infini. Par ailleurs, je prouve la conjecture de Moczurad *et al.* [MTZ00] supposant que la plupart des tautologies (formées avec l'implication) sont intuitionnistes (quand le nombre de variables tend vers l'infini) ; cela a été présenté à CSL'07 [FGGZ07]. Une étude plus fine permet par la suite d'obtenir plus de détails quant aux tautologies ; ceux-ci ont été exposés lors des conférences Types'08 [GKZ07] et au colloque MCS'08 [GKM08]. Du fait que l'étude soit plus détaillée, de nombreuses familles

sont présentées et nécessitent le calcul de leur proportions : des outils basés uniquement sur un squelette de ces familles nous donne directement la proportion et me permettent de quantifier aisément les différences entre logique intuitionniste et classique. Un dernier article concernant la logique dans le système propositionnelle complet a été présenté à LFCS'09 [GK09]. Ce résultat a par ailleurs été soumis au journal Annals of Pure and Applied Logic.

Enfin, la dernière partie concerne les arbres équilibrés. Je commence par rendre exhaustive la classification de Brodsky et Pippenger [BP05] pour les arbres construits avec un unique connecteur binaire. Puis j'étudie les arbres Et/Ou équilibrés : l'amplification probabiliste basée sur deux connecteurs aléatoires est une nouvelle approche et permet d'étudier la distribution de probabilité engendrée sur les fonctions Booléennes induite par des expressions équilibrées. Cette sous-classe d'arbres du modèle des grands arbres ne contient que peu d'arbres ce qui a pour conséquence que la distribution de probabilité sur les fonctions Booléennes est très différente de celle obtenue via les arbres Et/Ou généraux. Cette partie a été exposée à ANALCO'09 [FGG09].

Chapitre 2

Deux modèles d'expressions aléatoires

Paris *et al.* [PVW94], puis Lefman et Savický [LS97] ont cherché à définir une distribution de probabilité "naturelle" sur les fonctions Booléennes, induite par des expressions construites avec les deux connecteurs *et* et *ou* et un nombre k fixé de variables – ce système est complet, i.e. toute fonction Booléenne sur k variables est expressible dans ce système. L'approche de Lefman et Savický utilise des arbres de taille infinie qui sont élagués. En 2004, Chauvin *et al.* présentent une autre approche [CFGG04] se basant sur une suite de proportions d'arbres de taille donnée et démontrent que la limite de cette suite définit une distribution de probabilité sur les fonctions Booléennes, de plus cette distribution est identique à celle de Lefman et Savický. Par ailleurs, leur article présente une seconde distribution de probabilité induite par des arbres construits via un processus de branchement critique, suivi d'un étiquetage aléatoire et indépendant des noeuds. L'article de synthèse de Gardy [Gar06] montre l'existence de ces deux distributions de probabilité pour toute une classe de systèmes.

Par ailleurs, on trouve, dans cet article de synthèse, des résultats numériques pour un nombre restreint de variables. Nous remarquons clairement que les deux distributions de probabilité sont biaisées vers les fonctions qui s'expriment le plus simplement – i.e. avec des arbres les représentant relativement petits. En raison du grand nombre de fonctions Booléennes à k variables : 2^{2^k}, on se trouve très rapidement dans l'impossibilité d'obtenir des résultats numériques lorsque k augmente.

Le même article [Gar06] présente dès lors les outils d'analyse combinatoire nous permettant de nous intéresser au cas où le nombre de variables k tend vers l'infini. Ces outils sont étudiés très largement dans les livres de Flajolet et Sedgewick [FS96] et [FS09]. Citons-en quelques notions classiques qui seront largement utilisées par la suite et dont la présentation est exhaustive dans ces ouvrages : arbres de Catalan, arbres de Galton-Watson, fonction génératrice énumérant les expressions, théorème de Drmota-Lalley-Woods, ...

Enfin, les articles [LS97], [CFGG04] et [Koz08] présentent des bornes sur la probabilité d'une fonction donnée, dépendant directement de la *complexité* de la fonction Booléenne – i.e. la taille de ses plus petits arbres. Ces résultats semblent biaisés vers les fonctions de petite complexité, mais ne donnent pas beaucoup d'information sur les fonctions de grande complexité. Ainsi, ils ne nous permettent pas directement de conclure quant à l'effet Shannon sur ces deux distributions de probabilité.

Ces distributions biaisées, le sont-elles en raison du système de connecteurs utilisés, ou

le résultat est-il plus fortement lié à la structure d'arbre sous-jacente ? L'effet Shannon a-t-il aussi lieu pour ces distributions de probabilités ?

Nous aborderons entre autres ces questions via l'étude d'un système différent, basé sur l'unique connecteur binaire *implication*. Nous travaillerons sur les deux principaux modèles d'expressions dont la définition formelle suit :

– Grands arbres ;
– Processus de branchement.

Pour chaque modèle, les expressions peuvent être représentées sous la forme d'arbres binaires complets (i.e. les noeuds ont zéro fils ou deux fils) dont les noeuds internes sont étiquetés par un connecteur et les feuilles par un littéral. Nous définissons une application *taille* qui à chaque arbre associe son nombre de feuilles. La figure 2.1 présente un arbre dans le système de l'implication – un seul connecteur : $\rightarrow$.

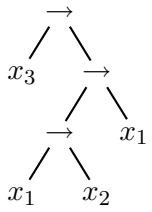

FIG. 2.1 – Exemple d'un arbre binaire complet et étiqueté de taille 4.

Cette expression s'écrit $x_3 \rightarrow ((x_1 \rightarrow x_2) \rightarrow x_1)$ et calcule (ou représente) la fonction Booléenne $\bar{x}_3 \vee \bar{x}_2 \vee x_1$. En effet, rappelons que $x_i \rightarrow x_j$ est équivalent à $\bar{x}_i \vee x_j$, où $\bar{x}_j$ représente la négation de x_j. La fonction Booléenne $\bar{x}_3 \vee \bar{x}_2 \vee x_1$ est de complexité 3 et l'expression $x_2 \rightarrow (x_3 \rightarrow x_1)$ est une des ses expressions *minimales* (i.e. de taille la plus petite possible) pour le système de l'implication. Enonçons quelques définitions formelles :

Taille. La taille $|A|$ d'un arbre A est le nombre de ses feuilles. Dans le cas des arbres binaires complets, rappelons qu'il y a exactement une feuille de plus que de noeuds internes. Par ailleurs, lorsque l'usage de parenthèses s'avère nécessaire pour l'écriture en ligne d'une expression, elles ne sont pas comptées dans la taille de l'expression.

Complexité. Pour une fonction f expressible dans le système, nous appellerons *complexité* de la fonction, et noterons $L(f)$, la taille du plus petit arbre du système la calculant. Les arbres calculant f et de taille $L(f)$ seront appelés *arbres minimaux* de f. L'ensemble de arbres minimaux sera noté $\mathcal{M}_f$.

Le modèle dit des *grands arbres* est basé sur l'ensemble de tous les arbres constructibles à l'aide d'un ensemble fini de connecteurs (qui permet d'étiqueter les noeuds internes) et d'un ensemble fini de variables (qui permet d'étiqueter les feuilles). A taille n fixée, nous pouvons définir la proportion d'un ensemble d'arbres vérifiant certaines propriétés parmi tous les arbres de cette taille. Pour une fonction Booléenne fixée, nous pouvons, par exemple, calculer la proportion d'arbres de taille n la représentant parmi tous les arbres de taille n. La limite de la suite de proportions (lorsque n varie), permet dans certains cas, de définir la probabilité de la fonction Booléenne en question. Afin qu'un système de connecteurs et de variables définissent une distribution de probabilité sur les fonctions Booléennes, il faut que ce système vérifie des conditions provenant des théorèmes de Drmota [Drm97], Lalley [Lal93] et Woods [Woo97], qui sont trois versions du même résultat. Dans l'article de Chauvin *et al.* [CFGG04], le lecteur trouvera une application de ces théorèmes au cas des arbres *Et/Ou*. Dans le chapitre suivant,

nous vérifierons que les conditions sont réunies pour le système de l'implication et k littéraux positifs. Donnons la définition formelle de la distribution de probabilité.

Fraction limite. La *fraction limite* d'un sous-ensemble $\mathcal{A}$ d'arbres est défini par

$$\mu_k(\mathcal{A}) = \lim_{n\to\infty} \frac{|\{A \in \mathcal{A}, |A| = n\}|}{|\{A \in \mathcal{F}_k, |A| = n\}|},$$

lorsque cette limite existe. Notons $\mathcal{F}_k(f)$ l'ensemble des arbres calculant une fonction f ; dès lors nous définissons la *probabilité $\mu_k(f)$ d'une fonction f* par la fraction limite de l'ensemble $\mathcal{F}_k(f)$. Pour un ensemble $\mathcal{A}$ donné, lorsque nous ne prouverons pas que la limite existe, nous considérerons les valeurs $\mu^-(\mathcal{A})$ et $\mu^+(\mathcal{A})$ qui seront respectivement les *limites inférieure et supérieure* de la proportion précédente, ainsi :

$$\mu_k^+(\mathcal{A}) = \limsup_{n\to\infty} \frac{|\{A \in \mathcal{A}, |A| = n\}|}{|\{A \in \mathcal{F}_k, |A| = n\}|},$$

$$\mu_k^-(\mathcal{A}) = \liminf_{n\to\infty} \frac{|\{A \in \mathcal{A}, |A| = n\}|}{|\{A \in \mathcal{F}_k, |A| = n\}|}.$$

Le modèle du *processus de branchement* considère une distribution de probabilité sur les arbres, définie par un processus de Galton Watson critique. Pour ce modèle, chaque noeud a une arité aléatoire, indépendante de celles des autres noeuds et qui vaut uniformément zéro ou deux. En outre, les noeuds sont étiquetés de manière aléatoire, indépendante et uniformément avec l'ensemble des connecteurs (pour les noeuds internes) ou avec l'ensemble des littéraux pour les feuilles. Ainsi, la probabilité d'une expression est égale à la probabilité de sa structure d'arbre multipliéé par la probabilité de l'étiquetage de l'arbre. Les arbres, obtenus par un processus de Galton-Watson critique, sont finis presque sûrement, ainsi nous définissons la probabilité d'une fonction Booléenne comme la somme des probabilités des arbres la représentant. Pour une référence aux processus de branchement, le lecteur pourra se tourner vers le livre de Athreya et Ney [AN72]. L'introduction de cette distribution de probabilité sur les arbres *Et/Ou* est faite dans l'article [CFGG04].

Voilà une définition formelle de la distribution de probabilité dans ce modèle :

Probabilité pour le processus de branchement. Pour un arbre A, nous avons

$$\pi_k(A) = \mathbb{P}(\text{structure de } A)\ \mathbb{P}(\text{étiquetage de } A) = \frac{1}{2^{2|A|-1}\ k^{|A|}}.$$

De manière immédiate, la probabilité d'un ensemble d'arbres est la somme des probabilités de chaque arbre appartenant à l'ensemble. Finalement la probabilité d'une fonction f donnée est la probabilité de l'ensemble $\mathcal{F}_k(f)$ d'expressions calculant f. Nous vérifions aisément que $\pi_k(\mathcal{F}_k) = 1$.

Deuxième partie

Système de l'implication

Chapitre 3

Généralités sur le système de l'implication

Dans toute cette partie, les formules sont construites avec l'unique connecteur implication que nous noterons $\rightarrow$, et l'ensemble de littéraux est réduit aux k littéraux positifs $\{x_1, \ldots, x_k\}$ – sauf Section 5.5.2. Les expressions peuvent être représentées par des arbres binaires complets contenant $\rightarrow$ dans les noeuds internes et un littéral dans chaque feuille. La Figure 3.1 en donne un exemple.

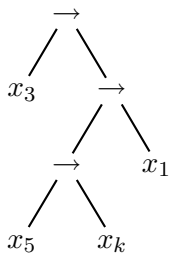

FIG. 3.1 – Représentation arborescente de l'expression $x_3 \rightarrow ((x_5 \rightarrow x_k) \rightarrow x_1)$.

Nous appellerons $\mathcal{B}_k$ l'ensemble des fonctions Booléennes sur k variables et $\mathcal{F}_k$ l'ensemble des expressions Booléennes que nous pouvons construire dans le système de l'implication et k variables. Par la suite nous utiliserons de manière synonyme les termes *expressions Booléennes* et *formules Booléennes*. Lorsque nous considérerons plus spécifiquement la structure d'une expression Booléenne, nous utiliserons aussi le terme *arbre*.

Nous remarquons tout d'abord que le système de l'implication n'est pas complet : toutes les fonctions de $\mathcal{B}_k$ ne sont pas expressibles dans le système. Les fonctions expressibles dans le système, sont les fonctions f pour lesquelles il existe une variable x_i et une fonction $g \in \mathcal{B}_k$ telles que $f = x_i \vee g$. Le nombre de fonctions expressibles dans le système est équivalent à $k \cdot 2^{2^{k-1}}$, lorsque k tend vers l'infini. Le fait qu'une fonction du système s'exprime sous la forme d'une disjonction d'une variable et d'une autre fonction découle de façon immédiate de la représentation suivante.

Commençons par définir ce que nous appellerons la *forme canonique* d'une expression. Soit A une expression, sa forme arborescente peut être décomposée le long de sa branche droite – voir Figure 3.2. Ainsi il est de la forme $A_1 \rightarrow (A_2 \rightarrow (\ldots \rightarrow (A_p \rightarrow r(T)))\ldots)$; que nous réécrirons dorénavant

$$A = A_1, \ldots, A_p \rightarrow r(A).$$

Les expressions A_i sont appelées les *prémisses* de A, et $r(A)$, la feuille la plus à droite de l'arbre est appelée le *but* de A. Evidemment l'expression $A = A_1 \to (A_2 \to (\ldots \to (A_p \to r(A)))\ldots)$ est équivalente, de façon logique, à $\overline{A_1} \vee \overline{A_2} \vee \cdots \vee \overline{A_p} \vee r(A)$, où $\overline{A_i}$ représente la négation de A_i. De façon analogue on définit les prémisses et le but d'un sous-arbre. Nous les nommerons *sous-buts* et *sous-prémisses*.

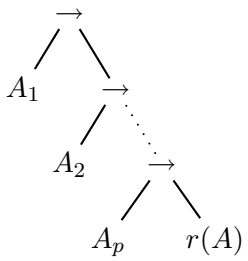

FIG. 3.2 – La décomposition canonique d'un arbre

Pour un arbre donné A, la fonction calculée par l'expression sera notée $[A]$. Pour une affectation a, complète ou partielle, des variables, nous noterons $A_{|a}$ l'arbre dans lequel les étiquettes qui sont évaluées par a sont remplacées par la valeur qui leur est associée par a. Par abus de langage, il nous arrivera de définir des fonctions Booléennes par des expressions Booléennes : il est clair que la fonction en question sera celle calculée par l'expression présentée.

Rappelons quelques résultats qui nous seront utiles par la suite.

Soit $C(z)$ la fonction génératrice énumérant les arbres binaires complets, où z marque les feuilles ; elle satisfait :

$$C(z) = z + C(z)^2,$$

et est égale à :

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

Ses coefficients sont :

$$[z^{n+1}]C(z) = C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Dans le système de l'implication, la fonction génératrice $f_k(z)$ énumérant l'ensemble des arbres satisfait l'équation suivante : $f_k(z) = kz + f_k(z)^2$, obtenue en utilisant la construction récursive des arbres. De plus le fait que $f_k(0) = 0$ nous permet de conclure :

$$f_k(z) = \frac{1 - \sqrt{1-4kz}}{2}.$$

Le $n$ième coefficient de f_k est lié au $(n-1)$ème nombre de Catalan $[z^n]f_k(z) = k^n C_{n-1} = k^n/n \cdot \binom{2n-2}{n-1}$ –la notation $[z^n]\phi(z)$ correspond au $n$ème coefficient de la série $\phi(z)$.

Rappelons que pour un ensemble d'arbres quelconque, la fraction limite n'est pas toujours définie : soit $\mathcal{D}$ l'ensemble suivant : $\mathcal{D} = \{\mathcal{A} \subset \mathcal{F}_k \mid \mu_k(\mathcal{A})\ existe\}$. Cet ensemble $\mathcal{D}$ est une algèbre Booléenne : si $\mathcal{A}$ et $\mathcal{B}$ appartiennent à $\mathcal{D}$, alors le complémentaire de $\mathcal{A}$, l'union $\mathcal{A} \cup \mathcal{B}$ et l'intersection $\mathcal{A} \cap \mathcal{B}$ appartiennent à $\mathcal{D}$.

Intéressons-nous aux distributions de probabilité des fonctions Booléennes définies dans le système de l'implication et k variables.

Théorème 1 *[CFGG04] Si f est une fonction de k variables, alors sa probabilité $\mu_k(f)$ existe. Si f n'est pas expressible dans le système, alors $\mu_k(f) = \pi_k(f) = 0$, sinon on peut écrire la fonction génératrice ϕ_f énumérant les expressions calculant f sous la forme*

$$\phi_f(z) = \alpha_f - \beta_f\sqrt{1-4kz} + O\left(z - \frac{1}{4k}\right);$$

$$\textit{et alors} \qquad \pi_k(f) = \frac{\alpha_f}{\sum_g \alpha_g} \qquad \mu_k(f) = \frac{\beta_f}{\sum_g \beta_g}.$$

Nous remarquons que la distribution $\pi_k(\cdot)$ existe toujours, puisque la structure d'arbre obtenue par le processus de branchement est finie presque sûrement.

Idées de preuve : L'utilisation des fonctions génératrices nous permet, via les résultats de Drmota [Drm97], Lalley [Lal93] et Woods [Woo97] de prouver aisément que la probabilité $\mu_k(\cdot)$ de chaque fonction Booléenne de $\mathcal{B}_k$ est définie. Ces théorèmes sont décrits dans le livre de Flajolet et Sedgewick [FS09].

Le but de cette preuve est de montrer l'existence de la probabilité $\mu_k(f)$ pour l'ensemble des arbres $\mathcal{F}_k(f) = \{A \in \mathcal{F}_k \mid [A] = f\}$ calculant une fonction f donnée. Pour des raisons techniques et uniquement dans cette preuve, la taille d'un arbre sera donnée par le nombre de ses noeuds internes – du fait qu'il ne diffère que d'un par rapport au nombre de feuilles considéré dans le reste de la thèse, l'existence et la valeur de $\mu_k(f)$ ne sont pas affectées par ce changement. Soient $f_1, \ldots, f_p$ les fonctions Booléennes pouvant être représentées par des expressions de $\mathcal{F}_k$. Nous notons ϕ_i la fonction génératrice énumérant les arbres calculant f_i ; ainsi, $\phi_i(z) = \sum_{n=0}^{\infty} \alpha_{i,n} z^i$ où $\alpha_{i,n}$ est le nombre d'arbre de taille n calculant f_i. Une autre manière de présenter cette fonction génératrice est d'utiliser la construction récursive des arbres : $\phi_i(z) = 1_{f_i \text{ littéral}} + z\sum_{f_j, f_k;\ f_j \to f_k = f_i} \phi_j(z)\phi_k(z)$. Soit Φ le système polynomial obtenu en considérant la structure récursive des expressions. Soit $y_n \in \mathbb{C}[[z]]$ une variable pour $1 \leqslant n \leqslant p$. Pour $1 \leqslant n \leqslant p$, soit $\Phi_n = \varepsilon_n + z\sum_{i,j} y_i y_j$ où $\varepsilon_n = 1$ si f_n est un littéral et 0 sinon, et où la somme est définie pour $1 \leqslant i, j \leqslant p$ tels que $f_i \to f_j = f_n$. Evidemment, $(\phi_1, \ldots, \phi_p)$ est une solution du système polynomial Φ (et d'ailleurs, c'est la seule).

Afin de démontrer l'existence de $\mu_k(f_i)$ pour tout i, il suffit de vérifier que le système polynomial Φ satisfait les conditions du théorème de Drmota, Lalley et Woods. Nous allons suivre la présentation faite dans le livre de Flajolet et Sedgewick [FS09, Chapitre VII] – le lecteur se référera également à ce livre pour les définitions.

Il est clair que le système Φ est *non linéaire* (il est quadratique) et *algébrique positif*. Vérifions qu'il est *algébrique propre*. Pour cela, montrons que pour $a, b \in \mathbb{C}[[z]]^p$, $\mathrm{val}(\Phi(a) - \Phi(b)) > \mathrm{val}(a - b)$. Soit $\ell \in \{1, \ldots, p\}$, nous avons $\Phi_\ell(a) - \Phi_\ell(b) = z\sum_{i,j}(a_i a_j - b_i b_j)$, où la somme est sur tous les i, j tels que $f_i \to f_j = f_\ell$. Remarquons que $a_i a_j - b_i b_j = \frac{1}{2}((a_i + b_i)(a_j - b_j) + (a_j + b_j)(a_i - b_i))$. Il s'en suit $\mathrm{val}(\Phi_\ell(a) - \Phi_\ell(b)) \geqslant 1 + \mathrm{val}(a - b)$. Ceci prouve que $\mathrm{val}(\Phi(a) - \Phi(b)) > \mathrm{val}(a - b)$.

Démontrons désormais que le système Φ est *algébrique irréductible*. Pour cela il suffit de démontrer qu'il existe un "chemin" de n'importe quelle f_i à *Vrai* et un second chemin de *Vrai* à toute f_j dans le système. La première partie provient du fait que $f_i \to f_i$ calcule *Vrai*, et la seconde partie est obtenue car Vrai $\to f_j$ calcule f_j.

Enfin, vérifions que Φ est *algébrique apériodique*. Soit A une expression calculant la fonction f_i. Les deux expressions $(x_1 \to x_1) \to A$ et $(x_1 \to (x_1 \to x_1)) \to A$ calculent f_i. Du fait que ces expressions sont de taille $|A| + 2$ et $|A| + 3$ et que 2 et 3 sont premiers entre eux, ceci montre que pour ℓ suffisamment grand, il existe une expression de taille ℓ calculant f_i. Donc l'apériodicité est prouvée.

D'après le théorème de Drmota, Lalley et Woods, en développant ϕ_f autour de la plus grande singularité (commune à toutes les fonctions génératrices ϕ_g, où g est expressible dans le système), les deux premiers coefficients sont directement reliés aux probabilités $\pi_k(f)$ et $\mu_k(f)$ – il suffit de les normaliser. Pour plus de détails, le lecteur peut se référer à [Gar06]. □

Chapitre 4

Résultats numériques

Pour des très petites valeurs de k, le nombre de fonctions expressibles dans le système est raisonnable, et nous pouvons donc résoudre de façon numérique (voire exacte) le système algébrique Φ, de degré 2, satisfait par les fonctions génératrices des fonctions Booléennes (voir Chapitre 3 [p. 13]). Par ailleurs, nous remarquons que deux fonctions égales à permutation des variables près ont la même fonction génératrice. Soit la relation d'équivalence suivante [Har65, p. 147] : deux fonctions sont équivalentes s'il existe une permutation des variables transformant la première en la seconde. Cette relation permet de définir des classes de fonctions et par conséquent, en passant aux fonctions génératrices, nous obtenons un système Φ' de taille inférieure à celle de Φ. Le tableau 4.1 associe au nombre k de littéraux le nombre de fonctions distinctes dans le système ainsi que le nombre de classes d'équivalence.

k	nb. fonctions	nb. classes
1	2	2
2	6	4
3	38	12
4	942	80
5	325262	3984
6	25768825638	37333248

FIG. 4.1 – Tableau présentant le nombre de fonctions pour les premières valeurs de k.

Sur ces quelques valeurs nous observons bien la croissance doublement exponentielle du nombre de fonctions Booléennes lorsque k augmente.

4.1 Système de l'implication avec une variable

Si notre système ne contient qu'une seule variable, seule deux fonctions y sont expressibles. Il s'agit des fonctions x et *Vrai*. Construisons grâce à la structure récursive des arbres les fonctions génératrices de ces fonctions. Soient $\phi_x(z)$ et $\phi_{Vrai}(z)$ les fonctions énumérant respectivement les littéraux et *Vrai*. Nous obtenons le système suivant :

$$\begin{cases} \phi_x(z) = z + \phi_{Vrai}(z)\phi_x(z) \\ \phi_{Vrai}(z) = \phi_x(z)^2 + \phi_x(z)\phi_{Vrai}(z) + \phi_{Vrai}(z)^2 \end{cases} \tag{4.1}$$

Par ailleurs, nous savons que les deux ensembles d'arbres calculant x et celui des arbres calculant *Vrai* sont disjoints et que leur union est l'ensemble de tous les arbres. Nous pouvons donc ajouter une équation supplémentaire au système :

$$\phi_x(z) + \phi_{Vrai}(z) = f_1(z) = \frac{1-\sqrt{1-4z}}{2}.$$

Cette équation, linéaire en les fonctions génératrices, nous permet de simplifier de façon considérable le système (4.1) que nous pourrons dès lors résoudre de façon exacte. Notons que $\phi_x(0) = \phi_{Vrai}(0) = 0$. Finalement :

$$\begin{cases} \phi_x(z) = -\frac{1}{4}\left(1+\sqrt{1-4z}-\sqrt{2+12z+2\sqrt{1-4z}}\right) = \frac{\sqrt{5}-1}{4} - \frac{5-\sqrt{5}}{20}\sqrt{1-4z} + O\left(1-4z\right) \\ \phi_{Vrai}(z) = \frac{1}{4}\left(3-\sqrt{1-4z}-\sqrt{2+12z+2\sqrt{1-4z}}\right) = \frac{3-\sqrt{5}}{4} - \frac{5+\sqrt{5}}{20}\sqrt{1-4z} + O\left(1-4z\right) \end{cases}$$

Ces solutions nous permettent de donner les valeurs exactes des probabilités de chaque fonction pour les deux modèles. En s'appuyant sur l'article [CFGG04] (basé sur les arbres *Et/Ou*), et rappelé dans le chapitre précédent (Théorème 1 [p. 15]), nous savons que pour une fonction f donnée, le probabilité de la fonction dans le modèle des grands arbres vaut $\mu_k(f) = \beta_f / \sum_g \beta_g$, et sa probabilité dans le modèle des processus de branchement $\pi_k(f) = \alpha_f / \sum_g \alpha_g$ – où les coefficients α_f et β_f sont ceux introduits dans le Théorème 1 [p. 15].

Fonction	Complexité	π_k	μ_k
x	1	$\frac{\sqrt{5}-1}{2} \approx 0.618$	$\frac{5-\sqrt{5}}{10} \approx 0.276$
Vrai	2	$\frac{3-\sqrt{5}}{2} \approx 0.382$	$\frac{5+\sqrt{5}}{10} \approx 0.724$

FIG. 4.2 – Valeurs exactes des probabilités de chaque fonction.

Nous nous rendons compte que, pour les deux distributions, chaque fonction a une probabilité non nulle. Toutefois, la fonction ayant la probabilité la plus élevée diffère selon le modèle d'arbres – grands arbres ou processus de branchement.

4.2 Système de l'implication avec deux variables

Intéressons-nous désormais au système contenant deux variables. Dans ce système, six fonctions sont expressibles. Cependant certaines fonctions Booléennes ont les mêmes fonctions génératrices. En utilisant la relation d'équivalence que nous avons présentée plus haut, nous obtenons 4 classes ayant pour représentants x, $x \vee \bar{y}$, $x \vee y$ et *Vrai*. Nous nous ramenons donc au système suivant :

$$\begin{cases} \phi_x(z) = z + \phi_{Vrai}(z)\phi_x(z) + \phi_{x\vee\bar{y}}(z)\phi_x(z) \\ \phi_{x\vee\bar{y}}(z) = \phi_x(z)^2 + \phi_{x\vee y}(z)\phi_x(z) + \phi_x(z)\phi_{x\vee\bar{y}}(z) \\ \qquad\qquad +\phi_{x\vee\bar{y}}(z)^2 + \phi_{x\vee y}(z)\phi_{x\vee\bar{y}}(z) + \phi_{Vrai}(z)\phi_{x\vee\bar{y}}(z) \\ \phi_{x\vee y}(z) = 2\phi_{x\vee\bar{y}}(z)\phi_x(z) + 2\phi_{x\vee\bar{y}}(z)\phi_{x\vee y}(z) + \phi_{Vrai}(z)\phi_{x\vee y}(z) \\ \phi_{Vrai}(z) = 2\phi_x(z)^2 + 2\phi_{x\vee\bar{y}}(z)^2 + \phi_{x\vee y}(z)^2 + 2\phi_x(z)\phi_{x\vee\bar{y}}(z) + 2\phi_x(z)\phi_{x\vee y}(z) \\ \qquad\qquad +\phi_x(z)\phi_{Vrai}(z) + \phi_{x\vee\bar{y}}(z)\phi_{Vrai}(z) + \phi_{x\vee y}(z)\phi_{Vrai}(z) + \phi_{Vrai}(z)^2 \end{cases} \tag{4.2}$$

Nous pouvons ajouter une équation supplémentaire au système, obtenue en comptabilisant le nombre de fonctions Booléennes appartenant à chaque classe d'équivalence :

$$2\phi_x(z) + 2\phi_{x\vee\bar{y}}(z) + \phi_{x\vee y}(z) + \phi_{Vrai}(z) = f_2(z) = \frac{1-\sqrt{1-8z}}{2}.$$

En procédant par élimination, on obtient une équation algébrique de degré 8 que nous n'avons pas résolue. Par conséquent, on utilise l'approximation de la fonction génératrice de f (énoncée au Théorème 1 [p. 15]). En évaluant ce nouveau système en la singularité $1/8$, commune à toutes les fonctions, on obtient un système polynômial de degré 2 en les α_g ayant une unique solution. Cette solution peut être approchée par itération en partant du vecteur nul. Une fois qu'on a une valeur approchée des α_g, il reste un système linéaire en les β_g à résoudre (en évaluant le système de départ en $z = 0$).

Le tableau suivant est composé des fonctions à deux variables ainsi que des valeurs numériques que nous avons obtenue via la méthode précédente. On a écrit un programme en Python de moins de 300 instructions qui construit le système puis obtient les α_g. Les itérations sont arrêtées lorsque la norme infinie de deux vecteurs (α_g) consécutifs est inférieure à 10^{-8}. Puis nous avons obtenu la valeurs des β_g en utilisant Maple. Enfin, les coefficients sont normalisés afin d'avoir les valeurs des probabilités.

Fonction	Complexité	π_k	μ_k
x	1	0.296	0.109
Vrai	2	0.242	0.519
$x \vee \bar{y}$	2	0.0698	0.102
$x \vee y$	3	0.0256	0.0583

FIG. 4.3 – Valeurs approchées des probabilités de chaque fonction.

A nouveau chaque fonction Booléenne a une probabilité non nulle dans chacun des modèles. Cependant, les fonctions de petite complexité semblent avoir une probabilité plus élevées.

4.3 Système de l'implication avec trois ou quatre variables

En utilisant la même démarche que pour le cas précédent (écriture du système de fonctions génératrices, puis résolution approchée des α_g puis des β_g), nous obtenons les tableaux suivants 4.4 et 4.5.

Nous observons que la complexité de la fonction semble liée à sa probabilité. Par ailleurs, en dehors des littéraux, nous remarquons que la probabilité d'une fonction dans le modèle de branchement est toujours inférieure à celle dans le modèle des grands arbres.

Nous avons souhaité obtenir des résultats numériques pour de plus grandes valeurs de k. Pour cela nous avons réécrit un programme en C (afin de gagner en rapidité) calculant le système algébrique. Dans le cas à 4 variables, les 942 fonctions se réduisent à 80 classes distinctes. De ce fait, la méthode précédente a permis d'obtenir des valeurs approchées des α_g et β_g. Les remarques précédentes pour le cas à 3 variables sont également vraies pour 4 variables.

Dans le cas à 5 variables, nous avons désormais 3984 classes de fonctions équivalentes. Ce nombre étant trop important, la méthode précédente n'est plus suffisamment rapide. Par

Fonction	Complexité	π_k	μ_k
x	1	0.191	0.0570
Vrai	2	0.173	0.386
$x \vee \overline{y}$	2	0.0264	0.0326
$x \vee \overline{y} \vee \overline{z}$	3	0.0112	0.0291
$x \vee y$	3	0.00626	0.0126
$x \vee (y \wedge \overline{z})$	3	0.00356	0.00789
$x \vee y \vee \overline{z}$	4	0.00383	0.0129
$x \vee (\overline{y} \wedge \overline{z})$	4	0.00169	0.00503
$x \vee (y \wedge z)$	4	0.00123	0.00395
$x \vee y \vee z$	5	0.00169	0.00658
$x \vee (y \oplus z)$	5	$1.40\ 10^{-4}$	$5.92\ 10^{-4}$
$x \vee (y \oplus \overline{z})$	7	$3.77\ 10^{-5}$	$2.07\ 10^{-4}$

FIG. 4.4 – Valeurs numériques des probabilités de chaque fonction à trois variables.

conséquent, le système a été écrit via un programme tournant en parallèle sur une machine à 6 microprocesseurs. Il a été obtenu après plusieurs dizaines d'heures de calculs et sa taille fait plus de 20 Gigaoctets. Une récente méthode développée par Pivoteau *et al.* [PSS08], et basée sur l'approximation de Newton, permet de déterminer les valeurs approchées des α_g par une méthode bien plus puissante que celles des itérations. Leur programme a permis de vérifier nos résultats numériques pour $k \leqslant 4$. Cependant, leur programme est pour le moment implémenté sous Maple et n'est pas en mesure de résoudre notre système à 5 variables. Toutefois, Darrasse est en train de transcrire la méthode en C, ce qui nous laisse espérer l'obtention des valeurs numériques pour $k = 5$.

Puisque les résultats numériques ne peuvent pas être obtenus pour des valeurs plus grandes, nous nous tournons désormais vers une étude théorique des distributions de probabilité pour chacun des deux modèles.

Fonction	π_k	μ_k	Fonction	π_k	μ_k
x	0.140	0.0344	$x \vee y \vee (z \oplus t)$	$4.11\ 10^{-6}$	$2.57\ 10^{-5}$
$Vrai$	0.132	0.299	$x \vee (y \wedge z \wedge \overline{t}) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$3.73\ 10^{-6}$	$2.05\ 10^{-5}$
$x \vee \overline{y}$	0.0133	0.0137	$x \vee (y \oplus \overline{z})$	$3.47\ 10^{-6}$	$1.60\ 10^{-5}$
$x \vee (y \wedge (\overline{z} \vee \overline{t}))$	0.00361	0.00805	$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge z \wedge \overline{t})$	$3.32\ 10^{-6}$	$1.86\ 10^{-5}$
$x \vee y$	0.00223	0.00381	$x \vee (y \wedge (z \vee \overline{t})) \vee (\overline{y} \wedge (\overline{z} \vee \overline{t}))$	$3.25\ 10^{-6}$	$2.27\ 10^{-5}$
$x \vee \overline{y} \vee \overline{z} \vee \overline{t}$	0.00207	0.00782	$x \vee (y \wedge (z \vee \overline{t})) \vee (\overline{y} \wedge z \wedge \overline{t})$	$2.66\ 10^{-6}$	$1.63\ 10^{-5}$
$x \vee (y \wedge \overline{z})$	0.00121	0.00223	$x \vee (y \wedge \overline{z} \wedge \overline{t}) \vee (\overline{y} \wedge (z \oplus \overline{t}))$	$2.39\ 10^{-6}$	$1.35\ 10^{-5}$
$x \vee y \vee \overline{z}$	0.000894	0.00264	$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge z)$	$2.23\ 10^{-6}$	$1.32\ 10^{-5}$
$x \vee y \vee \overline{z} \vee \overline{t}$	0.000666	0.00300	$x \vee (y \wedge (z \oplus t))$	$2.18\ 10^{-6}$	$1.39\ 10^{-5}$
$x \vee (\overline{y} \wedge \overline{z})$	0.000422	0.00108	$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \vee z \vee t)$	$2.00\ 10^{-6}$	$1.24\ 10^{-5}$
$x \vee (y \wedge z \wedge t)$	0.000366	0.00118	$x \vee (y \wedge \overline{z} \wedge \overline{t}) \vee (\overline{y} \wedge (\overline{z} \vee \overline{t}))$	$1.86\ 10^{-6}$	$1.27\ 10^{-5}$
$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge \overline{z})$	0.000356	0.00117	$x \vee (y \wedge \overline{z} \wedge \overline{t}) \vee (\overline{y} \wedge z \wedge \overline{t})$	$1.75\ 10^{-6}$	$1.17\ 10^{-5}$
$x \vee y \vee z$	0.000307	0.00107	$x \vee (y \wedge (\overline{z} \vee t)) \vee (\overline{y} \wedge z)$	$1.70\ 10^{-6}$	$1.00\ 10^{-5}$
$x \vee (y \wedge z)$	0.000285	0.000787	$x \vee (y \wedge (z \oplus \overline{t})$	$1.66\ 10^{-6}$	$1.07\ 10^{-5}$
$x \vee y \vee z \vee \overline{t}$	0.000282	0.00142	$x \vee y \vee (z \oplus \overline{t})$	$1.52\ 10^{-6}$	$1.14\ 10^{-5}$
$x \vee (y \wedge \overline{z} \wedge \overline{t})$	0.000183	0.000717	$x \vee (y \wedge (z \oplus \overline{t})) \vee (\overline{y} \wedge z \wedge t)$	$1.22\ 10^{-6}$	$7.49\ 10^{-6}$
$x \vee (y \wedge z \wedge t)$	0.000161	0.000690	$x \vee (y \wedge (\overline{z} \vee t)) \vee (\overline{y} \wedge \overline{t})$	$1.21\ 10^{-6}$	$7.44\ 10^{-6}$
$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge (\overline{z} \vee \overline{t}))$	0.000155	0.000631	$x \vee (y \wedge (z \oplus \overline{t})) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$1.17\ 10^{-6}$	$7.70\ 10^{-6}$
$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	0.000145	0.000452	$x \vee (y \wedge (z \vee \overline{t})) \vee (\overline{y} \wedge (z \oplus \overline{t}))$	$1.13\ 10^{-6}$	$6.98\ 10^{-6}$
$x \vee y \vee z \vee t$	0.000136	0.000746	$x \vee (y \wedge \overline{z} \wedge \overline{t}) \vee (\overline{y} \wedge z \wedge t)$	$1.01\ 10^{-6}$	$7.03\ 10^{-6}$
$x \vee (y \wedge (z \vee \overline{t})) \vee (\overline{y} \wedge \overline{t})$	0.000110	0.000475	$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge (\overline{z} \oplus t))$	$8.92\ 10^{-7}$	$5.63\ 10^{-6}$
$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge \overline{z} \wedge t)$	$9.91\ 10^{-5}$	0.000376	$x \vee (y \wedge z) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$7.36\ 10^{-7}$	$4.90\ 10^{-6}$
$x \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$9.21\ 10^{-5}$	0.000417	$x \vee (y \wedge (z \vee t)) \vee (\overline{y} \wedge z \wedge t)$	$5.78\ 10^{-7}$	$4.39\ 10^{-6}$
$x \vee y \vee (z \wedge \overline{t})$	$7.82\ 10^{-5}$	0.000304	$x \vee y \vee \overline{z}$	$4.68\ 10^{-7}$	$4.77\ 10^{-6}$
$x \vee (y \wedge \overline{z} \wedge \overline{t}) \vee (\overline{y} \wedge \overline{z})$	$6.45\ 10^{-5}$	0.000255	$x \vee (y \wedge z \wedge t) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$3.70\ 10^{-7}$	$2.87\ 10^{-6}$
$x \vee y \vee (\overline{z} \wedge \overline{t})$	$4.62\ 10^{-5}$	0.000212	$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge (z \oplus t))$	$3.61\ 10^{-7}$	$2.76\ 10^{-6}$
$x \vee (y \wedge (\overline{z} \vee \overline{t}))$	$4.31\ 10^{-5}$	0.000180	$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge z \wedge t)$	$3.52\ 10^{-7}$	$2.56\ 10^{-6}$
$x \vee (y \wedge (z \vee \overline{t}))$	$3.79\ 10^{-5}$	0.000154	$x \vee (y \wedge (\overline{z} \vee t)) \vee (\overline{y} \wedge z \wedge \overline{t})$	$2.23\ 10^{-7}$	$1.68\ 10^{-6}$
$x \vee y \vee (z \wedge t)$	$3.51\ 10^{-5}$	0.000166	$x \vee (y \wedge (z \oplus \overline{t})) \vee (\overline{y} \wedge \overline{t})$	$2.05\ 10^{-7}$	$1.66\ 10^{-6}$
$x \vee (y \wedge (\overline{z} \vee \overline{t}) \vee (\overline{y} \wedge z \wedge \overline{t})$	$2.45\ 10^{-5}$	$9.24\ 10^{-5}$	$x \vee (y \wedge (z \vee t)) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$8.82\ 10^{-8}$	$7.46\ 10^{-7}$
$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge z)$	$2.17\ 10^{-5}$	$7.59\ 10^{-5}$	$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge (z \oplus \overline{t}))$	$6.69\ 10^{-8}$	$5.65\ 10^{-7}$
$x \vee (y \wedge (z \vee t)$	$1.65\ 10^{-5}$	$8.04\ 10^{-5}$	$x \vee (y \wedge (z \vee t)) \vee (\overline{y} \wedge (z \oplus \overline{t}))$	$6.03\ 10^{-8}$	$6.48\ 10^{-8}$
$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge (z \vee \overline{t}))$	$1.58\ 10^{-5}$	$8.61\ 10^{-5}$	$x \vee (y \wedge (z \oplus t)) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$4.02\ 10^{-8}$	$3.43\ 10^{-7}$
$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge \overline{t})$	$1.07\ 10^{-5}$	$5.04\ 10^{-5}$	$x \vee (y \wedge (z \vee \overline{t})) \vee (\overline{y} \wedge (z \oplus t))$	$2.17\ 10^{-8}$	$2.09\ 10^{-7}$
$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$7.28\ 10^{-6}$	$3.78\ 10^{-5}$	$x \vee (y \wedge (z \oplus t)) \vee (\overline{y} \wedge z \wedge t)$	$2.04\ 10^{-8}$	$1.95\ 10^{-7}$
$x \vee (y \wedge \overline{z}) \vee (\overline{y} \wedge z \wedge t)$	$6.52\ 10^{-6}$	$3.18\ 10^{-5}$	$x \vee (y \wedge (z \oplus \overline{t})) \vee (\overline{y} \wedge z \wedge \overline{t})$	$1.64\ 10^{-8}$	$1.61\ 10^{-7}$
$x \vee (y \wedge (z \vee \overline{t}) \vee (\overline{y} \wedge z \wedge t))$	$6.51\ 10^{-6}$	$3.09\ 10^{-5}$	$x \vee (y \wedge (z \oplus \overline{t})) \vee (\overline{y} \wedge (\overline{z} \vee \overline{t}))$	$1.35\ 10^{-8}$	$1.42\ 10^{-7}$
$x \vee (y \wedge z) \vee (\overline{y} \wedge z \wedge \overline{t})$	$5.84\ 10^{-6}$	$2.90\ 10^{-5}$	$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge (z \oplus t))$	$1.24\ 10^{-8}$	$1.28\ 10^{-7}$
$x \vee (y \wedge (z \vee \overline{t})) \vee (\overline{y} \wedge \overline{z} \wedge \overline{t})$	$4.98\ 10^{-6}$	$2.55\ 10^{-5}$	$x \vee (y \wedge (z \oplus t)) \vee (\overline{y} \wedge (z \oplus \overline{t}))$	$5.4\ 10^{-10}$	$6.35\ 10^{-9}$
$x \vee (y \wedge (\overline{z} \vee \overline{t})) \vee (\overline{y} \wedge (z \oplus t))$	$4.25\ 10^{-6}$	$2.37\ 10^{-5}$	$x \vee (y \wedge (z \oplus \overline{t})) \vee (\overline{y} \wedge (z \oplus t))$	$2.4\ 10^{-10}$	$3.01\ 10^{-9}$

FIG. 4.5 – Valeurs numériques des probabilités de chaque fonction à quatre variables.

Chapitre 5

Grands arbres

Le but de ce chapitre est d'étudier une sous-famille des expressions calculant une fonction donnée, dans le modèle des grands arbres. Nous allons établir que lorsque le nombre de variables k tend vers l'infini, la plupart des expressions calculant cette fonction fixée, fait partie de cette famille. Par ailleurs les expressions de cette famille comportent un motif qui est un des arbres minimaux de la fonction auquel on greffe un fils gauche ne modifiant pas la fonction calculée par le motif. Pour cette raison, nous dirons que les expressions de cette famille sont simples.

Les expansions valides d'un arbre. Nous allons définir trois règles, les *règles d'expansion* d'un arbre qui permettent, à partir d'un arbre A, d'obtenir une famille d'arbres, de taille supérieure à celle de A, qui calculent la même fonction que A. Soient A un arbre, B un de ses sous-arbres et notons ν la racine de B.

La première expansion est appelée *expansion valide avec "tautologie"*. Soit A' l'arbre obtenu en remplaçant B par $C \to B$ dans A, où C est une tautologie (une expression calculant toujours vrai). Il est clair que A' calcule la même fonction que A puisque $[C \to B] = [B]$.

La seconde expansion est appelée *expansion valide avec "but α"*. Si le fait de remplacer B par un arbre $C \to B$ donne un arbre A' calculant $[A]$ pour tout arbre C de but α, alors nous dirons que les arbres A' sont obtenus à partir de A via une expansion valide de type "but α" au noeud ν.

La troisième expansion est appelée *expansion valide avec "prémisse α"*. Si le fait de remplacer B par un arbre $C \to B$ donne un arbre A' calculant $[A]$ pour tout arbre C contenant une prémisse réduite à une feuille étiquetée par α, alors nous dirons que les arbres A' sont obtenus à partir de A via une expansion valide de type "prémisse α" au noeud ν.

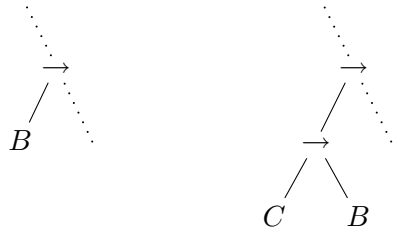

FIG. 5.1 – Expansion valide avec le sous-arbre C dans la racine de B.

La Figure 5.1 représente la structure de l'arbre obtenu après une expansion valide en la racine de B. Etant donné un arbre A, nous noterons $E(A)$ l'ensemble des arbres obtenus à

partir de A avec une seule expansion valide, d'un des trois types que nous venons de définir. Remarquons que tous les arbres de $E(A)$ calculent la même fonction que A. Nous étendons la définition de E à toute famille d'arbres $\mathcal{A} \subseteq \mathcal{F}_k$ en prenant : $E(\mathcal{A}) = \bigcup_{A \in \mathcal{A}} E(A)$. Par ailleurs, nous définissons $E^0(\mathcal{A}) = \mathcal{A}$, $E^i(\mathcal{A}) = E(E^{i-1}(\mathcal{A}))$ et enfin $E^*(\mathcal{A}) = \bigcup_{i \in \mathbb{N}} E^i(\mathcal{A})$.

Présentons deux exemples (Figure 5.3) d'arbres obtenus après une expansion valide de la fonction Booléenne $f \sim x_1 \vee x_3 \vee \bar{x}_2$ – la Figure 5.2 en représente une formule. Les expansions ont les deux lieu dans la prémisse soulignée, au départ réduite à x_2.

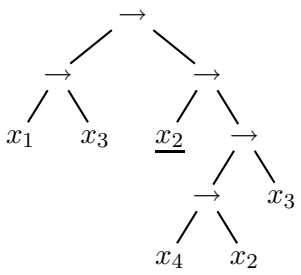

FIG. 5.2 – Une formule A représentant $f \sim x_1 \vee x_3 \vee \bar{x}_2$.

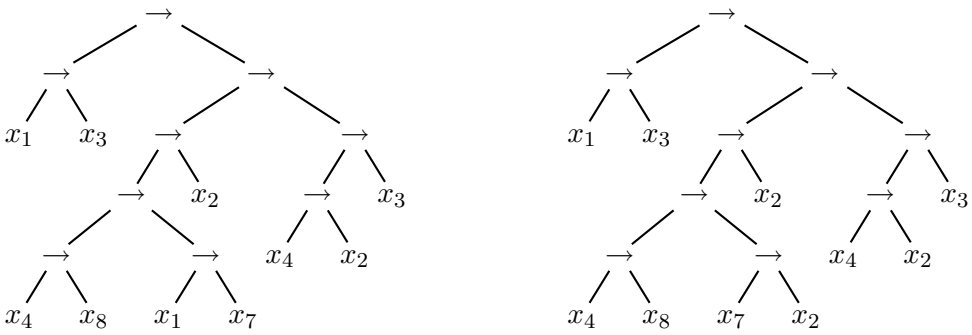

FIG. 5.3 – Deux arbres obtenus par des expansions valides de A de type "prémisse x_1" (à gauche), et de type "but x_2" (à droite).

Etant donné un arbre A, nous définissons $\lambda(A)$ comme le nombre d'expansions valides de A ; plus précisément, il s'agit de la somme des trois nombres suivants : le nombre de paires (ν, α), où ν est un noeud de A (soit un noeud interne, soit une feuille) et $\alpha \in \{x_1, \ldots, x_k\}$ est un littéral tel qu'une expansion de type "but α" soit valide en ν ; le nombre de paires (ν, α), où ν est un noeud de A (soit un noeud interne, soit une feuille) et $\alpha \in \{x_1, \ldots, x_k\}$ tel qu'une expansion de type "prémisse α" soit valide en ν ; le nombre $2|A| - 1$ qui compte le nombre d'expansions valides de type "tautologie" dans A – il y en a autant que de noeuds dans l'arbre.

Pour une fonction Booléenne f dépendant d'un nombre fini de variables, nous définissons $\lambda(f)$ comme la somme de tous les $\lambda(M)$, lorsque M parcourt l'ensemble $\mathcal{M}_f$ des arbres minimaux de f. Nous prouverons que $\lambda(f)$ ne dépend pas du nombre total k de variables.

Nous sommes désormais en mesure d'étudier les distributions de probabilité étant donné le modèle des grands arbres.

Théorème 2 *Pour le modèle des grands arbres, la plupart des tautologies ont une structure*

simple et par conséquent :

$$\mu_k(\text{Vrai}) = \frac{1}{k} + O\left(\frac{1}{k^2}\right).$$

Soit f une fonction Booléenne distincte de Vrai. *Pour le modèle des grands arbres, la plupart des arbres calculant f sont obtenus après une unique expansion valide des arbres minimaux de f :*

$$\mu_k(f) = \mu_k(E(\mathcal{M}_f)) + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Etant capable de calculer la fraction limite $\mu_k(E(\mathcal{M}_f))$, nous en concluons que la probabilité de f est asymptotiquement (lorsque $k \to \infty$) égale à :

$$\mu_k(f) = \frac{\lambda(f)}{4^{L(f)}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Le Théorème 2 sera démontré dans la Section 5.3 ; nous nous intéresserons dans la section suivante à la fonction *Vrai*, puis, Section 5.2, aux autres fonctions.

5.1 Fraction limite des tautologies

Le but de cette section est d'obtenir la probabilité de la fonction *Vrai*. Pour ce faire, nous allons définir trois ensembles d'expressions regroupant "presque toutes" les expressions, quand k tend vers l'infini.

5.1.1 Expressions tautologiques et non-tautologiques

Dans cette partie nous allons définir trois ensembles d'expressions que nous énumérerons dans la partie suivante.

Définition 1 *Définissons les sous-ensembles de $\mathcal{F}_k$ suivants :*

- *Cl_k est l'ensemble des* tautologies classiques *, i.e. les expressions s'évaluant à vrai pour toute affectation des variables.*

- *SN_k est l'ensemble des* non-tautologies simples *, du type*

  $$A = A_1, \ldots, A_p \to r(A),$$

  tel que pour tout i, $r(A_i) \neq r(A)$.

- *G_k est l'ensemble des* tautologies simples *, du type*

  $$A = A_1, \ldots, A_p \to r(A),$$

  où il existe i, tel que la prémisse A_i est égale à la variable $r(A)$.

- *LN_k est l'ensemble des* non-tautologies moins simples *, du type*

  $$A = B_1, \ldots, B_{i-1}, C, B_i, \ldots, B_p \to r(A),$$

tel que

$$C = C_1, C_2, \ldots, C_q \to r(C),$$

où $r(C) = r(A)$, $q \geqslant 1$, *et*

$$C_1 = D_1, D_2, \ldots, D_s \to r(D),$$

où $r(D) \neq r(A)$, $s \geqslant 0$, *et les conditions suivantes sont satisfaites : pour tout* j, $r(B_j) \notin \{r(A), r(D)\}$ *et* $r(D_j) \notin \{r(A), r(D)\}$.

Le fait d'ajouter un exposant n *au nom de ces ensembles signifie qu'on ne considère que les expressions de taille* n *appartenant à ces ensembles.*

Remarquons que les éléments de SN_k ne sont pas des tautologies. En effet, soit $A \in SN_k$, il suffit d'évaluer $r(A)$ à 0 et toutes les autres variables à 1 afin que A s'évalue à 0 – chacune des prémisses de A s'évalue à 1 puisque leurs buts sont distincts de $r(A)$. Par ailleurs, les expressions de G_k sont des tautologies. Il suffit d'évaluer successivement le but de ces expressions à 1 et à 0 et de s'apercevoir que dans les deux cas l'expression s'évalue à 1. Enfin, les expressions de LN_k ne sont pas des tautologies. En effet, soit $A \in LN_k$. Il suffit d'évaluer $r(A)$ et $r(D)$ à 0 et toutes les autres variables à 1 pour obtenir une évaluation de A à 0.

Etablissons quelques relations immédiates vérifiées par ces ensembles.

$$\begin{aligned} SN_k \cup LN_k &\subset \mathcal{F}_k \setminus Cl_k, \\ SN_k \cap LN_k &= \emptyset, \\ G_k \subsetneq Cl_k &\subsetneq \mathcal{F}_k \setminus (SN_k \cup LN_k). \end{aligned}$$

5.1.2 Enumération des trois familles

Nous allons énumérer les trois classes d'expressions : SN_k, G_k et LN_k. Le calcul des fractions limites est fait d'une manière systématique. Nous établissons pour chaque ensemble une construction récursive à partir d'ensembles plus simples : ceci nous permet d'écrire les fonctions génératrices énumérant les expressions de ces ensembles par rapport à leur taille (nombre de feuilles) et d'obtenir ensuite la fonction génératrice ϕ énumérant la classe en question. Par la suite nous extrayons le coefficient $[z^n]\phi(z)$ et, afin d'obtenir la fraction limite de la classe, nous calculons $\lim_{n\to\infty}[z^n]\phi(z)/[z^n]f_k(z)$ – rappelons que f_k est la fonction génératrice énumérant toutes les expressions du système.

Non-tautologies simples

Considérons tout d'abord l'ensemble SN_k des expressions simples qui ne sont pas des tautologies. Si $A \in SN_k$, alors A est du type

$$A = A_1, \ldots, A_p \to r(A),$$

tel que pour tout i, $r(A_i) \neq r(A)$.

Etablissons la fonction génératrice $SN(z)$ associée à SN_k. Etant donné une variable Booléenne α, prenons en compte tous les arbres de but α. Un tel arbre est une non-tautologie si et seulement si toutes ses prémisses A_i vérifient $r(A_i) \neq \alpha$. La fonction génératrice énumérant toutes les prémisses possibles vaut $\frac{k-1}{k}f_k(z)$. Une non-tautologie simple est une suite de

telles prémisses suivie d'une feuille α. Ainsi la fonction génératrice SN^α énumérant les non-tautologies simples de but α est égale à

$$SN^\alpha(z) = \frac{z}{1 - \frac{k-1}{k} f_k(z)}.$$

Puisque la variable α peut être choisie de façon arbitraire parmi les k littéraux, nous avons $SN(z) = k \cdot SN^\alpha(z)$, ce qui donne

$$SN(z) = \frac{kz}{1 - \frac{k-1}{k} f(z)}.$$

Proposition 1 *La fraction limite des non-tautologies simples existe et vaut*

$$\mu_k(SN_k) = \frac{k(k-1)}{(k+1)^2}.$$

Lorsque k est grand, cette fraction limite est égale à $1 - 3/k + O(1/k^2)$.

Preuve: Ce résultat a déjà été présenté dans l'article de Moczurad *et al.* [MTZ00, page 586], avec une preuve différente de celle que nous présentons ici. Si la fraction limite existe, elle est donnée par la formule suivante :

$$\mu_k(SN_k) \quad = \quad \lim_{n\to\infty} \frac{|SN_k^n|}{|\mathcal{F}_k^n|} = \lim_{n\to\infty} \frac{[z^n]SN(z)}{[z^n]f_k(z)}.$$

Après modification du dénominateur de $SN(z)$, nous obtenons :

$$SN(z) = \frac{k(k+1)z + kz(1-k)\sqrt{1-4kz}}{2(1+z(k-1)^2)}.$$

Le dénominateur de la fraction rationnelle $SN(z)$ admet un unique zéro $\rho = -1/(k-1)^2$. Cependant cette valeur annule également le numérateur puisque

$$k(k+1)\rho + k(1-k)\rho\sqrt{(-\rho)((k-1)^2+4k)} = 0.$$

Ainsi ρ n'est pas un pôle de $SN(z)$. Par conséquent, l'unique singularité qui compte asymptotiquement est $z = 1/(4k)$. Nous obtenons

$$[z^n]SN(z) = -\frac{2k^2(k-1)}{(k+1)^2}[z^{n-1}]\sqrt{1-4kz}\,(1+O(1/n))\,.$$

Rappelons,

$$[z^n]\sqrt{1-4kz} = -2k^n C_{n-1}. \tag{5.1}$$

Ceci nous permet de conclure :

$$[z^n]SN(z) = \frac{4k(k-1)}{(k+1)^2} k^n C_{n-2}\,(1+O(1/n))\,.$$

En outre, rappelons que pour tous i et j strictement positifs,

$$\lim_{i\to\infty} \frac{C_i}{C_{i+j}} = \frac{1}{4^j}. \tag{5.2}$$

Ceci, associé au résultat précédent donne

$$\lim_{n\to\infty} \frac{|SN_k^n|}{|\mathcal{F}_k^n|} = \frac{4k(k-1)}{(k+1)^2} \lim_{n\to\infty} \frac{C_{n-2}}{C_{n-1}} = \frac{k(k-1)}{(k+1)^2},$$

ainsi, la fraction limite $\mu_k(SN_k)$ existe et est égale à $k(k-1)/(k+1)^2$. □

Tautologies simples

Si A est une tautologie simple, alors A s'écrit

$$A = A_1, \ldots, A_p \to r(A),$$

avec un des sous-arbres A_i égal à $r(A)$.

Déterminons la fonction génératrice énumérant les tautologies simples. Un arbre A n'est pas une tautologie simple si et seulement si toutes ses prémisses sont distinctes de $r(A)$ – voir Figure 5.4. La fonction génératrice pour l'ensemble $\mathcal{F}_k \setminus G_k$ est égale à $kz/(1-(f_k(z)-z))$. Par conséquent, la fonction génératrice de G_k est

$$G(z) = f_k(z) - \frac{kz}{1+z-f_k(z)}.$$

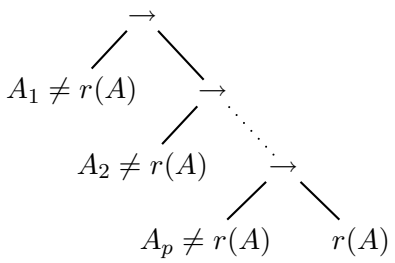

FIG. 5.4 – Arbres qui ne sont pas des tautologies simples.

Proposition 2 *La fraction limite des tautologies simples à k variables existe et vaut*

$$\mu_k(G_k) = \frac{4k+1}{(2k+1)^2}.$$

Lorsque k est grand, cette fraction limite est égale à $1/k - 3/4k^2 + O(1/k^3)$.

Preuve: A nouveau ce résultat avait été présenté avec une preuve différente dans l'article [MTZ00, page 584]. La fonction génératrice $G(z)$ peut être écrite ainsi :

$$G(z) = \frac{P(z) - (1+z)\sqrt{1-4kz}}{2(1+k+z)},$$

où $P(z)$ est un polynôme adéquat. Soit ρ le pôle de la fraction rationnelle ; $\rho = -1-k$. Mais ρ est de module plus grand que celui de la singularité algébrique $1/(4k)$; ainsi $1/(4k)$ est la singularité dominante de $G(z)$. Finalement, nous obtenons

$$[z^n]G(z) = \left(-\frac{2k}{(2k+1)^2}[z^n]\sqrt{1-4kz} - \frac{2k}{(2k+1)^2}[z^{n-1}]\sqrt{1-4kz} \right)(1+O(1/n)).$$

L'équation (5.1 [p. 27]) donne

$$[z^n]G(z) = \left(\frac{4k}{(2k+1)^2}k^n C_{n-1} + \frac{4}{(2k+1)^2}k^n C_{n-2} \right)(1+O(1/n)).$$

Démontrons l'existence et calculons la fraction limite de G_k^n.

$$\lim_{n\to\infty} \frac{|G_k^n|}{|F_k^n|} = \frac{4k}{(2k+1)^2} + \frac{1}{(2k+1)^2}.$$

D'après l'équation (5.2), la fraction limite $\mu_k(G_k)$ existe et est égale à $(4k+1)/(2k+1)^2$. Pour k suffisamment grand, elle est égale à $1/k + O(1/k^2)$. □

Non-tautologies moins simples

Dans la famille SN_k des non-tautologies simples, nous n'avons pas considéré les arbres avec une prémisse ayant pour but le même but que le but global de l'arbre. Dans la famille LN_k, nous considérons des arbres avec exactement une telle prémisse.

Rappelons qu'un arbre A est une non-tautologie moins simple s'il est du type

$$A = B_1, \ldots, B_{i-1}, C, B_i, \ldots, B_p \to r(A),$$

où $C = C_1, \ldots, C_q \to r(C)$, avec $r(C) = r(A)$, $q \geqslant 1$, et $C_1 = D_1, D_2, \ldots, D_s \to r(D)$ est tel que $r(D) \neq r(A)$, $s \geqslant 0$, enfin la condition suivante est réalisée : pour tout j, $r(B_j) \notin \{r(A), r(D)\}$ et $r(D_j) \notin \{r(A), r(D)\}$. Voir la représentation gauche de la Figure 5.5 pour la forme générale de l'arbre la représentation de droite pour le sous-arbre C. Sur ces représentations, si un sous-arbre S est souligné, cela signifie qu'il est soumis à la contrainte $r(S) \notin \{r(A), r(D)\}$.

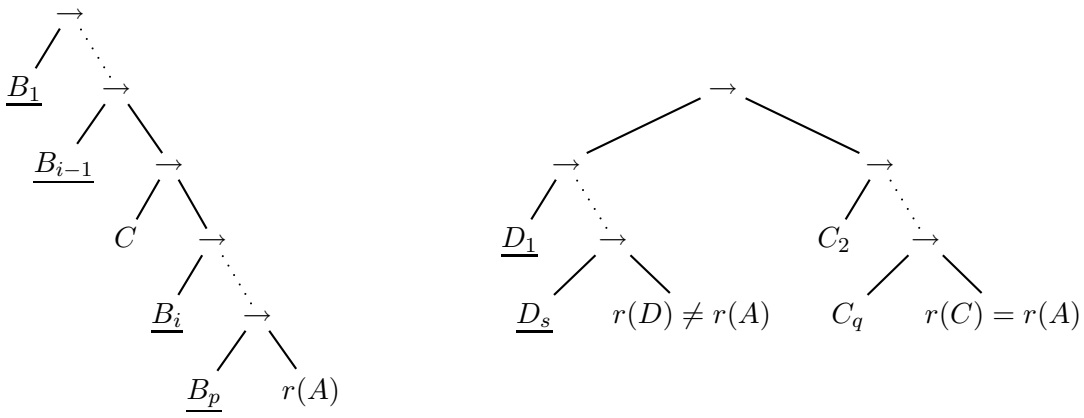

FIG. 5.5 – A gauche la forme générale de A et à droite son sous-arbre C.

Calculons la fonctions génératrice associée à LN_k. Fixons deux variables distinctes α et β. Nous allons tout d'abord déterminer la fonction génératrice énumérant les arbres de $LN_k^{\alpha,\beta}$, sous-ensemble de LN_k tel que $r(A) = \alpha$ et $r(D) = \beta$. Par symétrie, $\psi(z)$ est indépendante des choix de α et β.

Soit $b(z)$ la fonction génératrice des arbres $A \in \mathcal{F}_k$ tels que $r(A) \notin \{\alpha, \beta\}$. Nous avons $b(z) = (k-2)/k \cdot f_k(z)$. Cette fonction génératrice énumère les sous-arbres B_j mais aussi les éventuels D_j. Ainsi, la fonction génératrice pour l'ensemble des arbres D est $d(z) = z/(1-b(z))$, puisqu'il s'agit d'une suite d'arbres D_j tels que $r(D_j) \notin \{\alpha, \beta\}$, suivis d'une feuille β. De la même manière, la fonction génératrice associée au sous-arbre C est $c(z) = d(z) \cdot 1/(1-f_k(z)) \cdot z$. Remarquons qu'un arbre de $LN_k^{\alpha,\beta}$ est construit comme une suite

d'arbres B_j tels que $r(B_j) \notin \{\alpha, \beta\}$, suivi d'un sous-arbre C décrit précédemment, suivi d'une seconde suite d'arbres B_j avec $r(B_j) \notin \{\alpha, \beta\}$, suivis par une feuille α. Du fait que cette décomposition est unique, il en découle que la fonction génératrice pour $LN_k^{\alpha,\beta}$ est égale à

$$\psi(z) = \frac{1}{1-b(z)} c(z) \frac{1}{1-b(z)} z.$$

Il est désormais clair que la famille LN_k est l'union disjointe des $LN_k^{\alpha,\beta}$. Ainsi, étant donné un arbre $A \in LN_k$, alors α est le but de A, la prémisse C de A est définie de manière unique (puisqu'il s'agit de la seule prémisse de but égal à $r(A)$). Par conséquent, β est définie de manière unique puisqu'il s'agit du but de la première prémisse de C. Finalement $\phi(z) = k(k-1)\psi(z)$.

Proposition 3 *La fraction limite des non-tautologies moins simples vaut*

$$\mu_k(LN_k) = \frac{2k(k-1)^2}{(k+2)^4}.$$

Lorsque k est grand, cette fraction limite vaut $2/k + O(1/k^2)$.

Preuve: Après modification du dénominateur de la fonction génératrice $\phi(z)$, nous obtenons :

$$\phi(z) = \frac{P(z) + k(k-1)(-k^2 + (2k^3 - 6k^2 + 8)z)z^2\sqrt{1-4kz}}{2(2+(k-2)^2 z)^3},$$

où $P(z)$ est un polynôme adéquat. Le dénominateur de la fraction rationnelle $\phi(z)$ a un zéro $\rho = -2/(k-2)^2$. Cependant cette valeur annule également le numérateur de l'expression (et ses deux premières dérivées). Par conséquent, ρ n'est pas un pôle de l'expression. Ainsi, la seule singularité qui importe asymptotiquement est $z = 1/(4k)$. Nous obtenons :

$$\begin{aligned}[z^n]LN(z) &= \left(-\frac{k^3(k-1)}{2(2+\frac{(k-2)^2}{4k})^3}[z^{n-2}]\sqrt{1-4kz} \right. \\ &\quad \left. + \frac{k(k-1)(2k^3-6k^2+8)}{2(2+\frac{(k-2)^2}{4k})^3}[z^{n-3}]\sqrt{1-4kz} \right)(1+O(1/n)).\end{aligned}$$

En utilisant l'équation (5.1 [p. 27]) nous avons (en écartant le terme d'erreur) :

$$[z^n]LN(z) = \left(\frac{k^{n+1}(k-1)}{(2+\frac{(k-2)^2}{4k})^3} C_{n-3} - \frac{k^{n-2}(k-1)(2k^3-6k^2+8)}{(2+\frac{(k-2)^2}{4k})^3} C_{n-4} \right)(1+O(1/n)).$$

Démontrons l'existence de la fraction limite de LN_k et calculons-la.

$$\lim_{n\to\infty} \frac{|LN_k^n|}{|F_k^n|} = \frac{4k^4(k-1) - k(k-1)(2k^3-6k^2+8)}{(k+2)^6} = \frac{2k(k-1)^2}{(k+2)^4}.$$

La fraction limite $\mu_k(LN_k)$ existe et vaut $2k(k-1)^2/(k+2)^4$. Pour k suffisamment grand, elle est égale à $2/k + O(1/k^2)$. □

Les calculs que nous venons de présenter montrent que les trois familles d'expressions recouvrent l'ensemble des expressions jusqu'à l'ordre 2, lorsque k tend vers l'infini. Regroupons ces résultats sur la Figure 5.6

Ces résultats nous donnent le premier ordre de la probabilité de la fonction *Vrai*, lorsque le nombre k de variables devient grand. Intéressons-nous désormais aux autres fonctions.

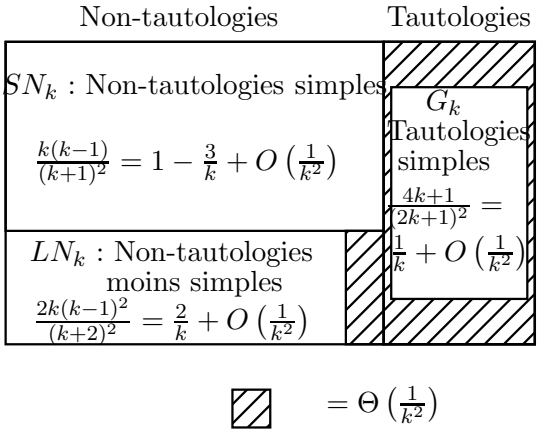

FIG. 5.6 – Couverture des expressions jusqu'à l'ordre 2.

5.2 Fraction limite des fonctions autres que *Vrai*

Nous dirons qu'un sous-arbre d'un arbre A est un *sous-arbre gauche* s'il est le fils gauche d'un noeud de A. Par ailleurs, nous définissons la *profondeur gauche* d'un noeud de A comme le nombre d'arêtes gauches nécessaires pour rejoindre le noeud en question depuis la racine de A. La Figure 5.7 représente un arbre dont les noeuds sont étiquetés par la valeur de leur profondeur gauche. Enfin, nous parlerons de profondeur gauche d'un sous-arbre comme étant la profondeur gauche de sa racine.

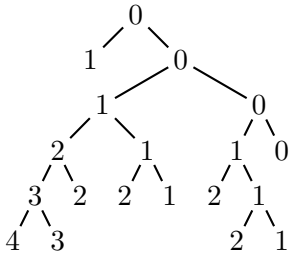

FIG. 5.7 – Noeuds étiquetés par la valeur de leur profondeur gauche.

5.2.1 Expansion et élagage

Nous allons, dans cette partie, étudier les règles d'expansions énoncées dans l'introduction du chapitre. Etant donné un arbre A et un de ses sous-arbres gauches B, nous notons $A \setminus B$ l'arbre obtenu après avoir supprimé B. Plus précisément, du fait que B est un sous-arbre gauche, il existe dans A un sous-arbre C tel que $B \to C$ est un sous-arbre. L'arbre $A \setminus B$ est obtenu en remplaçant $B \to C$ par C dans A – voir Figure 5.8. Les trois lemmes suivants donnent des conditions (nécessaires et) suffisantes pour qu'un arbre soit obtenu après une expansion valide d'un arbre plus petit.

Lemme 1 *Soient A un arbre et B un de ses sous-arbres gauches. Si B est une tautologie, alors A est obtenu par une expansion valide de type "tautologie" de $A \setminus B$.*

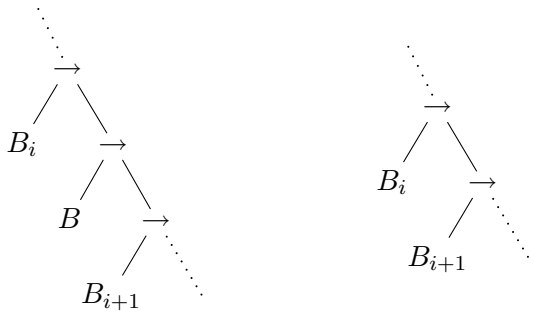

FIG. 5.8 – Suppression du sous-arbre B.

Preuve: Ceci est immédiat du fait de la définition d'une expansion valide de type "tautologie". □

Lemme 2 *Soient A un arbre et B un de ses sous-arbres gauches. Soit β le but de B. Si la substitution de B par 1 ou β dans A donne un arbre calculant $[A]$ dans les deux cas, alors A est obtenu par une expansion valide de type "but α" de $A \setminus B$.*

Preuve: Soit A_1 l'arbre A où B est remplacé par β, et A_2 l'arbre A où B est remplacé par 1. Soit B' un arbre quelconque avec un but β, et A' l'arbre obtenu à partir de A où B est remplacé par B'. Evidemment on a $\beta \leqslant [B'] \leqslant 1$. Puis, par récursion sur la taille de la formule, nous obtenons $[A] = [A_1] \leqslant [A'] \leqslant [A_2] = [A]$ ou $[A] = [A_1] \geqslant [A'] \geqslant [A_2] = [A]$, suivant que la profondeur gauche de la racine de B soit paire ou impaire. Dans tous les cas, $[A'] = [A]$. De plus, $[A \setminus B] = [A]$ puisque $[A \setminus B] = [A_2]$. □

Lemme 3 *Soient A un arbre et B un de ses sous-arbres gauches. Supposons que B ait une prémisse de taille 1 et étiquetée par β. Si la substitution de B par 1 ou $\bar{\beta}$ dans A donne un arbre calculant $[A]$ dans les deux cas, alors A est obtenu par une expansion valide de type "prémisse β" de l'arbre $A \setminus B$.*

Preuve: La démonstration est analogue à la précédente. □

Lorsqu'à partir de A on obtient $A \setminus B$ avec l'un des trois lemmes précédents, nous dirons que $A \setminus B$ est obtenu après élagage du sous-arbre gauche B dans A. A certains moments, nous aurons besoin de supprimer un sous-arbre dans un arbre donné. Voilà la différence entre suppression et élagage dans un arbre. Soient A un arbre et B un de ses sous-arbres gauches. Nous dirons que le sous-arbre B est *supprimé* de A lorsque nous enlèverons B sans aucune condition sur B. Nous dirons que le sous-arbre B est *élagué* de A lorsque A est un expansion valide de $A \setminus B$. Cependant, les deux arbres finaux seront notés $A \setminus B$.

Un arbre ne pouvant pas être élagué est appelé *arbre irréductible* . Evidemment, les arbres minimaux d'une fonction f sont irréductibles. Cependant, l'implication inverse n'est pas vraie. En effet, considérons la fonction $f = x_1 \vee (\bar{x}_2 \wedge \bar{x}_3) \vee (\bar{x}_2 \wedge x_4)$. On peut vérifier que l'expression $(x_4 \to x_2) \to (((x_2 \to x_3) \to x_3) \to x_1)$ calcule f et est irréductible. Toutefois elle n'est pas minimale puisque $((x_3 \to x_4) \to x_2) \to x_1$ est plus petite et calcule aussi f. Enfin, nous faisons remarquer que l'ensemble des règles d'élagage n'est pas confluent.

Définissons désormais une méthode afin d'obtenir de plus grands arbres à partir de plus petits. Cette fois, les arbres obtenus n'ont pas besoin de calculer la même fonction que celle calculée par l'arbre de départ ; le but de cette méthode est d'obtenir une borne supérieure sur la fraction limite d'une suite d'expansions. Cette nouvelle application X est nommée *extension* (elle est donc différente de l'expansion). L'application X est définie récursivement comme suivant : pour un arbre A étant une unique feuille α, $X(\alpha)$ est l'ensemble des arbres dont le but est étiqueté par α. Si $A = L \to R$, nous définissons

$$X(L \to R) = \left\{ A_1 \to \left(\cdots \to \left(A_p \to \left(\tilde{L} \to \tilde{R} \right) \right) \cdots \right) \mid A_1, \ldots, A_p \in \mathcal{F}_k,\ \tilde{L} \in X(L),\ \tilde{R} \in X(R) \right\}.$$

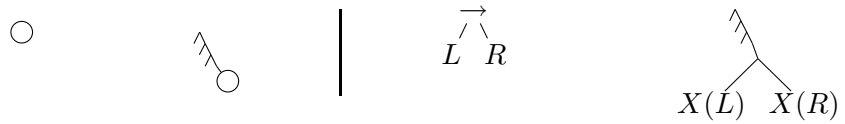

FIG. 5.9 – La définition récursive de l'application *extension*.

La Figure 5.9 présente graphiquement la définition récursive de l'application X et la Figure 5.10 donne la forme générale d'extensions d'un arbre donné : les fils gauches des peignes sont des arbres quelconques (non représentés pour ne pas surcharger la figure).

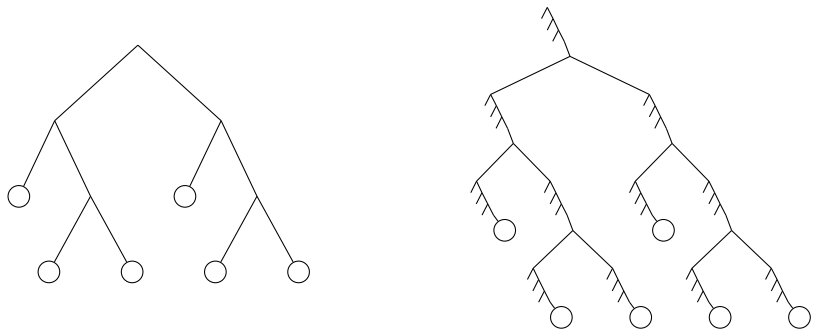

FIG. 5.10 – Un arbre A sur la gauche et l'ensemble $X(A)$ qu'il définit, sur la droite.

Nous étendons de façon naturelle X aux ensembles d'arbres $\mathcal{A} \subseteq \mathcal{F}_k$ en posant $X(\mathcal{A}) = \bigcup_{A \in \mathcal{A}} X(A)$. Remarquons que $X(X(\mathcal{A})) = X(\mathcal{A})$ pour tout ensemble $\mathcal{A} \subseteq \mathcal{F}_k$. La relation entre extensions et expansions est donnée ci-dessous.

Lemme 4 *Soit A un arbre. Alors $E^*(A) \subseteq X(A)$.*

Preuve: Soit A un arbre. Puisque $X(X(A)) = X(A)$, tout ce qu'il reste à démontrer est $E(A) \subseteq X(A)$. Rappelons que $A' \in E(A)$ est obtenu en substituant un sous-arbre B de A avec un arbre de la forme $C \to B$. Il est clair, vu la définition des extensions que $C \to B \in X(B)$, et par conséquent $A' \in X(A)$. □

Nous souhaitons démontrer que les ensembles contenant "trop" de répétitions de certaines variables a une profondeur gauche fixée ont une petite fraction limite. Toutefois, comme nous ne démontrerons pas l'existence de la fraction limite, nous utiliserons la limite supérieure : μ_k^+ pour donner des bornes supérieures.

Soit $\mathcal{V}$ un sous-ensemble fixé de variables $\{x_i \mid i > 0\}$, indépendant du nombre k de variables que nous considérons. Soient p et q deux entiers. Soit $\mathcal{A}_q^p(\mathcal{V})$ l'ensemble d'arbres de $\mathcal{F}_k$ contenant au moins p feuilles étiquetées dans $\mathcal{V}$, chacune d'elle étant à une profondeur gauche au plus q. La suite de la sous-section consiste à démontrer une borne supérieure de $\mu_k^+\left(E^*\left(\mathcal{A}_q^p(\mathcal{V})\right)\right)$. Pour cela, introduisons $\mathcal{B}_q^p(\mathcal{V})$, l'ensemble d'arbres $B \in \mathcal{F}_k$ tels que $p \leqslant |B| \leqslant pq+1$ et qui contiennent au moins p feuilles étiquetées par $\mathcal{V}$. Remarquons que $\mathcal{B}_q^p(\mathcal{V})$ dépend implicitement du nombre k de variables de notre système. Les ensembles $\mathcal{A}_q^p(\mathcal{V})$ et $\mathcal{B}_q^p(\mathcal{V})$ sont reliés via le lemme suivant.

Lemme 5 *Pour un ensemble fixe de variables $\mathcal{V}$ et deux entiers p et q, nous avons $\mathcal{A}_q^p(\mathcal{V}) \subseteq X(\mathcal{B}_q^p(\mathcal{V}))$ et par conséquent $E^*(\mathcal{A}_q^p(\mathcal{V})) \subseteq X(\mathcal{B}_q^p(\mathcal{V}))$.*

Preuve: Soit $A \in \mathcal{A}_q^p(\mathcal{V})$. Soient $\nu_1, \ldots, \nu_p$ les p feuilles de A étiquetées avec les variables de $\mathcal{V}$, à profondeur gauche inférieure ou égale à q. Soient $C_1, \ldots, C_r$ les ensembles maximaux (par rapport à l'inclusion) des sous-arbres gauches de A ne contenant aucun noeud ν_i. Soit B un arbre obtenu à partir de A en supprimant tous les C_i, i.e. $B = A \setminus \{C_1, \ldots, C_r\}$. Clairement $A \in X(B)$, et il peut être vérifié que $B \in \mathcal{B}_q^p(\mathcal{V})$: en effet, le plus grand arbre B que l'on puisse obtenir est celui dont tous les noeuds ν_i ont une profondeur gauche q et appartiennent à des prémisses distinctes de A, alors dans ce cas, $|B| = pq+1$. Donc $A \in X(\mathcal{B}_q^p(\mathcal{V}))$. La seconde partie du lemme provient du Lemme 4 et du fait que $X(X(B_q^p(\mathcal{V}))) = X(\mathcal{B}_q^p(\mathcal{V}))$. □

Lemme 6 *Pour un ensemble fixe de variables $\mathcal{V}$ et deux entiers p et q, nous avons*

$$\mu_k^+\left(X\left(\mathcal{B}_q^p(\mathcal{V})\right)\right) = O\left(\frac{1}{k^p}\right).$$

Preuve: Tout d'abord, pour deux fonctions génératrices $f, g \in \mathbb{R}[[z]]$, nous noterons $f \prec g$ si $[z^n]f \leqslant [z^n]g$ pour tout $n \in \mathbb{N}$. Soit $\gamma = |\mathcal{V}|$. Soit $\phi_{\mathcal{B}_q^p(\mathcal{V})}(z,t)$ la fonction génératrice bivariée énumérant $\mathcal{B}_q^p(\mathcal{V})$, z marquant les feuilles et t tous les noeuds. Elle satisfait :

$$\phi_{\mathcal{B}_q^p(\mathcal{V})}(z,t) \prec \sum_{\ell=p}^{pq+1} C_{\ell-1}\binom{\ell}{p}\gamma^p k^{\ell-p} z^\ell t^{2\ell-1}.$$

Soit $\phi_{X(\mathcal{B}_q^p(\mathcal{V}))}(z)$ la fonction génératrice de l'ensemble $X(\mathcal{B}_q^p(\mathcal{V}))$; il est clair que

$$\phi_{X(\mathcal{B}_q^p(\mathcal{V}))}(z) = \phi_{\mathcal{B}_q^p(\mathcal{V})}\left(z, \frac{1}{1-f_k(z)}\right).$$

Nous avons :

$$\phi_{X(\mathcal{B}_q^p(\mathcal{V}))}(z) \prec \sum_{\ell=p}^{pq+1} C_{\ell-1}\binom{\ell}{p}\gamma^p k^{1-\ell-p} z^{1-\ell} f_k(z)^{2\ell-1}.$$

En utilisant l'inversion de Lagrange [FS96, Chapitre 3] nous obtenons

$$\begin{aligned}
[z^n]\phi_{X(\mathcal{B}_q^p(\mathcal{V}))}(z) &\leqslant \sum_{\ell=p}^{pq+1} C_{\ell-1}\binom{\ell}{p}\gamma^p k^{1-\ell-p}[z^{n+\ell-1}]f_k(z)^{2\ell-1} \\
&\leqslant \sum_{\ell=p}^{pq+1} C_{\ell-1}\binom{\ell}{p}\gamma^p k^{n-p}\frac{2\ell-1}{n+\ell-1}\binom{2n-2}{n+\ell-1}.
\end{aligned}$$

Rappelons que

$$\mu_k\left(X\left(\mathcal{B}_q^p(\mathcal{V})\right)\right) = \lim_{n\to\infty} \frac{[z^n]\phi_{X(\mathcal{B}_q^p(\mathcal{V}))}(z)}{[z^n]f_k(z)}.$$

Puisque $[z^n]f_k(z) = k^n C_{n-1}$, nous avons

$$\frac{[z^n]\phi_{X(\mathcal{B}_q^p(\mathcal{V}))}(z)}{[z^n]f_k(z)} \leqslant \sum_{\ell=p}^{pq+1} C_{\ell-1}\binom{\ell}{p}\gamma^p k^{-p}(2\ell-1)\frac{1}{C_{n-1}}\frac{1}{n+\ell-1}\binom{2n-2}{n+\ell-1}.$$

Or remarquons que pour tout $\ell \geqslant 0$,

$$\frac{1}{C_{n-1}}\frac{1}{n+\ell-1}\binom{2n-2}{n+\ell-1} = \frac{n}{n+\ell-1}\frac{\binom{2n-2}{n+\ell-1}}{\binom{2n-2}{n-1}} \leqslant 2.$$

Par conséquent

$$\mu_k^+\left(X\left(\mathcal{B}_q^p(\mathcal{V})\right)\right) \leqslant \frac{2\gamma^p}{k^p}\sum_{\ell=p}^{pq+1} C_{\ell-1}\binom{\ell}{p}(2\ell-1).$$

Ceci implique $\mu_k^+(X(\mathcal{B}_q^p(\mathcal{V}))) = O(1/k^p)$. □

Corollaire 1 *Pour un ensemble fixe de variables $\mathcal{V}$ et deux entiers p et q, nous avons*

$$\mu_k^+\left(E^*\left(\mathcal{A}_q^p(\mathcal{V})\right)\right) = O\left(\frac{1}{k^p}\right).$$

Preuve: La preuve est immédiate à partir des Lemmes 5 et 6. □

Nous avons désormais les outils nécessaires pour borner des ensembles d'expressions et donner aisément l'ordre de leurs fractions limites.

5.2.2 Arbres irréductibles et une seule expansion des arbres minimaux

Toute expression calculant une fonction est obtenue après un certain nombre d'expansions d'un arbre irréductible calculant cette fonction. Cette partie a pour but de partitionner l'ensemble des arbres irréductibles d'une fonction donnée, puis de calculer la fraction limite d'une seule expansion des arbre minimaux. Nous prouverons par la suite que toutes les autres expansions ne contribuent pas de façon significative à la fraction limite de la fonction.

Soit f une fonction Booléenne distincte de *Vrai* et expressible dans le système. La variable x est appelée *variable essentielle* pour f si les deux fonctions obtenues en évaluant x à 0 ($f_{|x=0}$) et à 1 ($f_{|x=1}$) sont deux fonctions distinctes. Sinon x est dite *variable inessentielle* pour f.

Soient A un arbre et ν un de ses noeuds (internes ou externes). Nous définissons $\Delta(\nu)$ comme le plus petit sous-arbre gauche de A contenant ν, où l'arbre global est considéré lui-même comme un sous-arbre gauche. Ainsi, $\Delta(\nu)$ est l'arbre global lorsque ν a un profondeur gauche égale à 0 et est réellement un sous-arbre gauche lorsque la profondeur gauche de ν est strictement positive. De la même manière, nous définissons $\Delta^2(\nu)$ comme le plus petit sous-arbre gauche contenant strictement $\Delta(\nu)$ – à nouveau l'arbre global est considéré comme un sous-arbre gauche dans cette définition (voir Figure 5.11). Nous noterons $\Delta(B)$ pour le sous-arbre B enraciné en ν, au lieu de $\Delta(\nu)$ – et de la même manière nous écrirons $\Delta^2(B)$ pour $\Delta^2(\nu)$).

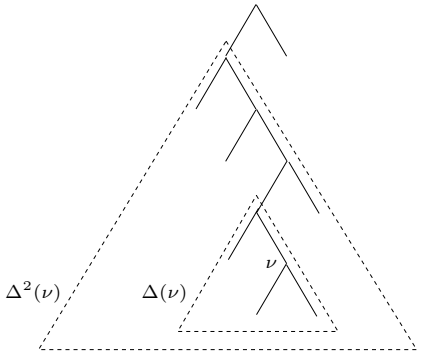

FIG. 5.11 – Les sous-arbres gauches $\Delta(\nu)$ et $\Delta^2(\nu)$ reliés au noeud ν d'un arbre.

Lemme 7 *Quel que soit l'arbre A, calculant la fonction f distincte de* Vrai, *il contient au moins $L(f)$ occurrences de variables essentielles de f.*

Preuve: Soit A un arbre calculant la fonction f distincte de *Vrai*. Remarquons tout d'abord que le but de A est une variable essentielle ; sinon A calculerait la constante *Vrai*. Par conséquent, $\Delta(\nu)$ n'est pas l'arbre tout entier, pour chaque feuille étiquetée avec une variable inessentielle. Soit $\{\Delta_1, \ldots, \Delta_p\}$ l'ensemble des $\Delta(\nu)$ maximaux pour l'inclusion, lorsque ν parcourt toutes les feuilles de A étiquetées par des variables inessentielles. Il est clair que les Δ_i sont disjoints. Si nous évaluons toutes les variables inessentielles à 1, alors tous les Δ_i s'évaluent à 1, car leurs buts sont des variables inessentielles. Ainsi $A' = A \setminus \{\Delta_1, \ldots, \Delta_p\}$ calcule f. Puisque A' ne contient que des variables essentielles et que $|A| \geqslant |A'| \geqslant L(f)$, il s'en suit qu'il y a au moins $L(f)$ variables essentielles dans A. □

La taille d'un arbre A calculant $f \neq Vrai$ peut être écrite $|A| = L(f) + e + i$, où $L(f) + e$ est le nombre de feuilles étiquetées par des variables essentielles et i est le nombre de feuilles étiquetées par des variables inessentielles ; remarquons que $e \geqslant 0$ en raison du Lemme 7. Etant donnée une fonction f distincte de *Vrai*, nous décomposons l'ensemble des arbres irréductibles calculant f en les ensembles disjoints suivants :

- $\mathcal{M}_f$ l'ensemble des arbres minimaux, i.e. arbres de taille $L(f)$ (cas $e = i = 0$) ;
- $\mathcal{P}_{f,1}$ l'ensemble des arbres irréductibles de taille strictement supérieure à $L(f)$, avec exactement $L(f)$ occurrences de variables essentielles et au moins une occurrence de variable inessentielle (cas $e = 0$, $i > 0$) ;
- $\mathcal{P}_{f,2}$ l'ensemble des arbres irréductibles de taille $L(f) + 1$, avec aucune occurrence de variable inessentielle (cas $e = 1$, $i = 0$) ;
- $\mathcal{P}_{f,3}$ l'ensemble des arbres irréductibles de taille strictement supérieure à $L(f) + 1$, avec exactement $L(f) + 1$ occurrences de variables essentielles et $i > 0$ occurrences de variables inessentielles *toutes distinctes* (cas $e = 1$, $i > 0$, première partie) ;
- $\mathcal{P}_{f,4}$ l'ensemble des arbres irréductibles de taille strictement supérieure à $L(f) + 2$, avec exactement $L(f) + 1$ occurrences de variables essentielles et $i > 0$ occurrences de variables inessentielles telles qu'au moins une des variables inessentielles est répétée (cas $e = 1$, $i > 0$, seconde partie) ;
- $\mathcal{P}_{f,5}$ l'ensemble des arbres irréductibles contenant au moins $L(f) + 2$ occurrences de variables essentielles (cas $e \geqslant 2$, $i \geqslant 0$).

De façon immédiate nous remarquons que chaque arbre calculant f appartient à l'ensemble des expansions d'un arbre irréductible calculant f (la preuve se fait en élaguant autant que possible l'arbre de départ). Le Théorème 2 [p. 24] s'obtient en évaluant les fractions limites des ensembles $E^*(\mathcal{C})$ pour chaque classe $\mathcal{C}$ que nous venons de décrire.

Commençons par introduire un ensemble $\mathcal{N}$ d'arbres. Pour chaque fonction Booléenne f, distincte de *Vrai*, soit Γ l'ensemble de ses variables essentielles.

$$\mathcal{N} = \mathcal{A}^{L(f)+2}_{L(f)+2}(\Gamma) \cup \bigcup_{\alpha \in \{x_1,\ldots,x_k\}} \mathcal{A}^{L(f)+3}_{L(f)+2}(\Gamma \cup \{\alpha\}) \cup \bigcup_{\alpha,\beta \in \{x_1,\ldots,x_k\}} \mathcal{A}^{L(f)+4}_{L(f)+2}(\Gamma \cup \{\alpha,\beta\}).$$

Remarquons que $\mathcal{N}$ dépend implicitement de la fonction f considérée. Nous allons utiliser constamment la proposition suivante :

Proposition 4

$$\mu^+_k(E^*(\mathcal{N})) = O\left(\frac{1}{k^{L(f)+2}}\right).$$

Preuve: En utilisant le Corollaire 1, nous avons $mu^+_k(\mathcal{A}^{L(f)+2}_{L(f)+2}) = O(\frac{1}{k^{L(f)+2}})$. Toujours avec le même corollaire, nous obtenons pour chaque variable α :

$$\mu^+_k(E^*(\mathcal{A}^{L(f)+3}_{L(f)+2}(\Gamma \cup \{\alpha\}))) = O\left(\frac{1}{k^{L(f)+3}}\right).$$

En outre, de la même manière, nous avons pour toutes les variables α et β :

$$\mu^+_k(E^*(\mathcal{A}^{L(f)+4}_{L(f)+2}(\Gamma \cup \{\alpha,\beta\}))) = O\left(\frac{1}{k^{L(f)+4}}\right).$$

□

Les lemmes suivants établissent quelques restrictions sur les types d'expansions possibles des ensembles $\mathcal{M}_f$ et $\mathcal{P}_{f,2}$.

Lemme 8 *Soient f une fonction Booléenne distincte de* Vrai*, et $A \in \mathcal{M}_f \cup \mathcal{P}_{f,2}$. Aucune expansion de type "but α" ou "prémisse α" par rapport à une variable inessentielle α n'est valide dans A.*

Preuve: Démontrons le résultat via une preuve par l'absurde. Soit $A \in \mathcal{M}_f \cup \mathcal{P}_{f,2}$. Posons ν un des noeuds en lequel il est possible d'appliquer une expansion valide de type "but α" ou "prémisse α" par rapport à une variable inessentielle α.

Remarquons que la profondeur gauche de ν est d'au moins 1 ; sinon f serait égale à *Vrai*. Par conséquent $\Delta(\nu)$ n'est pas l'arbre tout entier. Nous allons tout d'abord prouver que $|\Delta(\nu)| = 1$.

Supposons que l'expansion valide soit de type "but α". Soit A' l'arbre obtenu après une expansion en ν de A avec le sous-arbre gauche ajouté réduit à α. Alors nous avons $[A'_{|\alpha=0}] = [A \setminus \Delta(\nu)] = f$ et nous en concluons que $|\Delta(\nu)| = 1$ et $|A| = L(f) + 1$ – sinon nous aurions un arbre de taille strictement plus petite que $L(f)$ qui calculerait f. Supposons désormais que l'expansion valide en ν soit de type "prémisse α". Soit x le but de l'arbre A et A' l'arbre obtenu par expansion de A en ν avec le sous-arbre $\alpha \to x$. Alors nous avons $[A'_{|\alpha=1,\ x=0}] = [A \setminus \Delta(\nu)_{|x=0}] = f_{|x=0}$ et bien évidemment $[A \setminus \Delta(\nu)_{|x=1}] = 1 = f_{|x=1}$. A nouveau nous en

concluons que $[A \setminus \Delta(\nu)] = f$; par conséquent $|\Delta(\nu)| = 1$ et $|A| = L(f) + 1$, également dans ce cas.

En conséquence, peu importe le type d'expansion valide, nous savons que $\Delta(\nu)$ est une feuille. Soit y l'étiquette de $\Delta(\nu)$. Dans l'arbre A, le sous-arbre gauche $\Delta(\nu)$ calcule y. De plus, pour les deux types d'expansions, nous avons montré que $A \setminus \Delta(\nu)$ calculait toujours f ; en le disant autrement, la substitution de $\Delta(\nu)$ par 1 ou par son but y dans A ne change pas la fonction calculée. Il s'en suit, par le Lemme 2 que A est réductible, ce qui est absurde. □

Remarquons que le Lemme 8 entraîne le fait que le nombre $\lambda(f)$, défini dans l'introduction du chapitre, ne dépend pas du nombre k de variables que nous considérons. Intéressons à une expansion des arbres minimaux de f.

Lemme 9 *Soit f une fonction Booléenne distincte de* Vrai. *En appliquant une unique expansion des arbres minimaux, nous obtenons :*

$$\mu_k(E(\mathcal{M}_f)) = \frac{\lambda(f)}{4^{L(f)}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Preuve: Soient f une fonction Booléenne distincte de *Vrai* et $\mathcal{M}_f$ son ensemble d'arbres minimaux. Etant donnés $A \in \mathcal{M}_f$ et ν un noeud de A, soit $\lambda_\nu(A)$ le nombre d'expansions valides en ν de A et $E_\nu(A)$ l'ensemble des arbres obtenus après une unique expansion de A en ν.

Remarquons qu'étant donné un arbre B obtenu depuis $A \in \mathcal{M}_f$ en ajoutant un sous-arbre gauche de taille au moins $L(f)$, il est possible de déterminer A ; ceci car B ne possède qu'un seul sous-arbre de taille $|B| - L(f)$. Par conséquent, pour $n \geqslant 2L(f)$, on a :

$$|\{T \in E(\mathcal{M}_f)\ ;\ |T| = n\}| = \sum_{A \in \mathcal{M}_f, \nu \in A} |\{T \in E_\nu(A)\ ;\ |T| = n\}|.$$

Ce qui implique

$$\mu_k(E(\mathcal{M}_f)) = \sum_{A \in \mathcal{M}_f, \nu \in A} \mu_k(E_\nu(A)). \tag{5.3}$$

Pour $A \in \mathcal{M}_f$ et ν un noeud de A, nous souhaitons désormais estimer $\mu_k(E_\nu(A))$. Commençons par considérer la fraction limite des arbres obtenus par une expansion d'un type donné en ν. Pour t un type d'expansion ("tautologie", "but α", ou "prémisse α"), nous notons $E_\nu^t(A)$ l'ensemble de tous les arbres obtenus depuis A par une expansion en ν de type t. La fonction génératrice de $E_\nu^{taut}(A)$ est égale à $\phi_{Cl_k}(z) \cdot z^{L(f)}$ et le calcul de sa fraction limite donne

$$\mu_k(E_\nu^{taut}(A)) = \frac{1}{4^{L(f)} k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

La fonction génératrice de $E_\nu^{but\ \alpha}(A)$ est égale à $\frac{1}{k} f_k(z) \cdot z^{L(f)}$ donc sa fraction limite vérifie

$$\mu_k(E_\nu^{but\ \alpha}(A)) = \frac{1}{4^{L(f)} k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Via la même méthode, la fonction génératrice énumérant of $E_\nu^{prem\ \alpha}(A)$ vaut $\frac{z}{1 - f_k(z) + z} f_k(z) \cdot z^{L(f)}$, ainsi, sa fraction limite est

$$\mu_k(E_\nu^{prem\ \alpha}(A)) = \frac{1}{4^{L(f)} k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Soit $\mathcal{I} = E_\nu^{t_1}(A) \cap E_\nu^{t_2}(A)$ où $t_1 \neq t_2$ sont deux types d'expansions valides en ν. Du fait de la structure des tautologies (le but global doit se répéter dans un but des prémisses, sinon il ne s'agit pas d'une tautologie), on en conclut que $\mathcal{I} \subset \mathcal{N}$. La Proposition 4 [p. 37] nous donne $\mu_k^+(\mathcal{I}) = O(1/k^{L(f)+2})$ et par le principe d'inclusion-exclusion, nous obtenons :

$$\mu_k(E_\nu(A)) = \frac{\lambda_\nu(A)}{4^{L(f)}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right). \tag{5.4}$$

Les deux équations (5.3 [p. 38]) et (5.4 [p. 39]) concluent la preuve

$$\mu_k(E(\mathcal{M}_f)) = \frac{\lambda(f)}{4^{L(f)}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

□

5.2.3 Les expansions non significatives

Dans cette sous-section, nous démontrons que les expansions des arbres irréductibles dont nous n'avons pas encore tenu compte, n'augmentent pas de manière significative l'ensemble d'expressions obtenu par une expansion des arbres minimaux. Tout d'abord, démontrons que les ensembles $\mathcal{P}_{f,1}$ et $\mathcal{P}_{f,3}$ sont vides.

Lemme 10 *Pour toute fonction f distincte de* Vrai*, l'ensemble $\mathcal{P}_{f,1}$ est vide.*

Preuve: Soit f une fonction distincte de *Vrai*. Raisonnons par l'absurde. Supposons que $\mathcal{P}_{f,1}$ ne soit pas vide. Soit $A \in \mathcal{P}_{f,1}$. La taille de A est $L(f)+i$, avec exactement $L(f)$ occurrences de variables essentielles et $i > 0$ occurrences de variables inessentielles. Soit $\{\Delta_1, \ldots, \Delta_p\}$ l'ensemble des $\Delta(\nu)$ maximaux par rapport à l'inclusion, lorsque ν parcourt l'ensemble des noeuds dont l'étiquette est une variable inessentielle. Si nous évaluons toutes les variables inessentielles à 1, alors chaque Δ_i s'évalue à 1, puisque leur but est une variable inessentielle. De plus, du fait qu'il s'agit de sous-arbre gauches, l'arbre $A' := A \setminus \{\Delta_1, \ldots, \Delta_p\}$ calcule f. Le fait que A possède exactement $L(f)$ occurrences de variables essentielles, aucun des Δ_i ne contient de variables essentielles.

Supposons désormais qu'il n'existe aucune affectation des variables inessentielles telles que Δ_1 s'évalue à 0 : Alors Δ_1 est une tautologie et donc A est réductible, ce qui est absurde. Ainsi il existe une affectation a des variables inessentielles telle que Δ_1 s'évalue à 0 pour a. Remarquons que Δ_1 ne peut pas être une prémisse de A puisque $f \neq Vrai$, donc $\Delta_1^2 := \Delta^2(\Delta_1)$ n'est pas l'arbre tout entier et s'évalue à 1 pour a. Soit $S = \{\Delta_i \mid [\Delta_{i|a}] = 1\} \cup \{\Delta_i^2 \mid [\Delta_{i|a}] = 0\}$ – où $[C_{|a}]$ représente la fonction calculée par le sous-arbre C pour l'affectation a. L'ensemble S est composé de sous-arbres gauches, tous s'évaluant à 1 pour a. Par ailleurs S contient Δ_1^2 qui comporte au moins une variable essentielle (son but) – sinon Δ_1 n'aurait pas été maximal. Ainsi $A'' := A \setminus S$ est de taille au plus $L(f) - 1$ et calcule f, ce qui est absurde. □

En utilisant le même genre de raisonnement, nous obtenons le lemme suivant :

Lemme 11 *Pour toute fonction f distincte de* Vrai*, l'ensemble $\mathcal{P}_{f,3}$ est vide.*

Preuve: Raisonnons à nouveau par l'absurde, supposons $\mathcal{P}_{f,3}$ non vide et soit $A \in \mathcal{P}_{f,3}$. On a $|A| = L(f) + 1 + i$, A contenant i occurrences de variables inessentielles, toutes

distinctes. Définissons $\{\Delta_1, \ldots, \Delta_p\}$ comme dans la preuve du Lemme 10. Puisque $[A] = [A \setminus \{\Delta_1, \ldots, \Delta_p\}]$, l'ensemble $\bigcup_i \Delta_i$ contient au plus une occurrence d'une variable essentielle.

Supposons tout d'abord que $\bigcup_i \Delta_i$ ne contienne que des variables inessentielles. Remarquons que dans ce cas, il existe une affectation des variables inessentielles telle que tout Δ_i s'évalue à 0 (cette affectation existe car il n'y a aucune répétition des variables inessentielles). Par conséquent, $A \setminus \{\Delta_1^2, \ldots, \Delta_p^2\}$ est un arbre calculant f. Ainsi tous les Δ_i^2 sont égaux au même sous-arbre gauche de A. Donc les Δ_i sont tous des prémisses (et les uniques) d'un seul sous-arbre gauche de A dont le but est une variable essentielle α. Dans ce cas nous affirmons que A est obtenu à partir de $A \setminus \Delta_1^2$ par une expansion valide de type "but α". En effet, évaluons toutes les variables inessentielles à 1. Alors Δ_1^2 s'évalue à α et A calcule f pour cette affectation. Maintenant, choisissons une affectation complète des variables inessentielles telle que Δ_1 s'évalue à 0 (elle existe). Alors Δ_1^2 s'évalue à 1 et l'arbre A calcule f pour cette affectation. En utilisant le Lemme 2, A est réductible, ce qui est absurde.

Supposons désormais que $\bigcup_i \Delta_i$ contienne exactement une variable essentielle – disons qu'elle appartient à Δ_1. Il existe une affectation des variables inessentielles telle que Δ_1 s'évalue à 1 et tous les autres Δ_i s'évaluent à 0. Donc $A \setminus \{\Delta_1, \Delta_2^2, \ldots, \Delta_p^2\}$ calcule f. Mais sa taille est strictement inférieure à $L(f)$ si $p > 1$. Par conséquent $p = 1$. Par définition de Δ_1, α ne peut pas être son but. Donc α appartient à l'une des prémisses de Δ_1.

Cas 1 : Cette prémisse est de taille 1 (i.e. elle est réduite à α). Nous affirmons dans ce cas que A est obtenu à partir de $A \setminus \Delta_1$ par une expansion valide de type "prémisse α". Ainsi, évaluons toutes les variables inessentielles à 1. Alors Δ_1 s'évalue à 1 et l'arbre obtenu calcule f. Maintenant, choisissons une affectation des variables inessentielles telle que Δ_1 s'évalue à $\bar{\alpha}$ (elle existe). L'arbre obtenu calcule f. D'après le Lemme 3, A est réductible, ce qui est contradictoire.

Cas 2 : Cette prémisse est de taille au moins 2. Dans ce cas, nous pouvons trouver une affectation des variables inessentielles telle que Δ_1 s'évalue à 0. Ainsi, supprimer Δ_1^2 de A donne un arbre de taille strictement inférieure à $L(f)$ et qui calcule f. Ce qui est absurde. □

Il est facile de vérifier que les deux ensembles $\mathcal{P}_{f,4}$ et $\mathcal{P}_{f,5}$ ne sont pas vides quelle que soit la fonction f distincte de *Vrai*. D'un autre côté, suivant la fonction f, l'ensemble $\mathcal{P}_{f,2}$ est vide (ou non) : $\mathcal{P}_{f,2}$ est vide si $f = x_1$ alors que $\mathcal{P}_{g,2}$ n'est pas vide lorsque $g = x_1 \vee (\bar{x}_2 \wedge x_3 \wedge x_4) \vee (\bar{x}_2 \wedge \bar{x}_5)$. En effet, on peut vérifier que $L(g) = 6$ et que la formule $(x_3 \to (x_4 \to x_2)) \to (((x_5 \to x_2) \to x_2) \to x_1)$ appartient à $\mathcal{P}_{g,2}$. Notre prochaine étape consiste à démontrer que les expansions itérées de $\mathcal{P}_{f,2}$ construisent une famille incluse dans $\mathcal{N}$. Pour cela, nous supposons que f est telle que $\mathcal{P}_{f,2} \neq \emptyset$ – sinon la fraction limite de $E^*(\mathcal{P}_{f,2})$ est clairement égale à 0. Nous sommes donc prêts à borner la fraction limite de $E^*(\mathcal{P}_{f,2})$.

Lemme 12 *Pour toute fonction f distincte de* Vrai, $E^*(\mathcal{P}_{f,2}) \subseteq \mathcal{N}$.

Preuve: Notons tout d'abord que $\mathcal{P}_{f,2}$ est un ensemble fini. Intéressons-nous donc aux expansions de $\mathcal{P}_{f,2}$. Soit $A \in E(\mathcal{P}_{f,2})$; nous avons $A \in E(I)$ avec un arbre irréductible $I \in \mathcal{P}_{f,2}$. Nous allons prouver que A satisfait l'une des conditions suivantes :

- A contient au moins $L(f)+2$ occurrences de variables essentielles ayant une profondeur gauche au plus $L(f)+2$;
- A contient $L(f)+1$ occurrences de variables essentielles et deux occurrences de la même variable inessentielle, toutes ayant une profondeur gauche inférieure ou égale à $L(f)+2$.

Si A est obtenu depuis I par une expansion de type "but" ou "prémisse", alors ce doit être par rapport à une variable essentielle d'après le Lemme 8. Remarquons désormais que I est de taille $L(f)+1$, par conséquent tous ses noeuds ont une profondeur gauche au plus $L(f)$ (il y a au moins un noeud par profondeur gauche). Puisque les expansions préservent la profondeur gauche des noeuds présents dans l'arbre initial, nous en concluons que A satisfait la première des conditions ci-dessus. Supposons maintenant que A est obtenu depuis I par une expansion de type "tautologie". D'après la structure des tautologies, nous savons qu'elles contiennent deux occurrences de la même variable x parmi ses noeuds de profondeur gauche au plus 1. Si x est une variable essentielle de f, alors A satisfait la première condition. Sinon, A satisfait la seconde condition.

Les conditions satisfaites par A montrent que $A \in \mathcal{N}$. □

Prenons désormais $\mathcal{P}_{f,4}$ et $\mathcal{P}_{f,5}$ en compte. Dans le même esprit que pour la fraction limite des expansions répétées de $\mathcal{P}_{f,2}$, nous allons démontrer ce qui suit :

Lemme 13 *Pour toute fonction f distincte de* Vrai*, nous avons $E^*(\mathcal{P}_{f,4} \cup \mathcal{P}_{f,5}) \subseteq \mathcal{N}$.*

Preuve: Soit Γ l'ensemble des variables essentielles de f. Soit $A \in \mathcal{P}_{f,4} \cup \mathcal{P}_{f,5}$. Notons $|A| = L(f)+e+i$ la taille de A, où $L(f)+e$ correspond au nombre d'occurrences de variables essentielles et i correspond au nombre d'occurrences de variables inessentielles. Soit $e \geqslant 2$ et $i \geqslant 0$, soit $e = 1$ et $i \geqslant 2$ avec au moins deux occurrences de la même variable inessentielle. Comme précédemment, soit $\{\Delta_1, \ldots, \Delta_p\}$ l'ensemble des $\Delta(\nu)$ maximaux pour l'inclusion lorsque ν décrit l'ensemble des noeuds étiquetés par une variable inessentielle. Nous allons étudier dans l'ordre les cas suivants : $p = 0$, $p = 1$ et $p \geqslant 2$.

Cas $p = 0$. Ce cas a lieu lorsque $i = 0$ et $e \geqslant 2$. Dans ce cas, $A \in \mathcal{A}^{L(f)+2}_{L(f)+1}(\Gamma) \subseteq \mathcal{N}$. Donc la Proposition 4 [p. 37] donne le résultat.

Cas $p = 1$. Soit a une affectation partielle où chaque variable inessentielle est égale à 1. Le sous-arbre Δ_1 est un sous-arbre gauche et calcule 1 pour a. Donc

$$[A \setminus \Delta_1] = [A_{|a}] = [A].$$

Ainsi, l'arbre $A \setminus \Delta_1$, qui ne contient que des variables essentielles a une taille égale à $L(f)$. Nous allons maintenant diviser ce cas selon la taille de $A \setminus \Delta_1$.

- $A \setminus \Delta_1$ a une taille d'au moins $L(f) + 2$. Comme dans le cas $p = 0$, nous avons $A \in \mathcal{A}^{L(f)+2}_{L(f)+1}(\Gamma) \subseteq \mathcal{N}$.
- $A \setminus \Delta_1$ a une taille valant $L(f) + 1$. Notons B_i les prémisses de Δ_1 et β son but. Considérons deux cas :

  1. Il existe i tel que $r(B_i) \in \Gamma \cup \{\beta\}$. Alors nous avons $L(f) + 3$ occurrences de variables de $\Gamma \cup \{\beta\}$ dont les profondeurs gauches sont strictement plus petites que $L(f) + 2$. Ainsi $A \in \mathcal{A}^{L(f)+3}_{L(f)+2}(\Gamma \cup \{\beta\}) \subseteq \mathcal{N}$.
  2. Pour tout i, $r(B_i)$ n'est pas une variable essentielle et est différente de β. Soit a une affectation des variables inessentielles telle que $\beta = 0$ et toutes les autres variables inessentielles sont évaluées à 1. Alors, $[\Delta_{1\ |a}] = 0$ et il s'en suit que $[\Delta^2_{1\ |a}] = 1$ (Δ^2_1 existe car sinon Δ_1 serait une prémisse de l'arbre global et pour a, l'arbre

global calculerait 1). Ainsi $[\Delta_1^2{}_{|a}] = 1$ et $[A_{|a}] = f$. Maintenant le but de Δ_1^2 est une variable essentielle ; notons-la x. Remarquons que Δ_1^2 ne contient aucune autre variable essentielle puisque $[A \setminus \Delta_1^2] = f$. Considérons une affectation b des variables inessentielles satisfaisant $\beta = 1$; dans ce cas $[\Delta_1^2{}_{|b}] = x$ et $[A_{|b}] = f$. En utilisant le Lemme 2, nous concluons que A est réductible en $A \setminus \Delta_1^2$, contradiction.

– $A \setminus \Delta_1$ est de taille $L(f)$. Nous noterons B_i les prémisses de Δ_1 et β son but.

1. Pour tout i, $r(B_i) \notin \Gamma \cup \{\beta\}$. En prenant $\beta = 0$ et toutes les autres variables inessentielles à 1, Δ_1 calcule 0 et donc Δ_1^2 calcule 1. Ainsi $A \setminus \Delta_1^2$ calcule f avec au plus $L(f) - 1$ occurrences de variables essentielles – car le but de Δ_1^2 est une variable essentielle. Ceci est absurde.
2. Il existe $i \neq j$ tels que $r(B_i)$ et $r(B_j)$ sont les deux dans $\Gamma \cup \{\beta\}$. Alors il y a $L(f)+3$ occurrences de variables de $\Gamma \cup \{\beta\}$ dans A avec pour profondeur gauche au plus $L(f)+1$. Ainsi $A \in \mathcal{A}_{L(f)+1}^{L(f)+3}(\Gamma \cup \{\beta\}) \subseteq \mathcal{N}$.
3. Il existe un unique i tel que $r(B_i) \in \Gamma \cup \{\beta\}$. Supposons tout d'abord que B_i soit réduit à une feuille. Si $B_i = \beta$, alors Δ_1 est une tautologie simple ; il s'en suit que A est réductible : absurde. Donc B_i est une variable essentielle x. Soit a une affectation des variables inessentielles telle que $\beta = 0$ et toutes les autres variables inessentielles sont évaluées à 1. Pour a, Δ_1 calcule $\bar{x}$, et A calcule f. Considérons désormais une autre affectation b avec $\beta = 1$; pour b, Δ_1 calcule 1, et A calcule f. D'après le Lemme 3, nous en concluons que A est une expansion valide de $A \setminus \Delta_1$ de type "prémisse x" ; ceci est contradictoire.

   En conséquence, B_i n'est pas réduit à une feuille. Soit $B_i = C_1 \ldots, C_\ell \to \gamma$ avec $\ell \geqslant 1$ et $\gamma \in \Gamma \cup \{\beta\}$. Considérons les sous-cas suivants :
   – S'il existe j tel que $r(C_j) \in \Gamma \cup \{\beta\}$, alors A contient $L(f)+3$ occurrences de variables de $\Gamma \cup \{\beta\}$ avec profondeur gauche au plus $L(f)+2$. Par conséquent $A \in \mathcal{A}_{L(f)+2}^{L(f)+3}(\Gamma \cup \{\beta\}) \subseteq \mathcal{N}$.
   – S'il existe $j \neq k$ tels que $r(C_j) = r(B_k)$ ou $r(C_j) = r(C_k)$. Soit $\delta = r(C_j)$. Alors A contient $L(f)+4$ occurrences de variables de $\Gamma \cup \{\beta, \delta\}$ avec profondeur gauche au plus $L(f)+2$. Ainsi $A \in \mathcal{A}_{L(f)+2}^{L(f)+4}(\Gamma \cup \{\beta, \delta\}) \subseteq \mathcal{N}$.
   – Si tous les $r(C_k)$ sont distincts et pour tout j, $r(C_j) \notin \Gamma \cup \{\beta\} \bigcup_k \{r(B_k)\}$. Si C_1 est réduit à une feuille, considérons l'affectation suivante : $C_1 = 0$, $\beta = 0$ et toutes les autres variables inessentielles sont égales à 1. Alors Δ_1 calcule 0, ainsi Δ_1^2 calcule 1, et donc $A \setminus \Delta_1^2$ calcule f avec au plus $L(f) - 1$ occurrences de variables essentielles, ce qui est absurde. Donc C_1 n'est pas réduit à une feuille ; soit $C_1 = D_1, \ldots, D_m \to \delta$, avec $m \geqslant 1$ et δ une variable essentielle. S'il existe j telle que $r(D_j) \in \Gamma \cup \{\beta, \delta\}$, alors $A \in \mathcal{A}_{L(f)+3}^{L(f)+4}(\Gamma \cup \{\beta\}) \subseteq \mathcal{N}$.
     Sinon, il existe $j \neq k$ tels que $r(D_j) = r(B_k)$ ou $r(D_j) = r(C_k)$, ou $r(D_j) = r(D_k)$: Soit ϵ la variable inessentielle répétée. Alors A contient $L(f)+4$ occurrences de variables de $\Gamma \cup \{\beta, \epsilon\}$ avec profondeur gauche au plus $L(f)+3$. En conséquence, $A \in \mathcal{A}_{L(f)+3}^{L(f)+4}(\Gamma \cup \{\beta, \epsilon\}) \subseteq \mathcal{N}$.
     Si nous ne sommes pas dans les cas précédents, chaque D_j calcule 0 en évaluant son but à 0 et ses sous-buts à 1 (il n'y a pas de conflit entre les buts et les sous-buts puisqu'ils sont tous distincts). Donc pour cette affectation C_1 calcule 0, B_i s'évalue à 1 et finalement tout le sous-arbre Δ_1 s'évalue à 0. Par conséquent,

$A \setminus \Delta_1^2$ calcule f avec au plus $L(f)-1$ occurrences de variables inessentielles, ce qui est absurde.

Cas $p \geqslant 2$. Evidemment $[A \setminus \{\Delta_1, \dots \Delta_p\}] = f$ (évaluer toutes les variables inessentielles à 1). Ainsi $A \setminus \{\Delta_1, \dots \Delta_p\}$, qui ne contient que des variables essentielles a une taille d'au moins $L(f)$. Considérons différents cas, suivant la taille de $A \setminus \{\Delta_1, \dots \Delta_p\}$.

1. $A \setminus \{\Delta_1, \dots \Delta_p\}$ a pour taille $L(f)+2$. Alors $A \in \mathcal{A}_{L(f)+1}^{L(f)+2}(\Gamma) \subseteq \mathcal{N}$.
2. $A \setminus \{\Delta_1, \dots \Delta_p\}$ a pour taille $L(f)+1$. Soit S l'ensemble des buts et sous-buts des sous-arbres $\{\Delta_1, \dots, \Delta_p\}$. S'il existe un noeud de S étiqueté avec une variable essentielle, alors $A \in \mathcal{A}_{L(f)+2}^{L(f)+2}(\Gamma) \subseteq \mathcal{N}$. Sinon, il existe deux noeuds de S étiquetés avec la même variable inessentielle α, alors $A \in \mathcal{A}_{L(f)+2}^{L(f)+3}(\Gamma \cup \{\alpha\}) \subseteq \mathcal{N}$. Finalement si nous ne sommes pas dans les cas précédents, nous pouvons trouver une affectation a des variables inessentielles telle que tous les Δ_i calcule 0 pour a. Alors l'arbre $A \setminus \{\Delta_1^2, \dots, \Delta_p^2\}$ calcule f et contient $L(f)$ variables (tous les Δ_i^2 sont égaux au même sous-arbre gauche). Le but de Δ_1^2 est une variable essentielle notée x. Dans ce cas, nous prouvons que A est réductible. En fait nous affirmons que : A est obtenu depuis $A \setminus \Delta_1^2$ par une expansion valide de type "but x". En effet, évaluer toutes les variables inessentielles à 1. Donc Δ_1^2 s'évalue à x et A calcule f. Maintenant choisir une affectation (complète) des variables inessentielles telle que tout Δ_i s'évalue à 0 (elle existe puisque qu'il n'y a aucune répétition dans les étiquettes de S). Alors Δ_1^2 s'évalue à 1 et l'arbre global calcule f. D'après le Lemme 2, A est réductible, ce cas n'est pas possible.
3. $A \setminus \{\Delta_1, \dots \Delta_p\}$ a pour taille $L(f)$. A nouveau, soit S l'ensemble des buts et sous-buts des sous-arbres $\{\Delta_1, \dots, \Delta_p\}$. Si S contient au moins deux occurrences de variables essentielles, alors A appartient à $\mathcal{A}_{L(f)+1}^{L(f)+2}(\Gamma) \subseteq \mathcal{N}$. Sinon, s'il existe une occurrence de variable essentielle, et deux occurrences de la même variable inessentielle α parmi les étiquettes de S, alors$A$ appartient à $\mathcal{A}_{L(f)+1}^{L(f)+3}(\Gamma \cup \{\alpha\}) \subseteq \mathcal{N}$. Sinon, s'il existe trois occurrences de la même variable inessentielle α ou quatre occurrences de deux variables inessentielles distinctes α et β dans S, alors A appartient à $\mathcal{A}_{L(f)+1}^{L(f)+3}(\Gamma \cup \{\alpha\}) \subseteq \mathcal{N}$ ou $\mathcal{A}_{L(f)+1}^{L(f)+4}(\Gamma \cup \{\alpha, \beta\}) \subseteq \mathcal{N}$. Il y a trois autres cas : *Premier cas : S* contient une unique variable essentielle et aucune répétition parmi les variables inessentielles. Soit Δ_1 le sous-arbre contenant un sous-but étiqueté avec une variable essentielle. Il existe une affectation des variables inessentielles telle que Δ_1 calcule 1 et tous les autres Δ_i calcule 0. Donc nous en concluons que l'arbre $A \setminus \{\Delta_1, \Delta_2^2, \dots, \Delta_p^2\}$ calcule f avec au plus $L(f)-1$ occurrences de variables essentielles (puisque $p \geqslant 2$). Ceci est absurde. *Second cas :* Il n'existe aucune variable essentielle parmi S mais il y a exactement deux occurrences de la même variable inessentielle. Alors il existe une affectation des variables inessentielles telle que tous les Δ_i calculent 0 ou un unique Δ_j calcule 1 et tous les autres calculent 0. Donc nous en concluons que l'arbre $A \setminus \{\Delta_1^2, \dots, \Delta_p^2\}$ (ou $A \setminus \{\Delta_1^2, \dots, \Delta_{j-1}^2, \Delta_j, \Delta_{j+1}^2, \dots, \Delta_p^2\}$ dans l'autre cas) calcule f, avec au plus $L(f)-1$ occurrences de variables essentielles. Ceci est absurde. *Troisième cas :* Il n'existe aucune variable essentielle et aucune répétition parmi S. Alors, il existe une affectation des variables inessentielles telle que tous les Δ_i calculent 0. Nous en concluons que l'arbre $A \setminus \{\Delta_1^2, \dots, \Delta_p^2\}$ calcule f avec au plus $L(f)-1$ occurrences de variables essentielles. Ceci est absurde.

□

La dernière étape permettant de prouver le Théorème 2 consiste à étudier la fraction limite des expansions répétées des arbres minimaux calculant une fonction donnée.

Lemme 14 *Pour toute fonction f distincte de* Vrai, *nous avons* $E^*(\mathcal{M}_f) \setminus E(\mathcal{M}_f) \subseteq \mathcal{N}$.

Preuve: Tout d'abord, détaillons l'ensemble d'arbres suivant :

$$E^*(\mathcal{M}_f) \setminus E(\mathcal{M}_f) = \mathcal{M}_f \cup E^*\left(E^2(\mathcal{M}_f)\right) \setminus E(\mathcal{M}_f).$$

Commençons par démontrer cette inclusion :

$$E^*\left(E^2(\mathcal{M}_f)\right) \setminus E(\mathcal{M}_f) \subseteq E^*\left(E^2(\mathcal{M}_f) \setminus E(\mathcal{M}_f)\right). \tag{5.5}$$

Soient $A \in E^*(E^2(\mathcal{M}_f)) \setminus E(\mathcal{M}_f)$ et $\ell = \inf\{m \mid A \in E^m(E^2(\mathcal{M}_f))\}$. Alors A satisfait : $A \in E^\ell(E^2(\mathcal{M}_f))$ et $A \notin E^{\ell-1}(E^2(\mathcal{M}_f))$. Nous en concluons que : $A \notin E^\ell(E(\mathcal{M}_f))$. Finalement $A \in E^\ell\left(E^2(\mathcal{M}_f) \setminus E(\mathcal{M}_f)\right)$. Nous allons démontrer que $E^*\left(E^2(\mathcal{M}_f) \setminus E(\mathcal{M}_f)\right) \subset \mathcal{N}$. L'équation (5.5 [p. 44]) donne

$$\mu_k^+\left(E^*(\mathcal{M}_f \setminus E(\mathcal{M}_f)\right) \leqslant \mu_k^+(\mathcal{M}_f) + \mu_k^+(\mathcal{N}).$$

Puisque $\mathcal{M}_f$ est fini, $E^*(\mathcal{M}_f \setminus E(\mathcal{M}_f)) \subseteq \mathcal{N}$.

Nous allons maintenant prouver que $E^*\left(E^2(\mathcal{M}_f) \setminus E(\mathcal{M}_f)\right) \subseteq \mathcal{N}$. Soit $A_2 \in E^2(\mathcal{M}_f) \setminus E(\mathcal{M}_f)$. Il existe $A_0 \in \mathcal{M}_f$ et $A_1 \in E(\mathcal{M}_f)$ tel que $A_1 \in E(A_0)$ et $A_2 \in E(A_1)$. Soient B_1 le sous-arbre gauche qui a été ajouté à A_0 afin d'obtenir A_1, et B_2 le sous-arbre gauche qui a été ajouté à A_1 afin d'obtenir A_2.

D'après le Lemme 8 nous savons que A_1 n'est pas obtenu après une expansion valide de type "but α" ou "prémisse α" par rapport à une variable inessentielle α ; donc $A_1 \in \mathcal{A}_{L(f)+1}^{L(f)+1}(\Gamma) \cup \bigcup_\alpha \mathcal{A}_{L(f)+1}^{L(f)+2}(\Gamma \cup \{\alpha\})$ – A_1 appartient au second ensemble s'il est obtenu via une expansion valide de type "tautologie" dont le but est une variable inessentielle. Observons que $A_1 \in \mathcal{N}$ si B_1 est une tautologie qui n'est pas simple. En effet, nous démontrerons (Proposition 7 [p. 51]) qu'une tautologie non simple a trois occurrences de la même variable ou deux fois deux occurrences de deux variables distinctes parmi les feuilles de profondeurs gauches au plus 3. Donc si B_1 est une tautologie non simple, alors $A_1 \in \mathcal{N}$ et par conséquent A_2 appartient au même ensemble.

Supposons désormais que B_1 soit une tautologie simple. Soit S_1 l'ensemble des feuilles de B_1, dont la profondeur gauche – par rapport à B_1 est inférieure ou égale à 2. S'il existe deux noeuds de S_1 dont les étiquettes sont des variables essentielles ou trois noeuds dont les étiquettes sont une variable essentielle et deux occurrences de la même variable inessentielle, ou trois noeuds dont les étiquettes sont la même variable inessentielle, ou quatre noeuds contenant deux fois deux occurrences de deux variables inessentielles distinctes alors $A_1 \in \mathcal{N}$ et par conséquent A_2 appartient au même ensemble. Supposons que nous ne sommes pas dans le cas précédent. Nous allons étudier deux cas suivant l'emplacement de la seconde expansion par rapport à la première. Rappelons que dans les deux cas nous avons $A_1 \in \mathcal{A}_{L(f)+1}^{L(f)+1}(\Gamma) \cup \bigcup_\alpha \mathcal{A}_{L(f)+1}^{L(f)+2}(\Gamma \cup \{\alpha\})$.

Premier cas : Le noeud dans lequel a lieu la seconde expansion B_2 n'appartient pas à B_1. Si B_2 est une tautologie, alors nous en concluons que $A_2 \in \mathcal{N}$.

Si B_2 est ajouté via une expansion valide de type "but α" ou "prémisse α" par rapport à une variable essentielle α, alors $A_2 \in \mathcal{N}$. Supposons désormais que B_2 est ajouté par une expansion valide de type "but α" ou "prémisse α" ou α est inessentielle.

Soit S_2 l'ensemble des feuilles de B_2 dont la profondeur gauche (par rapport à B_2) est inférieure ou égale à 1. Si S_2 contient une variable essentielle, alors $A_2 \in \mathcal{N}$ et nous avons fini la preuve. Pour la même raison, s'il y a une répétition de variable inessentielle dans $S_1 \cup S_2$, alors $A_2 \in \mathcal{N}$. Pour la suite nous supposons que ce n'est pas le cas : il existe une affectation partielle des variables inessentielles telle que B_2 calcule 0 et aucune variable de S_1 ne soit évaluée. Alors, en évaluant les variables inessentielles de B_1, il n'est pas possible de trouver une affectation telle que B_1 calcule 1 – sinon $A \setminus \{B_1, \Delta^2(B_2)\}$ calculerait f avec une taille d'au plus $L(f)-1$. Puisqu'il n'est pas possible d'évaluer B_1 à 1, son but doit être une variable essentielle. Mais aucune autre variable essentielle n'étiquette les noeuds de S_1 et il n'y a aucune répétition parmi les étiquettes de S_1. Si B_1 n'était pas réduit à une feuille, on pourrait l'évaluer à 1. Puisque ce n'est pas le cas, B_1 est réduit à une feuille, nécessairement étiqueté par une variable essentielle. Nous concluons que l'arbre $A' := A_2 \setminus \Delta^2(B_2)$ calcule f et a pour taille $L(f)$, donc c'est un arbre minimal. Il existe une affectation des variables inessentielles telle que B_2 calcule 0 et une seconde telle qu'il calcule 1 – car $r(B_2)$ est inessentielle. Donc $\Delta^2(B_2)$ calcule respectivement 1 et $r(\Delta^2(B_2))$. D'après le Lemme 2, nous concluons que A_2 est obtenu après une expansion valide d'un arbre minimal A', ce qui est absurde puisque nous supposions que $A_2 \notin E(\mathcal{M}_f)$.

Second cas : Le noeud dans lequel a lieu la seconde expansion B_2 appartient à B_1. Si la première expansion de A_0 était de type "but x", alors A_2 peut être obtenu par une unique expansion valide depuis A_0 : ceci est absurde car $A_2 \notin E(\mathcal{M}_f)$. Si la première expansion était de type "prémisse x", alors le seul cas afin que $A_2 \notin E(\mathcal{M}_f)$, est obtenu si B_2 est une expansion de la prémisse réduite à x dans B_1. Par ailleurs, nous pouvons supposer qu'aucun élément de S_2 n'est étiqueté comme ceux de S_1 – sinon $A_2 \in \mathcal{N}$ et ce serait fini. Donc il existe une affectation des variables inessentielles telle que B_1 (qui contient B_2) calcule 0. Alors $A_2 \setminus \Delta^2(B_1)$ calcule f avec une taille d'au plus $L(f)-1$; ceci est absurde. Finalement, il reste le cas où la première expansion est une tautologie. S'il s'agit d'une tautologie simple (l'autre cas a déjà été étudié), alors le seul cas où $A_2 \notin E(\mathcal{M}_f)$ est tel que B_2 est une expansion de la prémisse, égale au but de la tautologie. Pour la même raison que le cas précédent, ceci n'est pas possible. □

5.3 Synthèse des expansions des arbres irréductibles

5.3.1 Démonstration

Preuve du Théorème 2 [p. 24] :

La première partie relative à la fonction *Vrai*, découle des Propositions 1 [p. 27], 2 [p. 28] et 3 [p. 30] puisque

$$G_k \subset Cl_k \subset \mathcal{F}_k \setminus (SN_k \cup LN_k),$$

et donc :

$$\mu_k(Cl_k) = \frac{1}{k} + O\left(\frac{1}{k^2}\right).$$

Dans la seconde partie, il s'agit de calculer la fraction limite des fonctions f fixées, autres que *Vrai*. Il est clair que les arbres calculant f tombent dans répétition d'expansions valides d'un arbre irréductible de f. Donc l'ensemble des arbres calculant f est exactement $\mathcal{F}_k(f) = E^*(\mathcal{M}_f \cup \mathcal{P}_1(f) \cup \mathcal{P}_{f,2} \cup \mathcal{P}_{f,3} \cup \mathcal{P}_{f,4} \cup \mathcal{P}_{f,5})$. Désormais, il est clair que

$$E(\mathcal{M}_f) \subseteq \mathcal{F}_k(f) \subseteq E(\mathcal{M}_f) \cup (E^*(\mathcal{M}_f) \setminus E(\mathcal{M}_f)) \cup \bigcup_{i\in\{1,\ldots,5\}} E^*(\mathcal{P}_i(f)).$$

Le résultat provient des Lemmes 10 [p. 39], 11 [p. 39], 12 [p. 40], 13 [p. 41], 9 [p. 38] et 14 [p. 44], puis on conclut en utilisant la Proposition 4 [p. 37]. □

5.3.2 Bornes sur le nombre d'expansions valides

Proposition 5 *Soient f une fonction Booléenne distincte de* Vrai*, et ℓ le nombre de ses variables essentielles ; alors :*

$$2 \cdot (2L(f) - 1) \cdot |\mathcal{M}_f| \leqslant \lambda(f) \leqslant (1 + 2\ell) \cdot (2L(f) - 1) \cdot |\mathcal{M}_f|.$$

Preuve: Soit M un arbre minimal calculant f et soit ν un noeud de M. Puisque M est de taille $L(f)$, l'arbre M possède $2L(f) - 1$ noeuds au total. Donc nous devons montrer que le nombre λ_ν de types d'expansions valides en ν satisfait $2 \leqslant \lambda_\nu \leqslant 1 + 2n$.

La borne supérieure est obtenue simplement en comptant toutes les expansions de tous les types en ν : 1 pour l'expansion de type "tautologie", ℓ pour les expansions de type "but" et autant pour les expansions de type "prémisse" – rappelons que d'après le Lemme 8 il n'y aucune expansion valide de type "but α" ou "prémisse α" par rapport à une variable inessentielle α.

La borne inférieure est obtenue en remarquant qu'en plus de l'expansion de type "tautologie", toujours valide en ν, l'expansion de type "prémisse x", où x est le but de M, est aussi toujours valide en ν. En effet, soit A l'arbre obtenu à partir de M en remplaçant le sous-arbre B enraciné en ν par $C \to B$ où C est un arbre contenant une prémisse égale à x. Il est clair que $[A_{|x=0}] = [M_{|_x=0}]$ car $[C_{|x=0}] = 1$. De plus, $[A|_{x=1}] = [M|_{x=1}] = 1$ car x est le but de M. Ainsi $[A] = [M]$ et nous en concluons que l'expansion de type "prémisse x" est valide en ν. □

Nous remarquons que pour les fonctions qui sont égales à l'un des littéraux (i.e. il existe i tel que $f(x_1, \ldots, x_k) = x_i$) les bornes inférieure et supérieure sur $\lambda(f)$ coïncident.

5.4 Fonctions *read-once*

Nous considérons ici le cas des fonctions *read-once*, i.e. les fonctions f avec $L(f)$ variables essentielles. Voilà une définition alternative : il s'agit des fonctions dont les arbres minimaux ne comportent aucune répétition de variables. Un arbre dans lequel toutes les feuilles seront étiquetées avec des variables distinctes sera appelé *arbre read-once*. Les arbres *read-once* sont exactement les arbres minimaux des fonctions *read-once*.

Soit $\mathcal{R}_{c,k}$ l'ensemble des arbres *read-once* de taille c. Nos buts sont de déterminer la probabilité de l'ensemble des fonctions *read-once* de complexité c sur $\{x_1, \ldots, x_k\}$ et la valeur moyenne de la probabilité d'une fonction *read-once* de complexité c. Un argument de symétrie nous permet pour cette-dernière, de ne considérer que l'ensemble $\mathcal{R}(x_1, \ldots, x_c)$ des arbres *read-once* sur $\{x_1, \ldots, x_c\}$. Remarquons que cette valeur moyenne ne dépend pas de la valeur du nombre k de variables que nous considérons – à partir du moment où $k \geqslant c$.

5.4.1 Expansions valides dans les arbres *read-once*

Commençons par définir la notion d'équivalence selon une permutation récursive des prémisses, notée $\equiv$. Deux arbres A et B de profondeur gauche 0 sont équivalents si et seulement si ils sont égaux. Deux arbres $A = (A_1, \ldots, A_p, \alpha)$ et $B = (B_1, \ldots, B_q, \beta)$ de profondeur gauche strictement positive sont équivalents si et seulement si $p = q$, $\alpha = \beta$ et il existe une permutation π de $\{1, \ldots, p\}$ telle que $A_i \equiv B_{\pi(i)}$ pour tout $i \in \{1, \ldots, p\}$.

Lemme 15 *Soit A un arbre* read-once *calculant f. L'ensemble des arbres* read-once *calculant f est exactement $\{B \mid B \equiv A\}$; i.e. il s'agit de l'ensemble des arbres obtenus à partir de A via une permutation récursive de ses prémisses.*

Preuve: Le résultat est obtenu par une récurrence par rapport à la taille de l'arbre *read-once*. Soient A et B deux arbres minimaux de f. L'étape initiale est immédiate car deux arbres de taille 1 qui calculent la même fonction f sont égaux, donc équivalents. Supposons que tous les arbres *read-once* de taille au plus r, calculant la même fonction f soient équivalents. Soient $A = (A_1, \ldots, A_p, \alpha)$ et $B = (B_1, \ldots, B_q, \beta)$ deux arbres *read-once* de taille $r + 1$ et calculant la même fonction f. Commençons par étudier les buts de A et B. Soit a une affectation telle que $\alpha = 0$ et toutes les autres variables sont évaluées à 1. Pour cette affectation, A s'évalue à 0. Puisque les deux arbres calculent f, B s'évalue aussi à 0 pour a. Donc le but de B doit être évalué à 0. Puisque α est la seule variable évaluée à 0, on a donc $\beta = \alpha$.

Concentrons-nous désormais sur les prémisses de A et de B. Pour tout i, notons α_i le but de A_i et β_i celui de B_i. Supposons que $\{\alpha_1, \ldots, \alpha_p\} \neq \{\beta_1, \ldots, \beta_q\}$. Par symétrie, nous pouvons supposer qu'il existe $\beta_\ell \in \{\beta_1, \ldots, \beta_q\} \backslash \{\alpha_1, \ldots, \alpha_p\}$. Soit a l'affectation partielle telle que $\alpha = \beta_\ell = 0$ et que toutes les autres variables soient évaluées à 1. Pour cette affectation, A s'évalue à 0 alors que B s'évalue à 1. Ce qui est contradictoire. Par conséquent $\{\alpha_1, \ldots, \alpha_p\} = \{\beta_1, \ldots, \beta_q\}$, et donc $p = q$ puisque A et B sont *read-once*.

Soit π la permutation de $\{1, \ldots, p\}$ telle que $\alpha_i = \beta_{\pi(i)}$ pour tout i. Soient $j \in \{1, \ldots, p\}$ et a une affectation telle que $\alpha = \alpha_j = 0$ et que toutes les autres variables soient évaluées à 1. Pour cette affectation, A calcule $\bar{A}_j$ et B calcule $\bar{B}_{\pi(j)}$. Ces deux sous-arbres doivent calculer la même fonction, sont *read-once* et ont une taille strictement inférieure à $r + 1$. Via l'hypothèse de récurrence, nous en concluons que $A_j \equiv B_{\pi(j)}$. Nous avons prouvé que $A \equiv B$. □

Le Lemme 15 permet de compter le nombre de fonctions *read-once* de complexité c et dépendant des variables $\{x_1, \ldots, x_c\}$. En effet, les classes d'équivalence des arbres *read-once* sur $\{x_1, \ldots, x_c\}$ par rapport à la permutation récursive des prémisses sont en bijection avec les arbres de Cayley étiquetés comportant c feuilles (dont la définition est donnée dans le livre de Flajolet et Sedgewick [FS09]). Enonçons la bijection : si A est une feuille, alors $\tilde{A} = A$. Sinon, si $A = A_1, \ldots, A_p \to r(A)$, alors $\tilde{A}$ est tel que sa racine est étiquetée par $r(A)$ et ses fils sont les sous-arbres $\tilde{A}_1$ jusqu'à $\tilde{A}_p$, où $\tilde{A}_i$ est l'arbre de Cayley obtenu à partir de A_i, pour tout $i \in \{1, \ldots, p\}$. Par conséquent, il y a c^{c-1} fonctions *read-once* dont les variables essentielles soient $\{x_1, \ldots, x_c\}$. La Figure 5.12 présente un exemple.

Pour une expression A, nous définissons $\lambda^{taut}(A)$ le nombre d'expansions valides de type "tautologie". De la même manière, nous notons les nombres d'expansions valides de type "but" et "prémisse" par $\lambda^{but}(A)$ et $\lambda^{prem}(A)$. Il est clair que $\lambda(A) = \lambda^{taut}(A) + \lambda^{but}(A) + \lambda^{prem}(A)$. Puis nous définissons $\lambda^t(\mathcal{R}(x_1 \ldots, x_c)) = \sum_{A \in \mathcal{R}(x_1, \ldots, x_c)} \lambda^t(A)$, le nombre d'expansions de type t de tous les arbres *read-once* dépendant de $\{x_1, \ldots, x_c\}$. Nous allons nous rendre compte

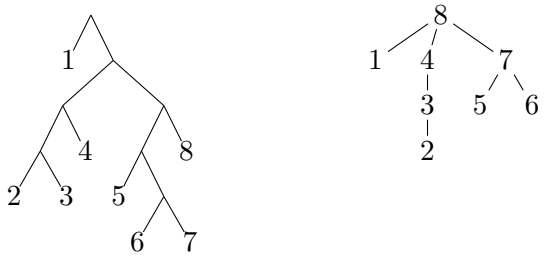

FIG. 5.12 – Un arbre *read-once* et l'arbre de Cayley lui correspondant sur la droite.

que pour un arbre A, les trois quantités $\lambda^t(A)$ ne dépendent pas de l'étiquetage de A mais seulement de sa forme – voir les Lemmes 17 et 16 ci-dessous.

Comme dans tout arbre, les expansions de type "tautologie" sont possibles dans tous les noeuds (internes et externes) dans les arbres *read-once*. Par conséquent $\lambda^{taut}(\mathcal{R}(x_1, \ldots, x_c)) = c!(2c-1)C_{c-1}$. Les deux lemmes suivants donnent une caractérisation des expansions valides de type "but α" ou "prémisse α" dans les arbres *read-once*.

Lemme 16 *Soit A un arbre* read-once. *Une expansion de type "prémisse α" est valide dans le noeud ν si et seulement si il existe une feuille τ dans A, étiquetée par α et telle que $\nu \in \Delta(\tau)$.*

Preuve: Supposons qu'il existe une feuille τ dans A étiquetée par α, et soit ν un noeud de $\Delta(\tau)$. Afin de prouver qu'une expansion de type "prémisse α" est valide dans ν, considérons le sous-arbre $\Delta(\tau)$. Si $\alpha = 1$, peu importe l'expansion avec une prémisse α faite dans le sous-arbre $\Delta(\tau)$, le nouveau sous-arbre $\Delta(\tau)$ calcule toujours 1. Si $\alpha = 0$, il est clair que le sous-arbre gauche ajouté par l'expansion contient une prémisse α et s'évalue donc à 1. Puisqu'il s'agit d'un sous-arbre gauche de $\Delta(\tau)$, cet arbre $\Delta(\tau)$ calcule toujours la même fonction.

Supposons qu'il existe un noeud ν dans A où une expansion de type "prémisse α" est valide. D'après le Lemme 8, nous en concluons qu'il existe un noeud τ dans A, étiqueté par α. Supposons que $\nu \notin \Delta(\tau)$. Prenons l'affectation partielle $\alpha = 1$. La fonction calculée par l'arbre global dépend de $c - |\Delta(\tau)|$ variables – si $|A| = c$. Dès lors, réalisons une expansion dans le noeud ν avec le sous-arbre $S = \alpha \rightarrow \beta$ telle que β soit inessentielle pour $[A]$. En prenant $\alpha = 1$ et $\beta = 0$, la fonction calculée par le nouvel arbre dépend de $c - |\Delta(\tau)| - |\Delta^2(S)|$. Donc l'expansion n'est pas valide. □

Lemme 17 *Soit A un arbre* read-once. *Une expansion de type "but α" est valide dans le noeud ν si et seulement si il existe une feuille gauche τ dans A, étiquetée par α et telle que $\nu \in \Delta^2(\tau) \setminus \{\tau\}$.*

Preuve: Supposons qu'il existe une feuille gauche τ dans A, étiquetée par α et soit ν un noeud de $\Delta^2(\tau) \setminus \{\tau\}$. Afin de prouver qu'une expansion de type "but α" est valide dans ν, considérons la fonction calculée par $\Delta^2(\tau)$. Si $\alpha = 0$, alors $\Delta^2(\tau)$ s'évalue à 1, même si une expansion a été faite dans $\Delta^2(\tau) \setminus \{\tau\}$. Si $\alpha = 1$, le sous-arbre gauche, ajouté par l'expansion avec un but α s'évalue à 1 et donc sa valeur ne change pas la valeur du nouvel arbre $\Delta^2(\tau)_{|\alpha=1}$.

Supposons maintenant qu'il existe un noeud ν dans A où une expansion de type "but α" soit valide. D'après le Lemme 8, nous en concluons qu'il existe un noeud τ dans A, étiqueté par α. Soit A' l'arbre obtenu en remplaçant ν par $\alpha \rightarrow \nu$ dans A. Soit $\beta = r(\Delta(\nu))$. Remarquons tout d'abord que $\beta \neq \alpha$: en effet, si $\beta = \alpha$, la fonction $[A'] = [A]$ ne dépendrait pas de α.

Supposons que τ soit une feuille droite. Dans ce cas, $[A_{|\alpha=0}]$ dépend toujours de β, alors que $[A'_{|\alpha=0}]$ n'en dépend plus : ceci est absurde. Par conséquent nous en concluons que τ est une feuille gauche. Nous avons déjà remarqué que $r(\Delta(\nu)) \neq \alpha$: ceci assure que $\tau \neq \nu$. Il reste à prouver que $\nu \in \Delta^2(\tau)$. Supposons le contraire : $\nu \notin \Delta^2(\tau)$. A nouveau $[A'_{|\alpha=0}]$ ne dépend pas de β alors que $[A_{|\alpha=0}]$ en dépend. Ceci est contradictoire. □

Les Lemmes suivants étudient $\lambda^{but}(\mathcal{R}(x_1,\ldots,x_c))$ et $\lambda^{prem}(\mathcal{R}(x_1,\ldots,x_c))$.

Lemme 18 *Soit $g_n = \sum_{|T|=n} \lambda^{but}(T)$ où la somme parcourt tous les arbres de Catalan (non étiquetés) de taille n. Cette suite est définie par la récurrence suivante :*

$$\begin{cases} g_1 = 0 \\ g_n = C_{n-1}(2n-2) + \sum_{t=2}^{n-1} C_{n-t} g_t \ \text{ pour } n > 1 \end{cases}$$

Preuve: Nous aurons besoin du nombre moyen de prémisses de taille t dans un arbre de Catalan de taille n. Cette valeur fait partie du folklore et est égale à $\bar{p}_t(n) = C_{t-1}C_{n-t}/C_{n-1}$.

Il est immédiat que $g_1 = 0$. Soient $n > 1$. Soit g_n^1 la contribution des prémisses de taille 1 (dans l'arbre global) et $g_n^{>1}$ la contribution des autres prémisses. Nous avons $g_n^1 = (2n-2)\bar{p}_1(n)C_{n-1} = (2n-2)C_{n-1}$, puisque pour chaque prémisse de taille 1, étiquetée par α, les expansions de type "but α" sont valides dans chaque noeud de l'arbre excepté dans la prémisse en question.

Etudions désormais $g_n^{>1}$, en regroupant les prémisses selon leur taille – nous ne considérons que les prémisses de taille au moins 2 puisque $\lambda^{but}(A) = 0$ lorsque $|A] = 1$. Pour un arbre T, nous notons $p(T)$ le nombre de prémisses de T et $T_1,\ldots,T_{p(T)}$ ses prémisses.

$$g_n^{>1} = \sum_{|T|=n} \sum_{i=1}^{p(T)} \lambda^{but}(T_i) = \sum_{|T|=n} \sum_{t=2}^{n-1} \sum_{i=1}^{p(T)} 1_{|T_i|=t} \lambda^{but}(T_i) = \sum_{t=2}^{n-1} \sum_{|T|=n} \sum_{i=1}^{p(T)} 1_{|T_i|=t} \lambda^{but}(T_i).$$

Via un argument de symétrie nous obtenons :

$$\sum_{|T|=n} \sum_{i=1}^{p(T)} 1_{|T_i=t|} \lambda^{but}(T_i) = \left(\sum_{|T|=n} p^t(T) \right) \bar{g}_t,$$

où $\bar{g}_t = g_t/C_{t-1}$. Ainsi nous avons $g_n^{>1} = \sum_{t=2}^{n-1} \bar{p}_t(n) C_{n-1} \bar{g}_t$. En utilisant le nombre moyen de prémisses d'un arbre de taille fixée, nous en concluons $g_n^{>1} = \sum_{t=2}^{n-1} C_{n-t} g_t$. □

Lemme 19 *Le nombre d'expansions valides de type "but" de l'ensemble des arbres* read-once *de taille c et dépendant des variables $\{x_1,\ldots,x_c\}$ vaut*

$$\lambda^{but}(\mathcal{R}(x_1,\ldots,x_c)) = c! \left(4^{c-1} - \binom{2c-2}{c-1} \right).$$

Preuve: Soit $g(z)$ la fonction génératrice suivante : $g(z) = \sum_n g_n z^n$. D'après le Lemme 18 nous obtenons l'équation fonctionnelle suivante :

$$g(z) = 2zC(z) - 2C(z) + \frac{C(z)g(z)}{z} - g(z),$$

où $C(z)$ énumère les arbres de Catalan. Bien sûr, $C(z) = (1-\sqrt{1-4z})/2$ et nous avons donc :

$$g(z) = \frac{z(1-\sqrt{1-4z})}{1-4z}.$$

Finalement $g_c = [z^c]g(z) = 4^{c-1} - c \cdot C_{c-1}$. Par ailleurs, nous pouvons étiqueter les feuilles de $c!$ manières différentes, donc

$$\lambda^{but}(\mathcal{R}(x_1,\ldots,x_c)) = c!\left(4^{c-1} - \binom{2c-2}{c-1}\right)$$

□

Lemme 20 *Le nombre d'expansions valides de type "prémisse" de l'ensemble des arbres* read-once *de taille c et dépendant des variables $\{x_1,\ldots,x_c\}$ vaut*

$$\lambda^{prem}(\mathcal{R}(x_1,\ldots,x_c)) = c!\ 4^{c-1}.$$

Preuve: Rappelons un peu de terminologie classique : la profondeur gauche d'un noeud correspond au nombre d'arêtes gauches qu'il faut traverser pour aller de la racine de l'arbre jusqu'au noeud en question. La *longueur de cheminement gauche* d'un arbre est défini comme la somme des profondeurs gauches de chacun de ses noeuds (internes et externes). Cette valeur peut être vue comme la somme des *longueur de cheminement gauche interne* et *longueur de cheminement gauche externe* d'un arbre : il s'agit respectivement de restreindre la somme des profondeurs gauches aux noeuds internes et aux noeuds externes. Les longueurs de cheminement droit sont définies de manière symétrique.

Soient p_n le nombre d'expansions de type "prémisse" valides dans l'ensemble des arbres de Catalan (sans étiquette) de taille n, et $p(z)$ la fonction génératrice qui l'énumère. Pour chaque noeud (interne ou externe) ν, nous notons $\delta(\nu)$ sa profondeur gauche. Pour un noeud ν, il existe exactement $\delta(\nu)+1$ feuilles droites τ telles que $\nu \in \Delta(\tau)$. D'après le Lemme 16, nous en concluons qu'il existe exactement $\delta(\nu)+1$ expansions valides de type "prémisse α" en ν. Par conséquent, $p_n = \sum_{A\in\mathcal{F}_k,|A|=n}\sum_{\nu\in A}\delta(\nu)+1$. Par ailleurs, la longueur de cheminement gauche de l'arbre A, notée $p(A)$, est égale à $\sum_{\nu\in A}\delta(\nu)$. Ainsi nous avons $p_n = (\sum_{A,|A|=n}p(A)) + (2n-1)C_{n-1}$.

Nous définissons respectivement $p^e(A)$ ($p^i(A)$) la longueur de cheminement gauche externe (resp. interne) d'un arbre A. Une récurrence sur la taille de A donne : $p^e(A) = p^i(A)+n-1$. En conséquence, $p_n = (\sum_{A,|A|=n}2p^i(A)) + (3n-2)C_{n-1}$. Pour un arbre A, nous définissons son *image miroir* A', obtenue à partir de A en permutant les sous arbres gauches et droits de chaque noeud interne. Il est clair que la longueur de cheminement gauche interne de A et la longueur de cheminement droit interne de A' sont égales. Par conséquent, la somme sur l'ensemble des arbres de leurs longueurs de cheminement gauche interne est exactement la moitié de la somme sur tous les arbres de leurs longueurs de cheminement(gauche et droite) interne. Cette somme est présentée par Flajolet et Sedgewick [FS96, Théorème 5.3, p. 242], par conséquent nous obtenons $p_n = 4^{n-1}$. En s'intéressant aux arbres de $\mathcal{R}(x_1,\ldots,x_c)$, il suffit d'extraire le coefficient d'ordre c de $p(z)$ et puisque nous pouvons étiqueter les feuilles de $c!$ manières différentes, nous avons

$$\lambda^{prem}(\mathcal{R}(x_1,\ldots,x_c)) = c!\ 4^{c-1}.$$

□

5.4.2 Probabilités des fonctions *read-once* de complexité fixée

Proposition 6 *Soit c un entier fixé. La probabilité des fonctions* read-once *de complexité c, lorsque k tend vers l'infini, vaut :*

$$\mu_k(\mathcal{R}_{c,k}) = \left(\frac{1}{2} + (1-c^{-1})\frac{\binom{2c-2}{c-1}}{4^c}\right)\frac{1}{k} + O\left(\frac{1}{k^2}\right).$$

Par ailleurs, la fraction limite moyenne d'une fonction read-once *de complexité c sur k variables, satisfait :*

$$\frac{\mu_k(\mathcal{R}_{c,k})}{|\mathcal{R}_{c,k}|} = \left(\frac{c}{2} + (c-1)\frac{\binom{2c-2}{c-1}}{4^c}\right)\frac{c!}{c^c}\frac{1}{k^{c+1}} + O\left(\frac{1}{k^{c+2}}\right).$$

Preuve: Soit c un entier. Commençons par calculer $\mu_k(\mathcal{R}(x_1,\ldots,x_c))$. Nous avons déjà remarqué que $\lambda^{taut}(\mathcal{R}(x_1,\ldots,x_c)) = c!(2c-1)C_{c-1}$. Ainsi les résultats des Lemmes 19 et 20 prouvent la valeur du nombre d'expansions des arbres *read-once*. Posons

$$\lambda(\mathcal{R}(x_1,\ldots,x_c)) = \lambda^{taut}(\mathcal{R}(x_1,\ldots,x_c)) + \lambda^{but}(\mathcal{R}(x_1,\ldots,x_c)) + \lambda^{prem}(\mathcal{R}(x_1,\ldots,x_c)).$$

Si deux fonctions f_1 et f_2 sont distinctes, alors $E(\mathcal{M}_{f_1})$ et $E(\mathcal{M}_{f_2})$ sont disjoints et le Théorème 2 [p. 24] nous permet de conclure :

$$\begin{aligned}\mu_k(\mathcal{R}(x_1,\ldots,x_c)) &= \frac{\lambda(\mathcal{R}(x_1,\ldots,x_c))}{4^c\ k^{c+1}} + O\left(\frac{1}{k^{c+2}}\right)\\ &= \left(\frac{1}{2} + (1-c^{-1})\frac{\binom{2c-2}{c-1}}{4^c}\right)c!\frac{1}{k^{c+1}} + O\left(\frac{1}{k^{c+2}}\right).\end{aligned}$$

Par symétrie, $\mu_k(\mathcal{R}_{c,k}) = \binom{c}{k}\mu_k(\mathcal{R}(x_1,\ldots,x_c))$, ce qui donne le résultat. Par ailleurs, le fait que le nombre de fonctions *read-once* dépendant de $\{x_1,\ldots,x_c\}$ vaut c^{c-1} prouve immédiatement la valeur moyenne de la probabilité des fonctions *read-once* sur $\{x_1,\ldots,x_c\}$. Par symétrie, elle est égale à la fraction limite moyenne des fonctions *read-once* de complexité c sur $\{x_1,\ldots,x_k\}$, ce qui prouve la seconde équation présentée dans l'énoncé de la proposition. □

5.5 Extensions

5.5.1 Tautologies non simples

Nous allons présenter dans cette sous-section une démonstration alternative permettant de montrer que la plupart des tautologies sont simples, lorsque le nombre de variables k devient grand. Les deux raisons principales qui font que nous présentions cette preuve résident dans le fait qu'elle permet de réduire la différence entre les bornes supérieure et inférieure de la fraction limite des tautologies, et que la preuve est intéressante en soi, car elle met en avant une technique pour exprimer des conditions nécessaires sur la structure des arbres calculant une fonction fixée.

Proposition 7 *La fraction limite des tautologies non simples, notées $\tilde{G}_k$ existe et est telle que*

$$\mu_k(\tilde{G}_k) \leqslant \frac{5}{k^2} + O\left(\frac{1}{k^3}\right).$$

Preuve: Le fait que cette fraction limite existe provient de l'existence des fractions limites des tautologies classiques et des tautologies simples.

Commençons par nous intéresser aux tautologies non simples qui comportent exactement une prémisse dont le but est égal au but global de l'arbre. Soit A une tautologie non simple de but $r(A) = \alpha$. Soit p le nombre de ses prémisses. Nous appelons B la prémisse de A de but $r(A)$, et $\alpha_1, \ldots, \alpha_{p-1}$ les buts des autres prémisses. Du fait des hypothèses, $\alpha_i \neq \alpha$ pour tout $i \in \{1, \ldots, p-1\}$. Puisque A n'est pas une tautologie simple, B ne peut pas être réduite à une feuille. Décomposons $B = (B_1, \ldots, B_m, \alpha)$, avec $m \geqslant 1$. Comme $\overline{B} = B_1 \wedge \ldots \wedge B_m \wedge \overline{\alpha}$, en développant l'expression A, nous obtenons que nécessairement, pour tout $j \in \{1, \ldots, m\}$,

$$B_j \vee \overline{\alpha}_1 \ldots \vee \overline{\alpha}_{p-1} \vee \alpha$$

calcule *Vrai*. Notons $\mathcal{C}_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$ l'ensemble des arbres tels que

$$C \vee \overline{\alpha}_1 \ldots \vee \overline{\alpha}_{p-1} \vee \alpha$$

calcule *Vrai*. Soit $C \in \mathcal{C}_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$.

- Si C est réduit à une feuille γ, alors nécessairement $\gamma \in \{\alpha_1, \ldots, \alpha_{p-1}\}$.
- Sinon, décomposons $C = (C_1, \ldots, C_s, \gamma)$ avec $s \geqslant 1$. Soient $\gamma_i = r(C_i)$. Alors,

$$\overline{\gamma_1} \vee \ldots \vee \overline{\gamma_s} \vee \gamma \vee \overline{\alpha_1} \ldots \vee \overline{\alpha_{p-1}} \vee \alpha$$

doit s'évaluer à *Vrai*. Donc $\alpha \in \{\gamma_1, \ldots, \gamma_s\}$ ou $\gamma \in \{\gamma_1, \ldots, \gamma_s, \alpha_1, \ldots, \alpha_{p-1}\}$.

Définissons la fonction génératrice $c_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$, énumérant plus d'arbres que ceux de $\mathcal{C}_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}$.

$$c_{(\alpha_1,\ldots,\alpha_{p-1},\alpha)}(z) = (p-1)z + \frac{1}{1-((k-1)/k)f_k(z)} \cdot \frac{f_k(z)}{k} \cdot \frac{1}{1-f_k(z)} \cdot kz + \sum_{s=1}^{\infty} f(z)^s \cdot (s+p-1)z.$$

Le premier terme correspond au premier point ci-dessus, et le second terme au deuxième cas $\alpha \in \{\gamma_1, \ldots, \gamma_s\}$ enfin le troisième cas à $\gamma \in \{\gamma_1, \ldots, \gamma_s, \alpha_1, \ldots, \alpha_{p-1}\}$. Cette fonction génératrice ne dépend que de p ; nous la noterons donc c_p. Définissons

$$b_p(z) = \frac{c_p(z)}{1-c_p(z)} \cdot z.$$

Cette fonction génératrice donne une borne supérieure du nombre d'arbres B (pour $p \geqslant 1$ et $\alpha, \alpha_1, \ldots, \alpha_{p-1}$ fixés) tels que

$$B \vee \overline{\alpha}_1 \ldots \vee \overline{\alpha}_{p-1} \vee \alpha$$

calcule *Vrai*. Bien sûr,

$$b_p(z) \leqslant \tilde{b}_p(z) := c_p(z) + \frac{c_p(z)^2}{1-f(z)}.$$

Enfin définissons

$$a_p(z) = p \cdot \left(\frac{k-1}{k} \cdot f(z)\right)^{p-1} \cdot \tilde{b}_p(z) \cdot z \cdot k.$$

Cette fonction génératrice a_p donne une borne supérieure du nombre de tautologies non simples à p prémisses et telle qu'une seule de ces prémisses a pour but $r(A)$. Le facteur z correspond à $r(A) = \alpha$, k correspond au choix de α parmi les littéraux et p correspond au choix de la position de la prémisse de but $r(A)$.

Ainsi, $a(z) = \sum_{p=1}^{\infty} a_p(z)$ énumère plus d'arbres que l'ensemble des tautologies non simples avec exactement une prémisse de but identique au but global. Les calculs basés sur $a(z)$ donnent une fraction limite majorée par $5/k^2 + O(1/k^3)$.

Passons aux tautologies non simples A ayant exactement deux prémisses B_1 et B_2 avec pour buts $r(A)$. Soient $\alpha_1, \ldots, \alpha_{p-2}$ les buts des autres prémisses. Puisque A n'est pas simple, ni B_1 ni B_2 ne sont réduites à une feuille. Soient C la première prémisse de B_1 et D celle de B_2. Soient γ le but de C et $\gamma_1, \ldots, \gamma_s$ les buts de ses prémisses (avec $s \geqslant 0$). Nous définissons de manière analogue les littéraux correspondants de $D : \delta, \delta_1, \ldots, \delta_t$. Du fait que A soit une tautologie, de la même manière que dans la première partie de la preuve, nous obtenons que nécessairement

$$\overline{\gamma_1} \vee \ldots \vee \overline{\gamma_s} \vee \gamma \vee \overline{\delta_1} \vee \ldots \vee \overline{\delta_t} \vee \delta \vee \overline{\alpha_1} \ldots \vee \overline{\alpha_{p-2}} \vee \alpha$$

calcule *Vrai*. Le même méthode que ci-dessus (non détaillée ici) donne une fraction limite majorée par $O(1/k^3)$.

Enfin, il reste les tautologies non simples contenant au moins trois prémisses avec pour but le but global de l'expression. La fonction génératrice les énumérant vaut

$$\left(\frac{1}{1-(k/(k-1))f_k(z)} \cdot \frac{f_k(z)}{k}\right)^3 \cdot \frac{1}{1-f_k(z)} \cdot kz.$$

A nouveau nous obtenons $O(1/k^3)$, ce qui conclut la preuve. □

Il est à remarquer que l'on pourrait obtenir exactement cette borne en utilisant les tautologies simples, les non-tautologies simples et en modifiant légèrement les non-tautologies moins simples. En effet, nous avions émis des conditions sur la première prémisse du sous-arbre C (voir Figure 5.5 [p. 29]), mais il suffit qu'une des prémisses de C vérifie ces conditions pour agrandir suffisamment la familles des non-tautologies moins simples et obtenir la même borne que celle qui vient d'être démontrée sur les tautologies non simples.

5.5.2 Tautologies et littéraux positifs et négatifs

La méthode que nous avons développée dans la Section 5.1 peut être étendue à la logique de l'implication avec les littéraux positifs et négatifs. Dans cette nouvelle configuration, nous allons démontrer à nouveau que lorsque le nombre de variables devient grand, la plupart des tautologies ont une structure très simple. Plus précisément, la plupart des tautologies ont une de leur prémisse égale au but (comme dans le modèle précédant) ou ont deux prémisses réduite à une feuille et étiquetées par des littéraux opposés.

Comme précédemment, le nombre k est le nombre de variables et nous considérons les arbres construits avec l'unique connecteur implication et ayant pour étiquette des feuilles : $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$. Afin que toutes les familles suivantes soient bien définies, nous prenons $k \geqslant 2$. Nous notons $\mathcal{F}_k$ l'ensemble de ces arbres. Ainsi, nous avons $C_{n-1}(2k)^n$ arbres de taille n et dépendant de k. Nous appelons à nouveau $f_k(z)$ la fonction génératrice énumérant tous les arbres. Il est clair que $f_k(z) = \frac{1-\sqrt{1-8kz}}{2}$. Pour un ensemble $\mathcal{A} \subseteq \mathcal{F}_k$, la fraction limite $\mu_k(\mathcal{A})$ de $\mathcal{A}$ dans $\mathcal{F}_k$ est définie comme précédemment.

Nous remarquons que ce système de connecteur et de variables n'est toujours pas complet (toute fonction Booléenne à k variable n'est pas expressible). Nous vérifions sans problème que les conditions du théorème de Drmota-Lalley-Woods sont vérifiées et se démontrent de la même manière que pour le Théorème 1 [p. 15].

Commençons par définir plus précisément les familles de tautologies simples. Elles sont de deux types : Les tautologies simples du premier type $G_k^{(1)}$ sont les expressions $A_1, \ldots, A_p \to \alpha$ où il existe $i \in \{1, \ldots, p\}$ tel que $A_i = \alpha$; il s'agit de toutes les expressions de la forme : $\ldots, \ell, \ldots \to \ell$, pour un littéral ℓ. Les tautologies simples du second type $G_k^{(2)}$ sont les expressions $A_1, \ldots, A_p \to \alpha$, où il existe i et j tels que A_i et A_j sont des littéraux opposés ; il s'agit de toutes les expressions de la forme : $\ldots, \ell, \ldots, \bar{\ell}, \ldots \to \cdot$, pour un littéral ℓ et tel que $\bar{\ell}$ est la négation de ℓ (i.e. $\bar{\ell} = \bar{x}_i$ if $\ell = x_i$ et $\bar{\ell} = x_i$ if $\ell = \bar{x}_i$).

Nous remarquons que les deux familles de tautologies simples ne sont pas disjointes. Finalement nous définissons $G_k = G_k^{(1)} \cup G_k^{(2)}$ et Cl_k l'ensemble de toutes les tautologies.

Lemme 21 *La fraction limite des tautologies simples du premier type existe et est égale à*

$$\mu_k(G_k^{(1)}) = \frac{1}{2k} + O\left(\frac{1}{k^2}\right).$$

Preuve: Une tautologie du premier type relativement au littéral ℓ peut être décomposée en l'expression non ambiguë suivante : une suite de prémisses différentes de ℓ, suivies de ℓ, puis une suite quelconque de prémisses et enfin un but ℓ. Nous pouvons l'écrire, dans le contexte des langages formels sous la forme :

$$G_k^{(1)}[\ell] = \mathcal{E}_\ell^* \cdot \ell \cdot \mathcal{F}_k^* \cdot \ell, \qquad \text{avec} \qquad \mathcal{E}_\ell = \mathcal{F}_k \setminus \{\ell\}.$$

Du fait que nous avons $2k$ choix pour le littéral ℓ, la fonction génératrice énumérant $G_k^{(1)}$ est égale à

$$2k \cdot \frac{z}{1 - (f_k(z) - z)} \cdot \frac{z}{1 - f_k(z)}.$$

Un calcul similaire à ceux présentés dans la Section 5.1 donne le résultat :

$$\mu_k(G_k^{(1)}) = \frac{8k+1}{(4k+1)^2} = \frac{1}{2k} + O\left(\frac{1}{k^2}\right).$$

□

Tournons-nous désormais vers les tautologies simples du second type. Avant d'en calculer la fraction limite, introduisons les définitions suivantes. Pour un arbre A, soit $\mathcal{G}(A)$ le multi-ensemble contenant les étiquettes du but et des sous-buts de A ; i.e. $\mathcal{G}(A_1, \ldots, A_p \to \alpha) = \{\{r(A_1), \ldots, r(A_p), \alpha\}\}$. Nous dirons qu'une variable $x \in \{x_1, \ldots, x_k\}$ a r *répétitions* dans le multi-ensemble de littéraux M si au moins r éléments de M appartiennent à $\{x, \bar{x}\}$. Soit $\mathcal{R}$ l'ensemble d'arbres $A \in \mathcal{F}_k$ tels que $\mathcal{G}(A)$ a au moins deux répétitions d'une variable, ou une répétition de deux variables distinctes.

Lemme 22 *Nous avons :* $\mu_k^+(\mathcal{R}) = O(1/k^2)$.

Preuve: Pour $x \in \{x_1, \ldots, x_k\}$, soit $\mathcal{R}^x$ l'ensemble des formules $A \in \mathcal{F}_k$ telles que x a deux répétitions dans $\mathcal{G}(A)$. En considérant le fait que la variable répétée apparaît dans trois sous-buts ou dans deux sous-buts et le but, la fonction génératrice $\phi_{\mathcal{R}^x}$ satisfait

$$\phi_{\mathcal{R}^x}(z) \prec \left(\frac{f_k(z)/(2k)}{1 - f_k(z)}\right)^3 \cdot f_k(z) + \left(\frac{f_k(z)/(2k)}{1 - f_k(z)}\right)^2 \frac{2z}{1 - f_k(z)}.$$

Des calculs simples nous donnent $\mu^+(\mathcal{R}^x) = O(1/k^3)$. De la même manière, pour deux variables distinctes $x \neq y$, l'ensemble $\mathcal{R}^{x,y}$ d'arbres $A \in \mathcal{F}_k$ tels que les deux variables x et y sont répétées dans $\mathcal{G}(A)$ satisfait $\mu^+(\mathcal{R}^{x,y}) = O(1/k^4)$. Puisque $\mathcal{R} = \bigcup_x \mathcal{R}^x \cup \bigcup_{x \neq y} \mathcal{R}^{x,y}$, nous en concluons que $\mu^+(\mathcal{R}) \leqslant k\ O(1/k^3) + \binom{k}{2}\ O(1/k^4) = O(1/k^2)$. □

Lemme 23 *La fraction limite des tautologies simples du second type existe et est égale à*

$$\mu_k(G_k^{(2)}) = \frac{3}{8k} + O\left(\frac{1}{k^2}\right).$$

Preuve: Pour un arbre A, nous définissons $\mathcal{U}(A)$ le multi-ensemble du but de A et des étiquettes de ses prémisses de taille 1. Donc

$$\mathcal{U}(A_1, \ldots, A_p \to \alpha) = \{\{A_i \ : \ |A_i| = 1\}\} \cup \{\{\alpha\}\}.$$

Soit $\tilde{G}_k^{(2)}$ l'ensemble d'arbres $A \in \mathcal{F}_k$ tels que deux de leurs prémisses sont des littérauxo opposés, et aucune autre répétition n'apparaît dans $\mathcal{U}(A)$. Plus précisément, $(A_1, \ldots, A_p \to \alpha) \in \tilde{G}_k^{(2)}$ s'il n'existe aucun littéral ℓ et $1 \leqslant i < j \leqslant p$ tels que $A_i = \ell$, $A_j = \bar{\ell}$, et tels qu'il n'y ait aucune répétition dans $\mathcal{U}(A) \setminus \{\{\bar{\ell}\}\}$. Remarquons que $\tilde{G}_k^{(2)} \subseteq G_k^{(2)} \setminus G_k^{(1)}$.

Soit $T_{\ell,\alpha,t}$ l'ensemble des expressions $A \in \tilde{G}_k^{(2)}$ satisfaisant les conditions suivantes :
– ℓ et $\bar{\ell}$ apparaissent exactement une fois en tant que prémisse dans A, dans cet ordre ;
– α est le but de A ;
– A a exactement $t+2$ prémisses de taille 1.
Le sous-ensemble $T_{\ell,\alpha,t}[y_1, \ldots, y_{t+2}]$ de $T_{\ell,\alpha,t}$ est défini comme l'ensemble d'expressions telles que y_i est la $i$ème prémisse parmi celle égale à un littéral – pour tout $1 \leqslant i \leqslant t+2$.

Soit $\mathcal{E}_L$ l'ensemble d'expressions différentes d'un littéral, i.e. de taille au moins 2. L'ensemble $T_{\ell,\alpha,t}[y_1, \ldots, y_{t+2}]$ est décrit par l'expression non ambiguë suivante :

$$\mathcal{E}_L^* \cdot y_1 \cdot \ldots \cdot \mathcal{E}_L^* \cdot y_{t+2} \cdot \mathcal{E}_L^* \cdot \alpha.$$

Nous pouvons désormais énumérer cet ensemble. La fonction génératrice énumérant $\mathcal{E}_L$ est $f_k(z) - 2kz$; en utilisant l'équation définissant f_k, nous pouvons réécrire cette fonction génératrice comme $f_k(z)^2$. En conséquence, la fonction génératrice pour $T_{\ell,\alpha,t}[y_1, \ldots, y_{t+2}]$ est :

$$\sigma(z) = \left(\frac{z}{1 - f_k(z)^2}\right)^{t+3}.$$

$$\frac{z}{1 - f^2(z)} = \frac{1 + 4kz - \sqrt{1 - 8kz}}{8k + 8k^2 z},$$

nous pouvons réécrire $\sigma(z)$ comme

$$\sigma(z) = \left(\frac{1}{8k + 8k^2 z}\right)^{t+3} \left(P(z) - \sqrt{1 - 8kz} \sum_{s=0}^{\lfloor \frac{t+2}{2} \rfloor} \binom{t+3}{2s+1} (1 - 8kz)^s (1 + 4kz)^{t-2s+2}\right),$$

où $P(z)$ est le polynôme adéquat. La singularité dominante de σ est $1/(8k)$. Ainsi, le comportement asymptotique de $[z^n]\sigma$ est donné par le premier terme de la somme ($s = 0$), et nous concluons

$$[z^n]\sigma(z) = \frac{-(t+3)(3/2)^{t+2}}{(9k)^{t+3}} [z^n]\{\sqrt{1 - 8kz}\} \ (1 + O(1/n)).$$

Comme dans la preuve du Lemme 21, nous obtenons que la fraction limite de $T_{\ell,\alpha,t}[y_1, \ldots, y_{t+2}]$ est bien définie et sa valeur vaut $\mu_k(T_{\ell,\alpha,t}[y_1, \ldots, y_{t+2}]) = (4/3)(t+3)/(6k)^{t+3} =: \theta_t$.

Puisque les ensemble $T_{\ell,\alpha,t}[y_1,\ldots,y_{t+2}]$ (pour tout $0 \leqslant t \leqslant k-2$ et tout $(\alpha,\ell,y_1,\ldots,y_{t+2})$ satisfaisant les bonnes conditions) forment une partition de $\tilde{G}_k^{(2)}$, la fraction limite de $\tilde{G}_k^{(2)}$ existe et peut être calculée : pour $t \in \{0,\ldots,k-2\}$, une classe $T_{\ell,\alpha,t}[y_1,\ldots,y_{t+2}]$ est obtenue en choisissant les littéral ℓ, les positions de ℓ et $\bar{\ell}$ dans la suite de prémisses de taille 1, puis les $t+1$ literraux $\alpha, y_1,\ldots,y_{t+2}$ différents de ℓ et $\bar{\ell}$. Par conséquent la fraction limite de $\tilde{G}_k^{(2)}$ est égale à :

$$\begin{aligned}\mu_k(\tilde{G}_k^{(2)}) &= \sum_{t=0}^{k-2} 2k\binom{t+2}{2}\cdot(k-1)\ldots(k-t-1)\cdot 2^{t+1}\cdot\theta_t \\ &= \sum_{t=0}^{k-2}\frac{(t+3)(t+2)(t+1)}{3^{t+4}k}\cdot\frac{(k-1)\ldots(k-(t+1))}{k^{t+1}}.\end{aligned}$$

Nous devons estimer le comportement asymptotique de la somme précédente. Le fait que $(k-1)\ldots(k-t-1)/k^{t+1} \leqslant 1$ donne une borne supérieure, ainsi, $\mu_k(\tilde{G}_k^{(2)}) \leqslant 3/(8k)$. D'un autre côté, $(k-1)\ldots(k-t-1)/k^{t+1} \geqslant ((k-t-1)/k)^{t+1} = (1-(t+1)/k)^{t+1} \geqslant 1-(t+1)^2/k$ donne une borne inférieure : $\mu_k(\tilde{G}_k^{(2)}) \geqslant 3/(8k) + O(1/k^2)$. Ainsi nous concluons

$$\mu_k(\tilde{G}_k^{(2)}) = \frac{3}{8k} + O\left(\frac{1}{k^2}\right).$$

De la même manière que nous l'avons fait pour $\tilde{G}_k^{(2)}$, nous pouvons partitionner $G_k^{(2)} \setminus (G_k^{(1)} \cup \tilde{G}_k^{(2)})$ en un nombre fini d'ensemble (dépendant de k), tous ayant une fraction limité bien définie, en considérant la suite des prémisses de taille 1 d'un arbre, tronquée à la position adéquate. Ainsi l'ensemble $G_k^{(2)} \setminus (G_k^{(1)} \cup \tilde{G}_k^{(2)})$ a aussi une fraction limite bien définie. Ceci montre que la fraction limite de $G_k^{(2)} \setminus G_k^{(1)}$ existe ; avec le Lemme 21, il s'en suit que $\mu_k(G_k^{(2)})$ est bien définie.

Afin d'estimer la fraction limite de $G_k^{(2)}$, nous prenons un raccourci : puisque $G_k^{(2)} \setminus (G_k^{(1)} \cup \tilde{G}_k^{(2)}) \subseteq \mathcal{R}$, d'après le Lemme 22 sa fraction limite $O(1/k^2)$. Ceci montre que $\mu_k(G_k^{(2)}) = 3/(8k) + O(1/k^2)$. □

Lemme 24 *Les tautologies simples G_k ont une fraction limite qui existe et qui vaut*

$$\mu_k(G_k) = \frac{7}{8k} + O\left(\frac{1}{k^2}\right).$$

Preuve: Il a été montré dans la preuve du Lemme 23 que $G_k^{(2)} \setminus G_k^{(1)}$ admet une fraction limite égale à $3/8 + O(1/k^2)$. Ajouté au Lemme 21, nous en concluons le résultat. □

Démontrons désormais que la fraction limite des autres tautologie est d'ordre inférieur.

Corollaire 2 *Les tautologies non simples ont une fraction limite qui existe et qui vaut $O(1/k^2)$.*

Preuve: L'existence de la fraction limite des tautologies non simples provient de l'existence des fractions limites de l'ensemble des tautologies et de l'ensemble des tautologies simples.

Soit $A = A_1,\ldots,A_p \to \ell$ une tautologie. Notons $\alpha_i = r(A_i)$. Nécessairement (voir Section 5.5.1), $\bar{\alpha}_1 \vee \ldots \vee \bar{\alpha}_p \vee \ell$ calcule *vrai*. Ainsi il existe i tel que $\alpha_i = \ell$ ou il existe $i \neq j$

tels que $\alpha_i = \bar{\alpha}_j$. Nous pouvons supposer qu'il s'agit de l'unique répétition de variable dans $\mathcal{G}(A) = \{\{\alpha_1, \ldots, \alpha_p, \ell\}\}$. En effet, via le Lemme 22, $\mu_k^+(\mathcal{R}) = O(1/k^2)$.

Le lemme découle des deux affirmations suivantes.

Affirmation 1. L'ensemble $\mathcal{N}_1$ des tautologies non simples $A \in Cl_k \setminus (G_k \cup \mathcal{R})$ telles qu'un de leur sous-but soit égal au but de A satisfait : $\mu_k^+(\mathcal{N}_1) = O(1/k^2)$.

Affirmation 2. L'ensemble $\mathcal{N}_2$ des tautologies non simples $A \in Cl_k \setminus (G_k \cup \mathcal{R})$ telles que deux de ses sous-buts sont des littéraux opposés satisfait : $\mu_k^+(\mathcal{N}_2) = O(1/k^2)$.

Démontrons l'Affirmation 1. Soit une tautologie non simple de la forme $A_1, \ldots, A_p \to \ell$ et satisfaisant les hypothèses de l'Affirmation 1. Soit $\alpha_i = r(A_i)$. Supposons qu'il existe un unique i_0 tel que $\alpha_{i_0} = \ell$. De plus, $A_{i_0} \neq \ell$. Soit $B = A_{i_0}$. L'arbre B est de la forme $B_1, \ldots, B_q \to \ell$ avec $q \geqslant 1$. Soit $C = B_1$; à nouveau développons C : $C = C_1, \ldots, C_s \to \gamma$. Soit $\gamma = r(C)$ et $\gamma_i = r(C_i)$ – notons que nous n'avons pas besoin de supposer $s \geqslant 1$, i.e. C pourrait être réduit à une feuille γ (correspondant au cas $s = 0$).

De la même manière que dans la Section 5.5.1, nous obtenons cette expression

$$\bigvee_{1 \leqslant i \leqslant s} \bar{\gamma}_i \vee \gamma \vee \bigvee_{1 \leqslant i \leqslant p,\ i \neq i_0} \bar{\alpha}_i \vee \ell \tag{5.6}$$

qui dit s'évaluer à *vrai*.

Il est clair que le sous-ensemble d'arbres avec $\gamma \in \{\ell, \bar{\ell}\}$ admet une fraction limite égale à $O(1/k^2)$. Par conséquent, il nous reste à borner l'ensemble des autres arbres, i.e. ceux avec la propriété additionnelle que $\gamma \notin \{\ell, \bar{\ell}\}$. Pour cela, nous allons calculer une fonction génératrice $a(z)$ telle que les coefficients $[z^n]a(z)$ sont des bornes supérieures des nombres d'arbres que nous considérons. Commençons par fixer p, s, ℓ et γ. Le nombre de sous-arbres C possibles est majoré par le nombre d'arbres énumérés par la fonction génératrice suivante :

$$c_{p,s}(z) = \left(\frac{1}{1 - (f_k(z) - z)}\right)^p \cdot z \cdot n_{p,s}$$

où $n_{p,s}$ est un nombre entier que nous préciserons plus tard. Ainsi, le nombre de sous-arbres B est borné par cette fonction génératrice

$$b_{p,s}(z) = c_{p,s}(z) \cdot \frac{1}{1 - f_k(z)} \cdot z,$$

et le nombre total d'arbres A, par rapport aux nombres fixés p, s, ℓ, γ, est borné par

$$a_{p,s}(z) = p \cdot \left(\frac{1}{1 - (f_k(z) - z)}\right)^{p-1} \cdot z \cdot b_{p,s}(z).$$

D'après l'équation (5.6 [p. 57]) nous avons (au moins) un des α_i ou γ_i égal à $\bar{\ell}$ ou $\bar{\gamma}$, ou (au moins) un couple de littéraux parmi $\{\alpha_i \mid 1 \leqslant i \leqslant p, i \neq i_0\} \cup \{\gamma_i \mid 1 \leqslant i \leqslant s\}$ sont égaux. Alors nous pouvons poser

$$n_{p,s} = 2(2k)^{p-2+s}(p - 1 + s) + 2k(2k)^{p-3+s}\binom{p-1+s}{2}.$$

Il reste à définir

$$a(z) = 2k(2k-2)\sum_{s=0}^{\infty}\sum_{p=1}^{\infty} a_{p,s}(z)$$

où $2k$ correspond au choix de ℓ et $2k-2$ au choix de γ. Cette fonction génératrice "borne" le nombre total de tautologies non simples considérées par l'Affirmation 1. Un calcul simple concernant la fonction génératrice $a(z)$ montre que ces arbres ont une fraction limite égale à $O(1/k^2)$. Ce qui conclut le preuve de l'Affirmation 1. Le preuve de l'Affirmation 2 est très similaire et est laissée au lecteur. □

Proposition 8 *Dans le modèle de l'implication avec les littéraux positifs et négatifs, asymptotiquement (quand le nombre de variables tend vers l'infini), toutes les tautologies sont simples, i.e.*

$$\lim_{k\to\infty} \frac{\mu_k(G_k)}{\mu_k(Cl_k)} = 1.$$

De plus, la fraction limite des tautologies est égale à $\mu_k(Cl_k) = 7/(8k) + O(1/k^2)$.

Preuve: La preuve est directe à partir du Lemme 24 et du Corollaire 2. □

Chapitre 6

Processus de branchement

Dans ce modèle, la distribution de probabilité sur les fonctions Booléenne est induite par une distribution de probabilité sur les arbres obtenus via un processus de Galton-Watson critique et un étiquetage des feuilles uniforme et indépendant avec l'ensemble $\{x_1, \ldots, x_k\}$. Un arbre est fini presque sûrement dans ce modèle [AN72]. Pour un arbre $A \in \mathcal{F}_k$, nous avons

$$\pi_k(A) = \mathbb{P}(\text{structure de } A)\mathbb{P}(\text{étiquetage de } A) = \frac{1}{2^{2|A|-1}\ k^{|A|}}.$$

Il est clair que la probabilité d'un ensemble d'arbres vaut la somme des probabilités de chaque arbre de l'ensemble. Nous définissons la distribution de probabilité induite sur les fonctions Booléennes. Pour une fonction Booléenne f sur $\mathcal{B}_k$, la probabilité de f est égale à celle de l'ensemble des expressions la représentant $\mathcal{F}_k(f)$. Nous remarquons aussi que $\pi_k(\mathcal{F}_k) = 1$ et qu'il n'y a pas de problème d'existence de probabilité d'un ensemble quelconque d'arbres, comme dans le cas des grands arbres (ici, nous avons bien une mesure de probabilité sur $\mathcal{F}_k$ ce qui n'était pas le cas avec le modèle des grands arbres).

Théorème 3 *Pour le modèle de branchement, la plupart des tautologies sont "simples" et par conséquent :*

$$\pi_k(\text{Vrai}) = \frac{1}{2k} + O\left(\frac{1}{k^2}\right).$$

Soit f une fonction Booléenne distincte de Vrai. *Pour le modèle de branchement, la plupart des arbres calculant f appartiennent aux arbres minimaux de f (lorsque $k \to \infty$) :*

$$\pi_k(f) = \pi_k(\mathcal{M}_f) + O\left(\frac{1}{k^{L(f)+1}}\right).$$

De plus, la probabilité de f est vérifie asymptotiquement les inégalités suivantes :

$$\pi_k(f) \geqslant \frac{|\mathcal{M}_f|}{2^{2L(f)-1}\ k^{L(f)}} + \frac{|\mathcal{P}_{f,2}| + 2L(f) - 1}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right),$$

$$\pi_k(f) \leqslant \frac{|\mathcal{M}_f|}{2^{2L(f)-1}\ k^{L(f)}} + \frac{|\mathcal{P}_{f,2}| + 4(2L(f) - 1)}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Nous allons démontrer que dans ce modèle, lorsque le nombre k de variables devient grand, la plupart des arbres calculant une fonction donnée sont ses arbres minimaux. Toutefois, le travail effectué dans le Chapitre 5 par rapport aux expansions des arbres irréductibles ne va pas être laissé de côté car il nous permettra d'obtenir des bornes inférieure et supérieure sur le second terme du développement de la probabilité d'une fonction.

6.1 Probabilité des tautologies

Dans cette partie nous utiliser à nouveau les trois ensembles d'expressions : les non-tautologies simples, les tautologies simples ainsi que les tautologies moins simples définies dans la partie précédente – cf. Définition 1 [p. 25]. Nous allons tout d'abord démontrer que ces familles regroupent la plupart des tautologies lorsque k tend vers l'infini.

Rappelons que les *non-tautologies simples* sont les arbres tels que leurs sous-buts sont tous distincts du but global α. Soit SN_k l'ensemble de ces arbres, déterminons-en la probabilité.

$$\begin{aligned}\pi_k(SN_k) &= \sum_{\alpha}\sum_{p=0}^{\infty}\frac{1}{2^p}\left(1-\frac{1}{k}\right)^p\frac{1}{2k} = \sum_{\alpha}\frac{1}{k+1}\\ &= 1-\frac{1}{k+1} = 1-\frac{1}{k}+O\left(\frac{1}{k^2}\right).\end{aligned}$$

Les *tautologies simples* sont les arbres dont l'une des prémisses est égale au but de l'arbre. Nous notons G_k l'ensemble de ces arbres. Un arbre n'est pas une tautologie simple si et seulement si toutes ses prémisses sont différentes de son but, donc :

$$\begin{aligned}\pi_k(\mathcal{F}_k \setminus G_k) &= \sum_{\alpha}\sum_{p=0}^{\infty}\frac{1}{2^p}\left(1-\frac{1}{2k}\right)^p\frac{1}{2k}\\ &= 1-\frac{1}{2k+1}.\end{aligned}$$

Ainsi nous en concluons :

$$\pi_k(G_k) = \frac{1}{2k+1} = \frac{1}{2k}+O\left(\frac{1}{k^2}\right).$$

Enfin, soient α et β deux variables distinctes. Posons $LN_k^{\alpha,\beta}$ l'ensemble des *non-tautologies moins simples*, qui sont les arbres dont exactement une prémisse C a comme but le but global de l'arbre et satisfait les deux conditions suivantes : sa première prémisse C_1 a pour but β et les autres sous-buts de C sont distincts de α et β. Les autres prémisses de l'arbre global ont des buts différents de α et β. Du fait que cette décomposition est unique nous pouvons aisément calculer la probabilité de chaque prémisse de ces arbres : Soit $\mathcal{C}_1$ l'ensemble des arbres vérifiant les conditions de la prémisse C_1.

$$\pi_k(\mathcal{C}_1) = \sum_{p=0}^{\infty}\frac{1}{2^p}\left(1-\frac{1}{k}\right)^p\frac{1}{2k} = \frac{1}{k+2}.$$

Calculons maintenant la probabilité de l'ensemble d'arbres $\mathcal{C}$ vérifiant les conditions sur C :

$$\pi_k(\mathcal{C}) = \frac{1}{2}\cdot\pi_k(\mathcal{C}_1)\cdot\frac{1}{k} = \frac{1}{2k(k+2)}.$$

Et finalement,

$$\begin{aligned}\pi_k(LN_k) &= \sum_{\{\alpha,\beta\}}\left(\sum_{p=0}^{\infty}\frac{1}{2^p}\left(1-\frac{2}{k}\right)^p\right)\frac{1}{2}\pi_k(\mathcal{C})\left(\sum_{p=0}^{\infty}\frac{1}{2^p}\left(1-\frac{2}{k}\right)^p\right)\frac{1}{2k}\\ &= \frac{1}{2k}+O\left(\frac{1}{k^2}\right).\end{aligned}$$

6.2 Probabilité des fonctions autres que *Vrai*

Soit f une fonction de $\mathcal{B}_k$, distincte de *Vrai*. Les ensembles $\mathcal{M}_f$ et $\mathcal{P}_{f,1}$ à $\mathcal{P}_{f,5}$ sont ceux définis dans la Section 5.2.2. Notons à nouveau $\lambda^t(f)$ le nombre d'expansions valides de type t des arbres de $\mathcal{M}_f$.

Nous suivons la structure de la Section 5.2 et une adaptation immédiate nous donnera la preuve du Théorème 3. L'application *extension* X et les ensembles $\mathcal{B}_q^p(\mathcal{V})$ et $\mathcal{A}_q^p(\mathcal{V})$ sont ceux présentés dans la même Section 5.2. Dans le modèle des grands arbres, l'idée principale est le fait que trop de répétitions de variables, avec une petite profondeur gauche, réduisent de façon significative le nombre d'arbres. Cette idée est toujours valide dans le modèle des processus de branchement ; voilà l'analogue du Lemme 6 [p. 34].

Lemme 25

$$\pi_k\left(X\left(\mathcal{B}_q^p(\mathcal{V})\right)\right) = O\left(\frac{1}{k^p}\right).$$

Preuve: Rappelons que la structure d'un arbre est obtenue par un processus de branchement, puis l'arbre est étiqueté de manière uniforme et indépendante. Nous obtenons donc

$$\begin{aligned}\pi_k\left(X\left(\mathcal{B}_q^p(\mathcal{V})\right)\right) &\leqslant \sum_{A\in\mathcal{B}_q^p(\mathcal{V})} \pi_k(X(A)) = \sum_{A\in\mathcal{B}_q^p(\mathcal{V})} \pi_k(A)\left(\sum_{i=0}^{\infty}\frac{1}{2^i}\right)^{2|A|-1}\\ &\leqslant \sum_{A\in\mathcal{B}_q^p(\mathcal{V})} \frac{1}{2^{2|A|-1}\ k^{|A|}} 2^{2|A|-1}.\end{aligned}$$

Par ailleurs, la taille de chaque arbre de $\mathcal{B}_q^p(\mathcal{V})$ est supérieure ou égale à p et l'ensemble $\mathcal{B}_q^p(\mathcal{V})$ est fini. Finalement,

$$\pi_k\left(X\left(\mathcal{B}_q^p(\mathcal{V})\right)\right) = O\left(\frac{1}{k^p}\right).$$

□

Le Lemme 5 [p. 34] relie plusieurs expansions de $\mathcal{A}_q^p(\mathcal{V})$ aux extensions de $\mathcal{B}_q^p(\mathcal{V})$. Il est clair que ce lemme dépend de la structure des arbres, il est toujours valide pour le modèle de branchement et nous pouvons directement établir le corollaire suivant :

Corollaire 3

$$\pi_k\left(E^*\left(\mathcal{A}_q^p(\mathcal{V})\right)\right) = O\left(\frac{1}{k^p}\right).$$

Dans le cas de la fraction limite, les tautologies, les arbres ayant un but α et ceux avec une prémisse α ont asymptotiquement la même fraction limite. Ce n'est plus le cas dans ce modèle. Toutefois, leurs probabilités sont toujours du même ordre et par conséquent, les expansions des trois types ("tautologie", but "α" et prémisse "α") vont contribuer au second ordre de la probabilité de f. Par contre, dans le cas des grands arbres, puisque les expressions qui comptent sont celles dont la taille est grande, chacun des types d'expansions définis des ensembles d'expressions disjoints. Ce n'est plus la cas dans ce nouveau modèle d'arbres, et pour cette raison, seules des bornes inférieure et supérieure sur le second terme de la probabilité d'une fonction vont être démontrées. La Figure 6.1 présente, sur la gauche, l'arbre obtenu après une expansion des deux arbres minimaux de droite. Le premier arbre minimal donne

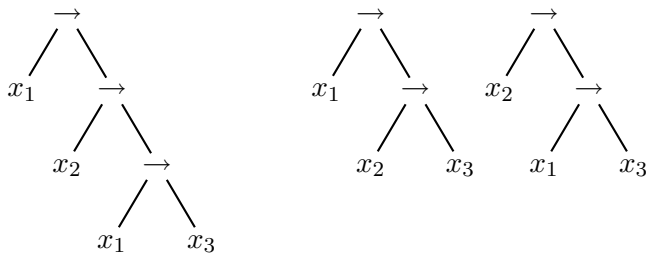

FIG. 6.1 – L'arbre de gauche est obtenu après une expansion de deux arbres minimaux distincts.

l'arbre de gauche après une expansion de type but "x_1" en le noeud étiqueté par x_3. Le second arbre minimal donne l'arbre de gauche après une expansion de type but "x_1" de sa racine.

Les Lemmes 7 [p. 36], 8 [p. 37], 10 [p. 39] et 11 [p. 39] décrivent des propriétés des arbres irréductibles et sont donc toujours valables pour ce contexte. A partir de maintenant, concentrons-nous sur les expansions des arbres irréductibles. Nous remarquons immédiatement que les "petits" arbres jouent un rôle important, du fait de leur probabilité élevée.

Lemme 26 *Soit f une fonction Booléenne distincte de* Vrai*,*

$$\pi_k\left(E^*(\mathcal{M}_f)\right) = \pi_k(\mathcal{M}_f) + \pi_k(E(\mathcal{M}_f)) + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Par conséquent

$$\pi_k\left(E^*(\mathcal{M}_f)\right)) \geqslant \frac{|\mathcal{M}_f|}{2^{2L(f)-1}\ k^{L(f)}} + \frac{2L(f)-1}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right),$$

$$\pi_k\left(E^*(\mathcal{M}_f)\right)) \leqslant \frac{|\mathcal{M}_f|}{2^{2L(f)-1}\ k^{L(f)}} + \frac{4(2L(f)-1)}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).$$

Preuve: Soit $A \in \mathcal{M}_f$, donc $|A| = L(f)$ et $\pi_k(A) = 1/(2^{2L(f)-1}k^{L(f)})$. En conséquence, $\pi_k(\mathcal{M}_f)$ correspond à la taille de $\mathcal{M}_f$ multipliée par la probabilité précédente.

Intéressons-nous à la borne inférieure sur $\pi_k(E(\mathcal{M}_f))$. Une expansion valide de type "tautologie" de deux arbres minimaux différents engendre deux ensembles d'arbres disjoints. Par conséquent :

$$\begin{aligned}
\pi_k(E(\mathcal{M}_f)) &\geqslant \frac{\lambda^{taut}(f)}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right), \\
\pi_k(E(\mathcal{M}_f)) &\geqslant \frac{2L(f)-1}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).
\end{aligned}$$

Le fait que des expansions valides ne forment pas des ensembles disjoints, en sommant les cardinalités des ensembles obtenus après expansions, nous obtenons une borne supérieure. Ainsi, en modifiant légèrement la preuve du Lemme 9 [p. 38], nous concluons :

$$\begin{aligned}
\pi_k(E(\mathcal{M}_f)) &\leqslant \frac{\lambda^{taut}(f) + \lambda^{prem}(f) + 2\lambda^{but}(f)}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right), \\
\pi_k(E(\mathcal{M}_f)) &\leqslant \frac{4(2L(f)-1)}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).
\end{aligned}$$

Finalement, des arguments analogues à ceux utilisés pour la preuve du Lemme 14 [p. 44], justifient que la probabilité des arbres obtenus après au moins deux expansions des arbres minimaux est égale à $O(1/k^{L(f)+2})$. □

Puisque les "petits" arbres ont une probabilité non négligeable (lorsque k tend vers l'infini), l'ensemble $\mathcal{P}_{f,2}$ devient désormais significatif pour la probabilité de f.

Lemme 27 *Soit f une fonction Booléenne distincte de* Vrai*,*

$$\begin{aligned}\pi_k(E^*(\mathcal{P}_{f,2})) &= \pi_k(\mathcal{P}_{f,2}) + O\left(\frac{1}{k^{L(f)+2}}\right)\\ \pi_k(E^*(\mathcal{P}_{f,2})) &= \frac{|\mathcal{P}_{f,2}|}{2^{2L(f)+1}\ k^{L(f)+1}} + O\left(\frac{1}{k^{L(f)+2}}\right).\end{aligned}$$

Preuve: Rappelons que les arbres de $\mathcal{P}_{f,2}$ sont de taille $L(f)+1$ et ont toutes leurs feuilles étiquetées par des variables essentielles. Par ailleurs, le terme d'erreur est obtenu de la même manière que celui du Lemme 12 [p. 40]. □

Enfin, tous les arbres de $\mathcal{P}_{f,4} \cup \mathcal{P}_{f,5}$ contiennent trop de répétitions avec une petite profondeur gauche et donc nous obtenons :

Lemme 28 *Soit f une fonction Booléenne distincte de* Vrai*,*

$$\pi_k(E^*(\mathcal{P}_{f,4} \cup \mathcal{P}_{f,5})) = O\left(\frac{1}{k^{L(f)+2}}\right).$$

Preuve: A nouveau, la preuve est similaire à celle du Lemme 13 [p. 41]. □

6.3 Synthèse des expansions d'arbres irréductibles

Nous avons désormais les outils permettant de démontrer le Théorème 3 [p. 59].

La première partie relative à la fonction *Vrai*, découle de la Section 6.1. Du fait que

$$G_k \subset Cl_k \subset \mathcal{F}_k \setminus (SN_k \cup LN_k),$$

on conclut :

$$\pi_k(Vrai) = \frac{1}{2k} + O\left(\frac{1}{k^2}\right).$$

Dans la seconde partie, il s'agit de calculer la fraction limite des fonctions f fixées, autres que *Vrai*. Il est clair que les arbres calculant f tombent dans répétition d'expansions valides d'un arbre irréductible de f. Donc l'ensemble des arbres calculant f est exactement $\mathcal{F}_k(f) = E^*(\mathcal{M}_f \cup \mathcal{P}_1(f) \cup \mathcal{P}_{f,2} \cup \mathcal{P}_{f,3} \cup \mathcal{P}_{f,4} \cup \mathcal{P}_{f,5})$. Désormais, il est clair que

$$\mathcal{M}_f \cup E^{taut}(\mathcal{M}_f) \cup \mathcal{P}_{f,2} \subseteq \mathcal{F}_k(f) \subseteq E^*\left(\mathcal{M}_f \cup \bigcup_{i=1,\ldots,5} \mathcal{P}_{f,i}\right),$$

où $E^{taut}(\mathcal{M}_f)$ correspond à une expansion de type "tautologie" des arbres minimaux de f. Dans la borne inférieure, les trois ensembles sont disjoints, donc nous en concluons :

$$\pi_k(\mathcal{M}_f) + \pi_k(E^{taut}(\mathcal{M}_f)) + \pi_k(\mathcal{P}_{f,2}) \leqslant \pi_k(f) \leqslant \pi_k(E^*(\mathcal{M}_f)) + \sum_{i=1}^{5} \pi_k(E^*(\mathcal{P}_{f,i})).$$

Ce qui nous donne, avec les Lemmes 26, 27 et 28 les bornes asymptotiques énoncées dans le Théorème 3. □

6.4 Probabilités des fonctions *read-once*

L'ensemble $\mathcal{R}_{c,k}$ est celui introduit dans la section précédente.

Proposition 9 *Soit c un entier fixé. La probabilité de l'ensemble des fonctions* read-once *de complexité c, lorsque k tend vers l'infini, vaut :*

$$\pi_k(\mathcal{R}_{c,k}) = \frac{1}{2^{2c-1}\ c}\binom{2c-2}{c-1} + O\left(\frac{1}{k}\right).$$

La fraction limite moyenne d'une fonction read-once *de complexité c parmi k variables, est égale à :*

$$\frac{\pi_k(\mathcal{R}_{c,k})}{|\mathcal{R}_{c,k}|} = \frac{c!}{2^{2c-1}\ c^c\ k^c} + O\left(\frac{1}{k^{c+1}}\right).$$

Preuve: La démonstration est analogue à celle de la Proposition 6 [p. 51]. □

A c fixé, nous remarquons que la probabilité des fonctions *read-once* de complexité c est strictement positive lorsque k tend vers l'infini. Par conséquent, l'effet Shannon n'est pas vérifié pour le modèle du processus de branchement. En effet, les fonction *read-once* de complexité fixée ayant une probabilité asymptotique strictement positive, il n'est pas possible que presque toutes les fonctions soient de complexité maximale.

Le tableau suivant 6.2 contient la probabilité des fonctions *read-once* de complexité inférieure ou égale à 10. En sommant ces valeurs, nous pouvons en conclure : *en choisissant une fonction aléatoirement selon le modèle de branchement (avec $k \to \infty$), avec une probabilité supérieure à 0.823, cette fonction est* read-once *de complexité inférieure ou égale à 10.*

Complexité	Probabilité
1	$1/2 = 0.5$
2	$1/8 = 0.125$
3	$1/16 = 0.0625$
4	$5/128 \approx 0.0390$
5	$7/256 \approx 0.0273$
6	$21/1024 \approx 0.0205$
7	$33/2048 \approx 0.0161$
8	$429/32768 \approx 0.0131$
9	$715/65536 \approx 0.0109$
10	$2431/262144 \approx 0.00927$

FIG. 6.2 – Probabilité des fonctions *read-once* de complexité inférieure à 10, lorsque $k \to \infty$.

Troisième partie

Logique classique vs logique intuitionniste

Chapitre 7

Logique intuitionniste

Dans un système logique donné, les tautologies (expressions qui calculent la fonction *Vrai*) sont souvent étudiées de prime abord pour différentes raisons. L'une d'elles est que la fonction *Vrai* est l'une des plus simples et l'on peut donc espérer que la structure des arbres la calculant n'est pas des plus compliquées. Une autre raison est le lien avec la logique dont la notion de démonstration correspond à l'idée de tautologie.

L'*intuitionnisme*, dont les fondements, en tant que branche de la logique formelle, ont été définis par Brouwer dans les année 30 consiste originellement en une position philosophique par rapport aux mathématiques. Certains raisonnements, jugés contre-intuitifs, sont rejetés dans cette logique. En effet, une règle du calcul propositionnel, définie en logique classique, n'est pas admise en logique intuitionniste : il s'agit du *tiers exclu* qui consiste à dire qu'étant donnée une proposition ϕ, soit on a *"ϕ est vrai"*, soit on a *"non(ϕ) est vrai"* ; formellement $\phi \vee \overline{\phi}$ est une tautologie. Le lecteur trouvera l'histoire et des motivations du développement de cette logique présentées par Van Dalen [vD86] et dans le chapitre 1 du livre de Troelstra et Van Dalen [TD88].

La logique intuitionniste se veut *constructive*. Ainsi on ne veut pas seulement connaître l'existence d'une solution à un problème, mais on veut aussi savoir la construire. Voilà un exemple de théorème : *il existe des nombres irrationnels a et b tels que a^b est un nombre rationnel.* Présentons-en une preuve : *si $\sqrt{2}^{\sqrt{2}}$ est irrationnel, alors $a = b = \sqrt{2}$ conviennent. Sinon, prenons $a = \sqrt{2}^{\sqrt{2}}$ et $b = \sqrt{2}$. Dès lors $a^b = \sqrt{2}^{\sqrt{2}^{\sqrt{2}}} = \sqrt{2}^2 = 2$.* Cette démonstration est rejetée par les intuitionnistes du fait que la solution n'est pas exhibée. Elle utilise le tiers exclu et par conséquent on ne connaît pas les valeurs de a et de b satisfaisant la propriété. Il existe une autre démonstration (nettement plus difficile) prouvant que $\sqrt{2}^{\sqrt{2}}$ est irrationnel, donc le problème peut être contourné. Toutefois, pour certains types de propositions, nous ne sommes pas en mesure actuellement de passer outre ce problème : le fait de savoir si une proposition est vraie ou fausse. Ainsi, le fait de savoir qu'elle est vraie ou que la proposition contraire est vraie n'apporte pas beaucoup d'information. En effet, soit la proposition : *Dans la suite des décimales de π, il existe un rang où les dix chiffres 0,1,2,3,4,5,6,7,8 et 9 apparaissent de manière consécutive et dans cet ordre.* Parmi les 200 premiers millions de décimales cette proposition est fausse, mais dans toute la suite des décimales, il est possible que jamais personne ne sache dire si cette proposition est vraie ou fausse.

Ainsi, la logique intuitionniste ne définit plus le concept d'*implication*, $a \rightarrow b$, comme formule équivalente à $\overline{a} \vee b$. En affirmant $a \rightarrow b$, on veut désormais un moyen de *construction*

d'une démonstration de b à partir d'une démonstration de a. Il est clair que les démonstrations en logique intuitionniste (i.e. les tautologies intuitionnistes) sont des démonstrations en logique classique (i.e. tautologies classiques). Le lecteur trouvera des informations sur la logique, et la logique intuitionniste en particulier, dans le livre de David *et al.* [DNR04]. Voilà quelques tautologies valides en logique classique mais non valide en logique intuitionniste.

$$\begin{aligned} a \vee \bar{a} &\quad : \quad \text{tiers exclu} \\ ((a \to b) \to a) \to a &\quad : \quad \text{formule de Peirce} \\ \bar{\bar{a}} \to a &\quad : \quad \text{double négation.} \end{aligned}$$

Les travaux récents, notamment la correspondance de Curry-Howard (voir le livre de Sørensen et Urzyczyn [SU98] par exemple), ont donné à la logique intuitionniste un statut central dans la logique et dans l'informatique, en en faisant historiquement la première des logiques constructives. C'est précisément la correspondance de Curry-Howard qui montre qu'une preuve constructive a toutes les propriétés d'un programme ; i.e. que les règles de démonstration permettent d'écrire des programmes corrects.

L'école polonaise, autour de Zaionc, a commencé à analyser de manière quantitative les différences entre logiques intuitionniste et classique. Ces chercheurs utilisent plutôt la terminologie *densité* ou *densité asymptotique* pour notre notion de *fraction limite*. Nous avons choisi d'utiliser cette appellation pour qu'il n'y ait aucun amalgame entre la notion mathématique de *fonction de densité* et leur notion de densité, issue de la physique.

En ce qui concerne la logique de l'implication avec une unique variable, et par la même occasion le *lambda-calcul simplement typé*, Statman [Sta80] a prouvé que toute tautologie classique est intuitionniste et a donné des bornes sur la fraction limite des tautologies. L'article de Moczurad *et al.* [MTZ00] donne la fraction limite exacte des tautologies dans ce système. Plus tard, Zaionc [Zai05] ajoute le connecteur *négation* au système précédent (qui le rend dès lors complet) et obtient à nouveau la fraction limite exacte des tautologies. Toujours pour le même langage, Kostrzycka et Zaionc [KZ04] ont donné la proportion exacte de la fraction limite des tautologies intuitionnistes par rapports aux tautologies classiques.

Par la suite, nous souhaitons discerner les tautologies intuitionnistes des tautologies non-intuitionnistes (habituellement appelées tautologies classiques). Pour ce faire, introduisons une caractérisation des tautologies intuitionnistes. Celle-ci nous permettra aisément de justifier si les tautologies à venir sont intuitionnistes ou non. Le lecteur trouvera des explications détaillées de ce qui suit dans le livre de Sørensen et Urzyczyn [SU98].

Logique classique. Pour commencer, intéressons-nous aux formules propositionnelles classiques. Jusque-là nous les avons liées aux deux valeurs de vérité 0 et 1 via l'algèbre Booléenne suivante :

Définition 2 *Soit* $\mathbb{B} =< \{0,1\}, max, min, \neg >$.

- *Une* évaluation *dans* $\mathbb{B}$ *est une application* $v : \{x_1, \ldots, x_k\} \longrightarrow \{0,1\}$ *;*
- *Etant donnée une évaluation* v*, son* interprétation *est l'application* $[[\cdot]]_v : \mathcal{F}_k \longrightarrow \{0,1\}$ *telle que :*
  1. $[[x_i]]_v = v(x_i)$*, pour tout* $i \in \{1, \ldots, k\}$ *;*
  2. $[[\bot]]_v = 0$*,* $\bot$ *correspond à la constante* Faux *;*
  3. $[[A \vee B]]_v = max\{[[A]]_v, [[B]]_v\}$*, pour* $A, B \in \mathcal{F}_k$ *;*

4. $[[A \wedge B]]_v = min\{[[A]]_v, [[B]]_v\}$, *pour* $A, B \in \mathcal{F}_k$;
5. $[[A \rightarrow B]]_v = max\{1 - [[A]]_v, [[B]]_v\}$, *pour* $A, B \in \mathcal{F}_k$.

– *Une formule* $A \in \mathcal{F}_k$ *est une tautologie (classique) si et seulement si* $[[A]]_v = 1$ *pour toute évaluation* v.

Cependant, nous aurions pu associer les formules propositionnelles classiques à l'algèbre Booléenne suivante. Il s'agit d'une sémantique alternative, liant les connecteurs logiques aux opérations de la théorie des ensembles.

Définition 3 *Soit* $\mathbb{B}' =< \mathcal{O}(\mathbb{R}), \cup, \cap, \neg, \emptyset, \mathbb{R} >$, *où* $\mathcal{O}(\mathbb{R})$ *consiste à l'ensemble des ouverts de* $\mathbb{R}$.

– *Une* évaluation *dans* $\mathbb{B}'$ *est une application* $v : \{x_1, \ldots, x_k\} \longrightarrow \mathcal{O}(\mathbb{R})$.
– *Etant donnée une évaluation* v, *son* interprétation *est l'application* $[[\cdot]]_v : \mathcal{F}_k \longrightarrow \mathcal{O}(\mathbb{R})$ *telle que :*
1. $[[x_i]]_v = v(x_i)$, *pour tout* $i \in \{1, \ldots, k\}$;
2. $[[\bot]]_v = \emptyset$;
3. $[[A \vee B]]_v = [[A]]_v \cup [[B]]_v$, *pour* $A, B \in \mathcal{F}_k$;
4. $[[A \wedge B]]_v = [[A]]_v \cap [[B]]_v$, *pour* $A, B \in \mathcal{F}_k$;
5. $[[A \rightarrow B]]_v = (\mathbb{R} \setminus [[A]]_v) \cup [[B]]_v$, *pour* $A, B \in \mathcal{F}_k$.

Les deux approches (selon l'algèbre Booléenne choisie) sont équivalentes et nous avons donc le résultat suivant :

Fait 1 *Pour l'algèbre* $\mathbb{B}'$, *une formule* $A \in \mathcal{F}_k$ *est une tautologie (classique) si et seulement si* $[[A]]_v = \mathbb{R}$ *pour toute évaluation* v.

Idées de preuve : La démonstration suit les idées générales suivantes :
– sens direct : démonstration par l'absurde.
– réciproque : une évaluation dans $\{0, 1\}$ peut être vue comme une évaluation dans $\mathcal{O}(\mathbb{R})$ qui n'affecte aux variables propositionnelles que $\mathbb{R}$ ou $\emptyset$.

□

Logique intuitionniste. Intéressons-nous désormais aux formules propositionnelles intuitionnistes. Une *algèbre de Heyting* a une structure algébrique satisfaisant les conditions suivantes :

Définition 4 $\mathbb{H}_0 =< H, \cup, \cap, \rightarrow, \neg, 0, 1 >$, *où*
– $\cup, \cap$ *sont associatives, commutatives et distributives l'une par rapport à l'autre ;*
– $a \cup 0 = a$ *et* $a \cap 1 = a$;
– $a \cup a = a$;
– $a \cap c \leqslant b$ *est équivalent à* $c \leqslant a \rightarrow b$ *(où* $a \leqslant b$ *signifie* $a \cup b = b$*) ;*
– $\neg a = a \rightarrow 0$.

Une évaluation v *est une fonction de l'ensemble des variables propositionnelles dans* $\mathbb{H}_0$. *Pour toute fonction d'évaluation* v, *nous définissons son* interprétation *:* $[[\cdot]]_v$ *de l'ensemble des formules dans* $\mathbb{H}_0$, *de la manière récursive suivante :*
– $[[x_i]]_v = v(x_i)$ *pour tout* $i \in \{1, \ldots, k\}$;
– $[[\bot]]_v = 0$;
– $[[A \wedge B]]_v = [[A]]_v \cap [[B]]_v$, *pour tout* $A, B \in \mathcal{F}_k$;

- $[[A \vee B]]_v = [[A]]_v \cup [[B]]_v$, *pour tout* $A, B \in \mathcal{F}_k$;
- $[[A \rightarrow B]]_v = [[A]]_v \rightarrow [[B]]_v$, *pour tout* $A, B \in \mathcal{F}_k$.

En particulier, toute algèbre Booléenne est une algèbre de Heyting où $a \rightarrow b$ est défini par $\neg a \cup b$.

Fait 2 *Une expression A est une tautologie intuitionniste si et seulement si $[[A]]_v = 1$ pour toute évaluation v et toute algèbre de Heyting.*

Idées de preuve : La démonstration suit les idées générales suivantes :
- sens direct : récurrence sur la taille des tautologies
- réciproque : démonstration par absurde. En supposant A non tautologie intuitionniste, on peut construire une évaluation v telle que $[[A]]_v \neq 1$.

□

Nous avons une première caractérisation des tautologies intuitionnistes, mais elle ne se manipule pas aisément. Etudions l'algèbre de Heyting suivante :

Définition 5 $\mathbb{H} =< \mathcal{O}(\mathbb{R}), \cup, \cap, \rightarrow, \neg, \emptyset, \mathbb{R} >$, *où*
- $\mathcal{O}(\mathbb{R})$ *est l'ensemble des ouverts de* $\mathbb{R}$;
- *les opérations* $\cup$ *et* $\cap$ *sont celles de la théorie des ensembles ;*
- $O_1 \rightarrow O_2 := Int((\mathbb{R} \setminus O_1) \cup O_2)$, *pour des ouverts* $O_1, O_2 \in \mathbb{R}$, *où* $Int(I)$ *correspond à l'intérieur de l'intervalle* I ;
- $\neg O = Int(\mathbb{R} \setminus O)$, *pour un ouvert* $O \in \mathbb{R}$.

Tout d'abord, remarquons que dans cette sémantique, pour O_1, O_2, deux ouverts de $\mathcal{O}(\mathbb{R})$, $\neg O_1 \cup O_2$ n'est pas égal à $O_1 \rightarrow O_2$.

Nous utiliserons la caractérisation suivante pour prouver désormais qu'une tautologie est intuitionniste :

Fait 3 *Une expression A est une tautologie intuitionniste si et seulement si $[[A]]_v = \mathbb{R}$ pour toute évaluation v sur $\mathbb{H}$.*

Preuve: La démonstration se trouve dans le livre de Rasiowa et Sikorski [RS63]. D'après le Fait 2, il suffit de montrer que si $[[A]]_v = \mathbb{R}$ pour toute évaluation v sur $\mathbb{H}$ alors A est une tautologie intuitionniste. En voilà les idées principales :
- toute formule de taille n est une tautologie intuitionniste si et seulement si elle est valide dans toutes les algèbres de Heyting de taille au plus 2^{2^n} ;
- soit $\mathcal{H}$ une algèbre de Heyting d'ouverts d'une espace V métrique dense. Toute algèbre finie peut être plongée dans l'ensemble des ouverts d'un sous-ensemble ouvert de V, et cette algèbre est homomorphe à $\mathcal{H}$;
- pour l'algèbre de Heyting $\mathcal{H}$, une formule ϕ est une tautologie intuitionniste si et seulement si pour toute évaluation v dans cette algèbre on à $[[\phi]]_v = 1$;
- l'algèbre $\mathbb{H}$ est une des algèbres de Heyting $\mathcal{H}$.

□

Possédant désormais un outil efficace pour déterminer la nature des tautologie, intéressons-nous au système de l'implication. *Que peut-on dire des tautologies intuitionnistes de ce système ?*

Chapitre 8

Système de l'implication

Le système que nous allons étudier dans ce chapitre est celui des arbres construits avec l'unique connecteur *implication* et l'ensemble des k littéraux positifs $\{x_1, \ldots, x_k\}$. Nous nous plaçons dans le modèle de probabilité des *grands arbres*. Soit Int_k l'ensemble des tautologies intuitionnistes, et Cl_k l'ensemble de toutes les tautologies. Nous conjecturons que l'ensemble des tautologies intuitionnistes admet une fraction limite. Après avoir exhibé certaines tautologies non intuitionnistes, nous allons déterminer les valeurs asymptotiques (quand k tend vers l'infini) de $\mu^-(Cl_k \setminus Int_k)$ et $\mu^+(Cl_k \setminus Int_k)$.

8.1 Tautologies simples

Nous avons défini dans le chapitre 5, l'ensemble des tautologies simples, qui correspondent aux expressions dont une des prémisses est égale au but de l'expression. Soit A une tautologie simple. Il existe des expressions $A_1, \ldots, A_p$, un entier $i \in \{1, \ldots, p\}$ et une variable $\alpha \in \{x_1, \ldots, x_k\}$ tels que α soit le but de A et que $A_i = \alpha$ (cf. Figure 8.1). Rappelons que nous nommons les sous-arbres A_j les prémisses de l'arbres.

Proposition 10 *Les tautologies simples sont des tautologies intuitionnistes.*

Preuve: Nous allons utiliser le Fait 3 [p. 72] introduit dans le chapitre précédent. Soit A une tautologie simple. Elle peut être représentée sous la forme de la Figure 8.1.

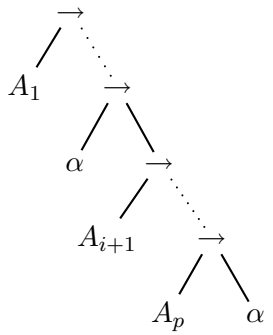

FIG. 8.1 – Une tautologie simple A.

Démontrons que A est une tautologie intuitionniste. Soit v une évaluation associant à α l'ouvert $O \in \mathcal{O}(\mathbb{R})$. Par récurrence sur la taille du sous-arbre $B = A_{i+1}, \ldots, A_p \to \alpha$, nous

avons $O \subseteq [[B]]_v = O'$. Par conséquent $[[\alpha \to B]]_v = Int((\mathbb{R} \setminus O) \cup O') = \mathbb{R}$. Une récurrence sur la somme des tailles des premières prémisses de A, nous permet de conclure que $[[A]]_v = \mathbb{R}$. Ainsi, A est une tautologie intuitionniste. □

Par ailleurs, nous avons prouvé Théorème 2 [p. 24] que lorsque le nombre de variables devient grand, presque toutes les tautologies sont simples. Nous en concluons le corollaire suivant :

Corollaire 4 *Asymptotiquement (lorsque le nombre de variables k devient grand), les tautologies classiques sont intuitionnistes, i.e.*

$$\lim_{k\to\infty} \frac{\mu_k^-(Int_k)}{\mu_k(Cl_k)} = 1.$$

Preuve: Du fait que $G_k \subset Int_k \subset Cl_k$, nous avons

$$\mu_k(G_k) = \lim_{n\to\infty} \frac{|G_k^n|}{|\mathcal{F}_k^n|} \leqslant \liminf_{n\to\infty} \frac{|Int_k^n|}{|\mathcal{F}_k^n|} \leqslant \limsup_{n\to\infty} \frac{|Int_k^n|}{|\mathcal{F}_k^n|} \leqslant \lim_{n\to\infty} \frac{|Cl_k^n|}{|\mathcal{F}_k^n|} = \mu_k(Cl_k).$$

Le résultat est obtenu du fait que les deux fractions limites $\mu_k(G_k)$ et $\mu_k(Cl_k)$ sont égales à $1/k + O(1/k^2)$. □

D'après les fractions limites des tautologies simples, des non-tautologies simples et des non-tautologies moins simples (familles définies dans la Section 5.1 [p. 25]), nous avons

$$\begin{array}{rcccl} \mu_k(G_k) & \leqslant & \mu_k(Cl_k) & \leqslant & 1 - \mu_k(SN_k) - \mu_k(LN_k) \\ \dfrac{4k+1}{(2k+1)^2} & \leqslant & \mu_k(Cl_k) & \leqslant & 1\dfrac{k(k-1)}{(k+1)^2} - \dfrac{2k(k-1)^2}{k+2)^4} \\ \dfrac{1}{k} - \dfrac{3}{4k^2} - O\left(\dfrac{1}{k^3}\right) & \leqslant & \mu_k(Cl_k) & \leqslant & \dfrac{1}{k} + \dfrac{15}{k^2} - O\left(\dfrac{1}{k^3}\right) \end{array}$$

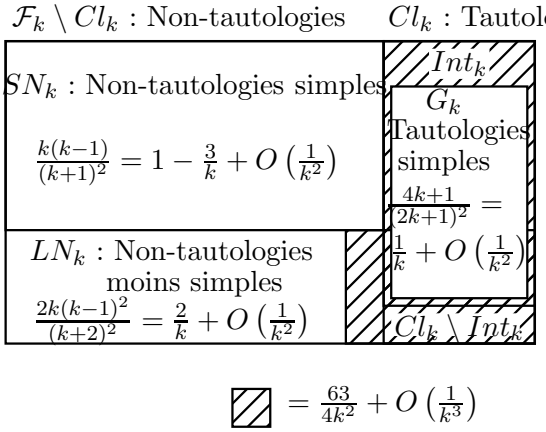

FIG. 8.2 – Fractions limites des tautologies simples, des non-tautologies simples et des non-tautologies moins simples.

Ce corollaire permet de donner une réponse positive au cas asymptotique de la conjecture de l'article de Moczurad *et al.* [MTZ00, page 593]. Par ailleurs, en étudiant le second terme

des développements de $\mu_k(G_k)$ et $\mu_k(Cl_k)$ nous pouvons également établir la valeur de la fraction limite (si elle existe) des tautologies non intuitionnistes. La suite du chapitre infirme la conjecture du même article [MTZ00, page 593] lorsque k est fixé.

8.2 Tautologies non intuitionnistes

Pour le moment, les tautologies que nous avons présentées étaient toutes intuitionnistes, et les calculs présentés ne prouvaient pas qu'il pouurait y avoir une différence, dans notre système de l'implication, entre les logiques intuitionniste et classique.

Dans le système de l'implication avec une unique variable, les tautologies classiques sont intuitionnistes. Statman [Sta80] en a donné en 1980 une démonstration par récurrence basée sur la taille de la tautologie. Ainsi, dans le système à une seule variable et un seul connecteur (l'implication), les deux logiques coïncident. A partir du moment où le système contient au moins deux variables, ce n'est plus vrai ; voici le genre d'expression que l'on peut construire :

$$((\alpha \to \beta) \to \alpha) \to \alpha.$$

Le lecteur aura sans doute reconnu la *loi de Peirce*.

Proposition 11 *La loi de Peirce est une tautologie non intuitionniste.*

Preuve: Une rapide vérification (via la table de vérité de α et β) nous laisse conclure que cette expression est une tautologie classique. Démontrons, en utilisant le Fait 3 [p. 72] que ce n'est pas une tautologie intuitionniste. Soit l'évaluation v qui à α associe $\mathbb{R} \setminus \{0\}$ et à β associe l'ensemble vide $\emptyset$. Nous avons $[[\alpha \to \beta]]_v = Int(\{0\} \cup \emptyset) = \emptyset$. Par conséquent $[[(\alpha \to \beta) \to \alpha]]_v = \mathbb{R}$. Ce qui nous permet de conclure que $[[((\alpha \to \beta) \to \alpha) \to \alpha]]_v = Int(\emptyset \cup \mathbb{R} \setminus \{0\}) = \mathbb{R} \setminus \{0\}$. Nous avons une évaluation v qui à la tautologie A n'associe pas $\mathbb{R}$, par conséquent A n'est pas une tautologie intuitionniste. □

Que peut-on dire de la fraction limite des tautologies non intuitionnistes ?

8.3 Fraction limite des expressions de Peirce

Nous avons prouvé que la plupart des tautologies sont simples par conséquent, elles sont intuitionnistes. Considérons l'ensemble des tautologies, auquel on soustrait les tautologies simples. Nous allons démontrer que dans cet ensemble, il y a une différence (lorsque k tend vers l'infini) entre les logiques intuitionniste et classique.

Soit $Peirce_k = Cl_k \setminus Int_k$. Les éléments de $Peirce_k$ seront appelés expressions de Peirce. Puisque nous ne savons pas si $\mu_k(Int_k)$ existe, il en est de même pour $\mu_k(Peirce_k)$. Nous allons donc nous intéresser à $\mu_k^+(Peirce_k)$ et $\mu_k^-(Peirce_k)$. D'après le corollaire 4 [p. 74], nous avons $\mu_k^+(Peirce_k) = o(\mu_k^-(Int_k))$. Par conséquent, $\mu_k^+(Peirce_k) = o(1/k)$.

Afin de déterminer précisément les termes prédominants de $\mu_k^+(Peirce_k)$ et $\mu_k^-(Peirce_k)$, quand $k \to \infty$, nous allons devoir introduire de nouvelles familles de tautologies qui ne sont pas des tautologies simples. Rappelons que ce dernier ensemble contient cependant la plus grande partie des tautologies puisque, asymptotiquement, le terme principal de sa fraction limite est égal à celui de l'ensemble des tautologies classiques.

L'étude qui va suivre, va permettre de donner le second terme exact de la fraction limite des tautologies classiques.

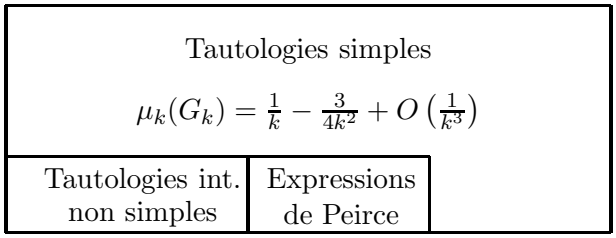

FIG. 8.3 – Ensemble des tautologies classiques

Afin de simplifier les calculs de fractions limites, nous allons introduire un lemme utilisant les définitions suivantes. Un *squelette* est un arbre unaire-binaire, fini, planaire et enraciné tel que :

- les feuilles sont étiquetées par les éléments d'un ensemble dénombrable de *variables de squelette* (symbolisée par des lettres grecques) ;
- les arêtes sont soit étiquetées par R (nous les appellerons *arêtes régulières*), soit par S (*arêtes de squelette*).

Par ailleurs, les arêtes joignant un fils gauche sont toutes des arêtes régulières et les arêtes provenant d'un noeud unaire sont toutes des arêtes de squelette.

On ne distingue pas les squelettes pouvant être transformés en un autre squelette via un renommage des variables de squelette. Dans les figures, les arêtes régulières sont représentées en traits continus et les arêtes de squelette en traits discontinus. En outre, en général on donne un nom aux arêtes de squelette – la Figure 8.4 en présente un exemple.

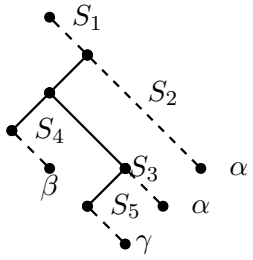

FIG. 8.4 – Exemple de squelette.

Pour un squelette T, soient $s(T)$ et $d(T)$ respectivement le nombre d'arêtes de squelette dans T et le nombre d'étiquettes différentes dans les feuilles de T. Le *nombre de répétitions* $rep(T)$ est la différence entre le nombre total de feuilles dans T et $d(T)$. Ainsi, l'expression $((\alpha \to \beta) \to \alpha) \to \alpha$ a deux répétitions. Nous écrirons s, d, rep au lieu de $s(T), d(T), rep(T)$ lorsque le squelette T sera clair via le contexte. Comme d'habitude, la taille de T est son nombre de feuilles, et est notée $|T|$.

Pour un squelette T, une *substitution* est un élément de $((\mathcal{F}_k)^*)^{s(T)} \times \{x_1, \ldots, x_k\}^{d(T)}$. Soit T un squelette, $\sigma = ((s_1, \ldots, s_a), (v_1, \ldots v_b))$ étant une substitution pour T, soient $(e_1, \ldots, e_a)$ et $(l_1, \ldots, l_b)$ les listes d'arêtes de squelette de T et celle des variables apparaissant dans T, les deux listes étant obtenues via un parcours quelconque de l'arbre – prenons le parcours en profondeur préfixé par exemple. L'application de la substitution σ au squelette T s'effectue de la manière suivante :

– chaque feuille étiquetée par l_i est réétiquetée par v_i ;
– chaque arête de squelette e_i est étendue localement par la suite $s_i = (t_1, \ldots, t_{k_i})$, comme représenté sur la Figure 8.5 :

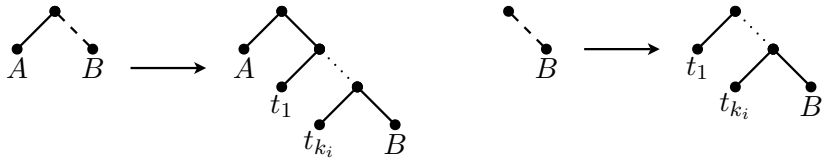

FIG. 8.5 – Substitution d'une arête de squelette.

Le traitement de la substitution est clair. Si la suite (t_i) est vide, lorsque le père est un noeud binaire, alors l'arête de squelette devient une arête régulière. Lorsque le père est un noeud unaire et que la suite est vide, le père est remplacé par le noeud fils. Bien entendu, tous les noeuds internes sont étiquetés par le connecteur $\rightarrow$. Finalement, le résultat d'une substitution de squelette est une expression Booléenne – voir les arbres obtenus sur la Figure 8.5.

On dira que la *substitution* $((s_1, \ldots, s_a), (v_1, \ldots v_b))$ *est propre* lorsque pour tout $1 \leqslant i < j \leqslant b$ on a $v_i \neq v_j$ et pour tout $i \in \{1, \ldots, a\}$ le but $r(s_i) \notin \{v_1, \ldots, v_b\}$.

Enfin, la famille d'arbres notée $\mathcal{F}_k^T$ est constituée des arbres obtenus à partir de T via une substitution propre. Soit $\widehat{\mathcal{F}_k^T}$ la famille d'expressions construites via T avec n'importe quelle substitution (pas forcément propre). Pour un ensemble $\mathcal{A}_k \subset \mathcal{F}_k$, s'il existe un squelette T tel que $\mathcal{F}_k^T \subset \mathcal{A}_k \subset \widehat{\mathcal{F}_k^T}$, alors nous dirons que la famille $\mathcal{A}_k$ *correspond* au squelette T.

Lemme 29 *Pour un squelette T, la fraction limite $\mu_k(\mathcal{F}_k^T)$ existe et vérifie :*

$$\mu_k(\mathcal{F}_k^T) = \frac{s(T)}{2^{2rep(T)+2d(T)-s(T)-1}\ k^{rep(T)}} + O\left(\frac{1}{k^{rep(T)+1}}\right).$$

Preuve: Soit T un squelette. Nous avons une bijection entre les substitutions propres pour T et les éléments de $\mathcal{F}_k^T$. Définissons la taille d'une substitution $((s_1, \ldots, s_{s(T)}), (v_1, \ldots v_{d(T)}))$ comme la somme de la taille de tous les arbres des suites $s_1, \ldots, s_{s(T)}$. On remarque que la taille d'une expression obtenue par une substitution est égale à la taille de la substitution augmentée de celle du squelette T. La fonction génératrice énumérant les arbres dont les buts sont distincts de $v_1, \ldots v_b$ est donnée par $b_k^{d(T)}(z) = \frac{k-d(T)}{k} f_k(z)$. En conséquence, la fonction génératrice énumérant des suites de tels arbres vaut $1/(1 - b_k^{d(T)}(z))$. Du fait que nous avons besoin de $s(T)$ telles suites et que la substitution est propre, la fonction génératrice énumérant $\mathcal{F}_k^T$ est

$$\left(1 - \frac{k-d(T)}{k} f_k(z)\right)^{-s(T)} \cdot k^{\underline{d(T)}} \cdot z^{|T|},$$

où $k^{\underline{d(T)}} = k(k-1)\ldots(k-d(T)+1)$. Pour $s(T) > 0$ et $d(T) < k$ (les autres cas sont évidents), il est clair que cette fonction génératrice a sa singularité dominante au même point que $f_k(z)$. Par conséquent :

$$\mu_k(\mathcal{F}_k^T) = \frac{k^{\underline{d(T)}} \cdot s(T) \cdot \left(1 - \frac{k-d(T)}{2k}\right)^{-s(T)-1} \cdot \frac{k-d(T)}{k}}{(4k)^{|T|}} = \frac{k^{d(T)} \cdot s(T) \cdot 2^{s(T)+1}}{(4k)^{|T|}} + O\left(\frac{k^{d(T)-1}}{k^{|T|}}\right).$$

Ceci, accompagné du fait que $|T| = d(T) + rep(T)$, conclut la preuve du lemme. □

Lemme 30 *Pour un squelette T, nous avons*

$$\mu_k^+(\widehat{\mathcal{F}_k^T}) = \mu_k(\mathcal{F}_k^T) + O\left(\frac{1}{k^{rep(T)+1}}\right).$$

Preuve: Le nombre d'éléments de $\widehat{\mathcal{F}_k^T}$ de taille n, est inférieure au nombre de substitutions (propres ou non) pour T de taille $n-|T|$. La fonction génératrice énumérant ces substitutions vaut $(1-f_k(z))^{-s(T)} \cdot k^{d(T)}$. Ainsi, le nombre d'expressions de taille n de $\widehat{\mathcal{F}_k^T}$ est inférieur ou égal à $[z^n]h(z)$, où $h(z)$ vérifie :

$$h_k(z) = (1-f_k(z))^{-s(T)} \cdot k^{d(T)} \cdot z^{|T|}.$$

Nous en concluons donc :

$$\mu_k^+(\widehat{\mathcal{F}_k^T}) \leqslant \lim_{n\to\infty} \frac{[z^n]h_k(z)}{[z^n]f_k(z)} = \frac{s(T)}{2^{2rep(T)+2d(T)-s(T)-1} \cdot k^{rep(T)}} + O\left(\frac{1}{k^{rep(T)+1}}\right).$$

Cette équation et le Lemme 29 concluent la preuve. □

La suite de ce chapitre est destinée à exhiber la plupart des tautologies qui ne sont pas simples (lorsque k tend vers l'infini). Cela revient à obtenir le coefficient du terme $1/k^2$ de la fraction limite des tautologies. En outre, nous allons exhiber les expressions de Peirce parmi ces tautologies. L'étude est ordonnée selon le nombre de prémisses ayant le même but que le but global des tautologies.

8.3.1 Familles de tautologies

Nous allons partitionner l'ensemble des tautologies selon le nombre de prémisses spécifiques qu'elles contiennent. Nous en donnerons en même temps la fraction limite.

Soit $\mathcal{S}^0$ l'ensemble des expressions dont les buts des prémisses sont tous différents du but global. Il s'agit des non-tautologies simples. Dans le chapitre 5, nous avons démontré que cet ensemble ne contient aucune tautologie.

Soit $\mathcal{G}^{1;0}$ l'ensemble des expressions possédant une unique prémisse dont le but est le même que le but global et telles que cette prémisse soit réduite à une feuille. Il s'agit des tautologies simples dont la fraction limite est égale à :

$$\mu_k(\mathcal{G}^{1;0}) = \frac{1}{k} - \frac{3}{4k^2} + O\left(\frac{1}{k^3}\right).$$

Soit $\mathcal{S}^{1;>1}$ l'ensemble des expressions telles que chaque arbre $A \in \mathcal{S}^{1;>1}$ vérifie les conditions suivantes :

- A n'est pas une tautologie simple, i.e. A ne possède pas de prémisse réduite à la feuille étiquetée par $r(A)$,
- A a exactement une prémisse B telle que $r(B) = r(A)$ et B a au moins deux prémisses.

Le squelette de $\mathcal{S}^{1;>1}$ est représenté sur la Figure 8.6. Soit $\mathcal{G}^{1;>1}$ l'ensemble des tautologies de $\mathcal{S}^{1;>1}$.

Lemme 31

$$\mu_k^+(\mathcal{G}^{1;>1}) = O\left(\frac{1}{k^3}\right).$$

Preuve: Commençons par calculer la fraction limite de $\mathcal{S}^{1;>1}$. Soit $\phi_{\mathcal{S}^{1;>1}}(z)$ la fonction génératrice énumérant les arbres de $\mathcal{S}^{1;>1}$.

$$\phi_{\mathcal{S}^{1;>1}}(z) = k\frac{1}{1-f_k(z)+z}\left(\frac{f_k(z)^2}{1-f_k(z)}z\right)\frac{1}{1-f_k(z)+z}z.$$

Après calculs nous obtenons :

$$\mu_k(\mathcal{S}^{1;>1}) = \frac{5}{4k} - \frac{3}{2k^2} + O\left(\frac{1}{k^3}\right).$$

En utilisant le squelette de la Figure 8.6, nous pouvons définir 4 sous-familles disjointes de non-tautologies, en imposant des conditions sur les substitutions permises. Pour une substitution $((S_1, S_2, S_3, S_4, S_5), (\alpha, \beta, \gamma))$, nous considérons les cas suivants (nous utilisons abusivement les mêmes noms pour les variables (resp. arêtes) de squelette et les variables (resp. suites) obtenue après substitution) :

(a) $\beta = \alpha$, $\gamma \neq \alpha$, $\alpha, \gamma \notin r(S_1) \cup \ldots \cup r(S_5)$;
(b) $\beta \neq \alpha$, $\alpha, \beta \notin r(S_1) \cup r(S_2) \cup r(S_4)$ (aucune restriction sur γ) ;
(c) $\beta \neq \alpha$, β apparaît exactement une fois parmi les buts des arbres de S_1, S_2, S_4, $\beta \notin r(S_3) \cup r(S_5)$, $\alpha, \gamma \notin r(S_1) \cup \ldots \cup r(S_5)$;
(d) $\beta \neq \alpha$, α apparaît exactement une fois parmi les buts des arbres de S_4 ; $\alpha \notin r(S_1) \cup r(S_2) \cup r(S_3) \cup r(S_5)$, $\gamma \notin r(S_1) \cup \ldots \cup r(S_5)$.

Nous remarquons que seule la condition (a) engendre des substitutions propres pour le squelette de la Figure 8.6.

Nous appelons $\mathcal{S}_a^{1;>1}, \mathcal{S}_b^{1;>1}, \mathcal{S}_c^{1;>1}, \mathcal{S}_d^{1;>1}$ les quatre sous-familles de $\mathcal{S}^{1;>1}$ correspondant aux restrictions (a), (b), (c) et (d). Chacune de ces familles ne contient que des non-tautologies.

En effet, les expressions de $\mathcal{S}_a^{1;>1}, \mathcal{S}_c^{1;>1}, \mathcal{S}_d^{1;>1}$ s'évaluent à 0 lorsque $\alpha = \gamma = 0$ et que les autres variables sont évaluées à 1.

Pour la famille $\mathcal{S}_b^{1;>1}$, il suffit d'évaluer α et β à 0 et les autres variables à 1.

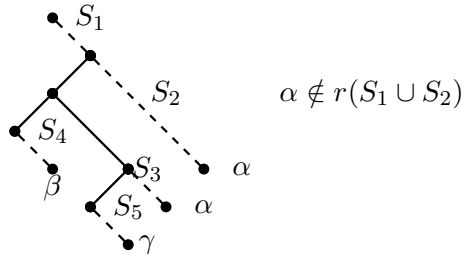

FIG. 8.6 – Squelette de la famille $\mathcal{S}^{1;>1}$.

La famille $\mathcal{S}_a^{1;>1}$ est construite à partir du squelette T de la figure 8.6 (en remplaçant β par α). L'application du lemme 29 avec les paramètres $s = 5$, $d = 2$ et $rep = 2$ nous donne :

$$\mu_k(\mathcal{S}_a^{1;>1}) \geqslant \frac{5}{4k^2} - O\left(\frac{1}{k^3}\right).$$

Afin de déterminer la fraction limite de $\mathcal{S}_b^{1;>1}$, nous avons besoin d'une estimation plus précise que celle fournie par la Lemme 29. Soit $\phi_{\mathcal{S}_b^{1;>1}}$ la fonction génératrice énumérant les arbres de $\mathcal{S}_b^{1;>1}$, elle vérifie :

$$\phi_{\mathcal{S}_b^{1;>1}} = k(k-1)\frac{1}{1-\frac{k-2}{k}f_k(z)}\left(\frac{z}{1-\frac{k-2}{k}f_k(z)}\frac{f_k(z)}{1-f_k(z)}z\right)\frac{1}{1-\frac{k-2}{k}f_k(z)}z.$$

Le calcul de la fraction limite de cette famille donne :

$$\mu_k(\mathcal{S}_b^{1;>1}) = \frac{5}{4k} - \frac{47}{4k^2} + O\left(\frac{1}{k^3}\right).$$

La famille $\mathcal{S}_c^{1;>1}$ est l'union disjoints des trois familles obtenues via les squelettes T_{c1}, T_{c2} et T_{c3}, décrits sur la Figure 8.7. Il est clair que

$$\left(\mathcal{F}_k^{T_{c1}} \cup \mathcal{F}_k^{T_{c2}} \cup \mathcal{F}_k^{T_{c3}}\right) \subset \mathcal{S}_b^{1;>1} \subset \left(\widehat{\mathcal{F}_k^{T_{c1}}} \cup \widehat{\mathcal{F}_k^{T_{c2}}} \cup \widehat{\mathcal{F}_k^{T_{c3}}}\right).$$

Pour chaque squelette, nous avons les paramètres $s = 7, d = 3$ et $rep = 2$. Les Lemmes 29 et 30 prouvent le résultat suivant :

$$\mu_k(\mathcal{S}_c^{1;>1}) = 3\frac{7}{4k^2} + O\left(\frac{1}{k^3}\right).$$

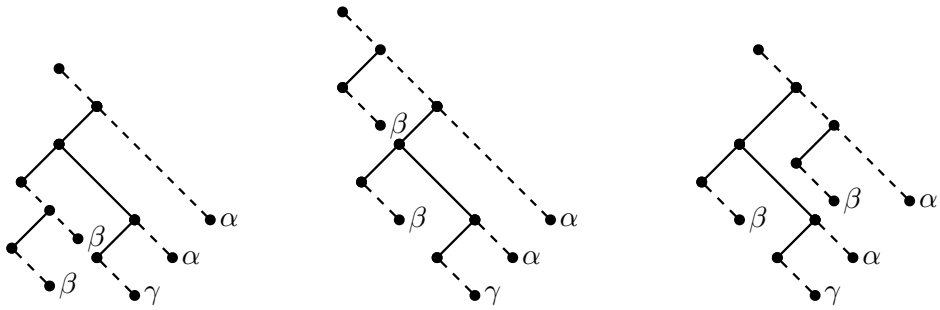

FIG. 8.7 – Squelettes permettant d'obtenir $\mathcal{S}_c^{1;>1}$.

Enfin, pour le dernier cas $\mathcal{S}_d^{1;>1}$, nous définissons les squelettes T_d sur la Figure 8.8. D'après les Lemmes 29 et 30, nous avons

$$\mu_k(\mathcal{S}_d^{1;>1}) = \frac{7}{4k^2} + O\left(\frac{1}{k^3}\right).$$

Finalement, en comparant les fractions limites de $\mathcal{S}^{1;>1}$ et $\mathcal{S}_a^{1;>1} \cup \mathcal{S}_b^{1;>1} \cup \mathcal{S}_c^{1;>1} \cup \mathcal{S}_d^{1;>1}$, nous en concluons que

$$\mu_k(\mathcal{G}^{1;>1}) = O\left(\frac{1}{k^3}\right).$$

□

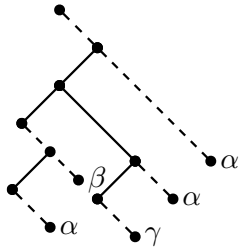

FIG. 8.8 – Squelette permettant d'obtenir $\mathcal{S}_d^{1;>1}$.

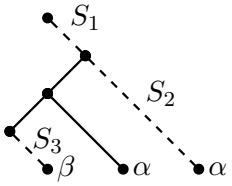

FIG. 8.9 – Squelette permettant d'obtenir $\mathcal{S}^{1;1}$.

Dès lors, nous allons considérer les expressions A ayant exactement une prémisse B vérifiant les deux conditions suivantes : $r(B) = r(A)$, et B a exactement une prémisse. Soit $\mathcal{S}^{1;1}$ l'ensemble de ces expressions. Sur la figure 8.9 nous en avons représenté un squelette permettant de générer cette famille – en considérant toutes les substitutions, y compris les non propres. A nouveau, soit $\mathcal{G}^{1;1}$ l'ensemble des tautologies de $\mathcal{S}^{1;1}$.

Lemme 32

$$\mu_k(\mathcal{G}^{1;1}) = \frac{2}{k^2} + O\left(\frac{1}{k^3}\right),$$

$$\mu_k(\mathcal{G}^{1;1} \cap Int_k) = \frac{3}{2k^2} + O\left(\frac{1}{k^3}\right).$$

Preuve: Déterminons tout d'abord la fraction limite de l'ensemble des expressions de $\mathcal{S}^{1;1}$. Soit $\phi_{\mathcal{S}^{1;1}}(z)$ la fonction génératrice énumérant ces arbres.

$$\phi_{\mathcal{S}^{1;1}}(z) = k\frac{1}{1-\frac{k-1}{k}f_k(z)}(f_k(z)z)\frac{1}{1-\frac{k-1}{k}f_k(z)}z.$$

Après calculs nous obtenons :

$$\mu_k(\mathcal{S}^{1;1}) = \frac{3}{4k} - \frac{5}{2k^2} + O\left(\frac{1}{k^3}\right).$$

Comme précédemment, nous allons définir quatre conditions sur les substitutions (propres ou non) du squelette et nous obtiendrons ainsi quatre nouvelles familles $\mathcal{S}_a^{1;1}$ à $\mathcal{S}_d^{1;1}$.

(a) $\beta = \alpha$, $\alpha \notin r(S_1) \cup r(S_2) \cup r(S_3)$,

(b) $\beta \neq \alpha$, $\alpha, \beta \notin r(S_1) \cup r(S_2) \cup r(S_3)$,

(c) $\beta \neq \alpha$, $\beta \notin r(S_1) \cup r(S_2) \cup r(S_3)$, α apparaît exactement une fois parmi les buts de S_3, $\alpha \notin r(S_1) \cup r(S_2)$,

(d) $\beta \neq \alpha$, $\alpha \notin r(S_1) \cup r(S_2) \cup r(S_3)$ et β apparaît exactement une fois parmi les buts de S_1, S_2 et S_3.

Nous remarquons que seules les conditions (a) et (b) engendrent des substitutions propres pour le squelette de la Figure 8.9. De ce fait, d'après les Lemmes 29 et 30, nous avons

$$\mu_k(\mathcal{S}_a^{1;1}) = \frac{3}{4k^2} + O\left(\frac{1}{k^3}\right).$$

Afin d'obtenir une fraction limite de $\mathcal{S}_b^{1;1}$ suffisamment précise, nous ne pouvons pas utiliser directement les Lemmes 29 et 30. Déterminons donc la fonction génératrice $\phi_{\mathcal{S}_b^{1;1}}(z)$ énumérant $\mathcal{S}_b^{1;1}$.

$$\phi_{\mathcal{S}_b^{1;1}}(z) = k(k-1)\frac{1}{1-\frac{k-2}{k}f_k(z)}\left(\frac{1}{1-\frac{k-2}{k}f_k(z)}z\right)\frac{1}{1-\frac{k-2}{k}f_k(z)}z.$$

Après calculs nous obtenons :

$$\mu_k(\mathcal{S}_b^{1;1}) = \frac{3}{4k} - \frac{33}{4k^2} + O\left(\frac{1}{k^3}\right).$$

En évaluant α à 0 (resp. α et β à 0) nous remarquons que toutes les expressions de $\mathcal{S}_a^{1;1}$ (resp. $\mathcal{S}_b^{1;1}$) s'évaluent à 0. Donc ces deux ensembles ne contiennent aucune tautologie.

Intéressons-nous désormais aux conditions (c). La Figure 8.10 présente deux squelettes différents suivant le fait que la sous-sous-prémisse est réduite à α, construisant la famille P_k, ou est un arbre de but α mais non réduit à une feuille (famille NT_c). Remarquons tout d'abord

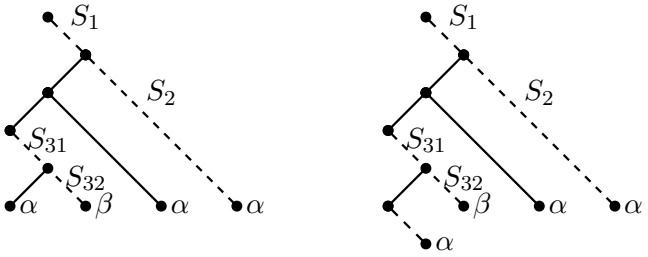

FIG. 8.10 – Squelettes permettant d'obtenir $\mathcal{S}_c^{1;1}$.

l'inclusion de P_k dans $Peirce_k$. Il suffit de remarquer que ces arbres s'évaluent à 1 quelle que soit la valeur de α pour conclure que ces arbres sont des tautologies classiques. Vérifions qu'il ne s'agit pas de tautologies intuitionnistes. Soit v l'évaluation qui associe l'ouvert $\mathbb{R} \setminus \{0\}$ à α, l'ensemble vide à β et $\mathbb{R}$ à toutes les autres variables. Pour chacune de ces tautologies v associe $\mathbb{R} \setminus \{0\}$. Par conséquent, $P_k \in Peirce_k$. Les Lemmes 29 et 30 nous donne la fraction limite de P_k :

$$\mu_k(P_k) = \frac{1}{2k^2} + O\left(\frac{1}{k^3}\right).$$

Afin de s'apercevoir que les expressions de NT_c ne sont pas des tautologies, il suffit d'évaluer à 0 α et β et à 1 les autres variables. Par ailleurs, les Lemmes 29 et 30 concluent :

$$\mu_k(NT_c) = \frac{3}{4k^2} + O\left(\frac{1}{k^3}\right).$$

Enfin, présentons six squelettes différents pour les expressions vérifiant les condition (d) (Figures 8.11 et 8.12). Ceux contenant la prémisse réduite à β, construisent la famille H_k, et ceux construisant les expressions avec un sous-arbre de but β mais non réduit à une feuille (famille NT_d).

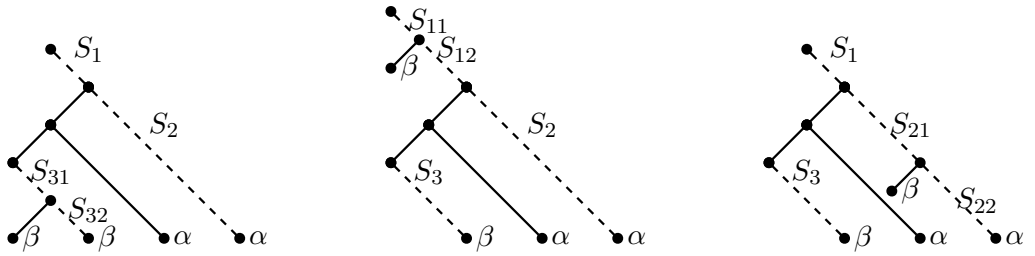

FIG. 8.11 – Squelettes de la famille H_k.

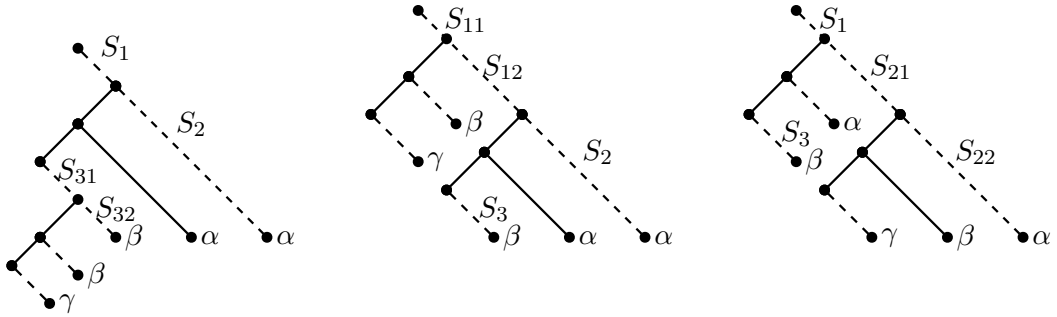

FIG. 8.12 – Squelettes de la famille NT_d.

Démontrons tout d'abord que les arbres de H_k sont des tautologies intuitionnistes. Soit A un arbre obtenu avec le premier squelette. L'arbre A possède une unique prémisse B telle que $r(B) = r(A)$. Par ailleurs cette prémisse B comporte une unique prémisse C qui est une tautologie simple et donc intuitionniste. Donc quelle que soit l'évaluation v, on a $[[C]]_v = \mathbb{R}$. Donc $[[B]]_v = [[r(B)]]_v = [[r(A)]]_v$. Nous en concluons que $[[B \to r(A)]]_v = \mathbb{R}$ et donc quelles que soient les autres prémisses de A, nous obtenons $[[A]]_v = \mathbb{R}$. Ainsi A est une tautologie intuitionniste. En ce qui concerne les expressions obtenues via les deux autres squelettes, remarquons tout d'abord que pour toute évaluation v, nous avons $[[\phi_1 \to (\phi_2 \to \phi_3)]]_v = [[\phi_2 \to (\phi_1 \to \phi_3)]]_v$ et il nous suffit donc de démontrer que les expressions obtenues via le troisième squelette sont des tautologies intuitionnistes, afin de conclure que $H_k \subset Int_k$. Soit A une expression obtenue avec le troisième squelette. Soit $B \to C$ la sous-expression de A telle que $r(B) = r(A)$. Notons B_1 la première prémisse de B. Quelle que soit l'évaluation v, nous avons $[[C]]_v \supset \text{Int}(\mathbb{R} \setminus [[r(B_1)]]_v \cup [[r(A)]]_v$. Par ailleurs $[[B]]_v \subset \text{Int}(\mathbb{R} \setminus [[r(B_1)]]_v \cup [[r(A)]]_v$. Par conséquent $[[B \to C]]_v = \mathbb{R}$. Or l'interprétation de l'implication est croissante, donc $[[A]]_v = \mathbb{R}$. En conclusion, les formules de H_k sont toutes des tautologies intuitionnistes. Une énumération des expressions construites à l'aide de ces six squelettes en utilisant les Lemmes 29 et 30 nous permet de prouver :

$$\mu_k(H_k) = 3\frac{1}{2k^2} + O\left(\frac{1}{k^3}\right),$$

$$\mu_k(NT_d) = 3\frac{3}{4k^2} + O\left(\frac{1}{k^3}\right).$$

Finalement, nous remarquons que

$$\mu_k(\mathcal{S}^{1;1}) = \mu_k(NT_a) + \mu_k(NT_b) + \mu_k(P_k) + \mu_k(NT_c) + \mu_k(H_k) + \mu_k(NT_d) + O\left(\frac{1}{k^3}\right).$$

En conséquence nous concluons :

$$\mu_k(\mathcal{G}^{1;1}) = \mu_k(P_k) + \mu_k(H_k) + O\left(\frac{1}{k^3}\right) = \frac{2}{k^2} + O\left(\frac{1}{k^3}\right),$$

$$\mu_k(\mathcal{G}^{1;1} \cap Int_k) = \mu_k(H_k) + O\left(\frac{1}{k^3}\right) = \frac{3}{2k^2} + O\left(\frac{1}{k^3}\right).$$

□

Soit $\mathcal{S}^2$ l'ensemble des expressions qui ne sont pas des tautologies simples et dont au moins deux prémisses ont le même but que le but global (voir Fig. 8.13). En ne substituant le squelette que par des substitutions propres, on n'obtient aucune tautologie (il suffit d'évaluer α à 0 et les autres variables à 1). Soit $\mathcal{G}^2$ les tautologies de $\mathcal{S}^2$. Ainsi on a

$$\mu_k(\mathcal{G}^2) = o(\mu_k(\mathcal{S}^2)).$$

D'après les Lemmes 29 et 30, on en conclut :

$$\mu_k(\mathcal{G}^2) = O\left(\frac{1}{k^3}\right).$$

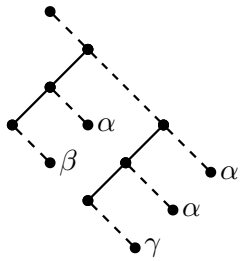

FIG. 8.13 – Squelette de la famille $\mathcal{S}^2$.

Soit $\mathcal{S}^{>2}$ l'ensemble des expressions qui ne sont pas des tautologies simples et dont au moins trois prémisses ont le même but que le but global. Soit $\mathcal{G}^{>2}$ les tautologies de $\mathcal{S}^{>2}$. D'après le Lemme 30,

$$\mu_k^+(\mathcal{S}^{>2}) = O\left(\frac{1}{k^3}\right)$$

et donc

$$\mu_k^+(\mathcal{G}^{>2}) = O\left(\frac{1}{k^3}\right).$$

8.3.2 Synthèse

Nous avons désormais les résultats nécessaires pour estimer la fraction limite des tautologies non intuitionnistes.

Proposition 12 *L'ensemble $Peirce_k$ des tautologies non intuitionnistes vérifie :*

$$\mu_k^-(Peirce_k) = \frac{1}{2k^2} + O\left(\frac{1}{k^3}\right) \text{ et } \mu_k^+(Peirce_k) = \frac{1}{2k^2} + O\left(\frac{1}{k^3}\right).$$

Preuve: La preuve de cette proposition découle immédiatement des Lemmes 31 et 32 ainsi que des remarques les précédant. □

Nous sommes donc désormais en mesure de reprendre la Figure 8.2 [p. 74], de nous concentrer uniquement sur la partie concernant les tautologies et de donner plus de détails que précédemment. Nous connaissons donc les deux premiers termes du développement de la fraction limite des tautologies. De plus, nous savons lesquelles sont intuitionnistes. Ces détails sont représentés dans sur la Figure 8.14.

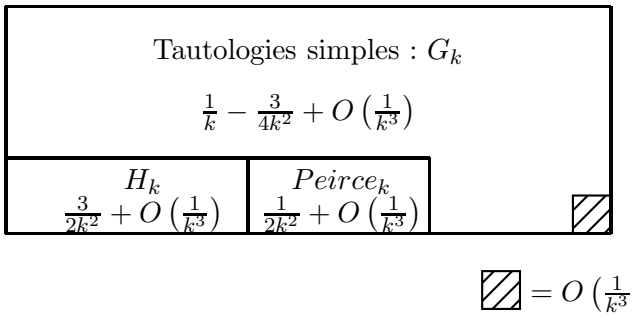

FIG. 8.14 – Ensemble des tautologies classiques : $\mu_k(Cl_k) = \frac{1}{k} + \frac{5}{4k^2} + O\left(\frac{1}{k^3}\right)$

Chapitre 9

Logiques plus riches

Le but de cette partie est d'enrichir l'ensemble des connecteurs utilisés, mais également d'autoriser la négation. Soient $\mathcal{C} = \{\rightarrow, \wedge, \vee\}$ l'ensemble des connecteurs et $\{x_1, \ldots, x_k, \perp\}$ l'ensemble des étiquettes possibles dans les feuilles des expressions. L'étiquette $\perp$ représente la constante *Faux* et donc $x_i \rightarrow \perp$ est équivalent à la négation de x_i. En substituant n'importe quelle expression A à x_i, l'expression $A \rightarrow \perp$ représente la négation de A. Nous nous plaçons donc dans un système propositionnel complet.

Par abus de notation nous réutilisons les mêmes noms d'ensembles que précédemment. Ainsi Int_k et Cl_k représentent désormais respectivement les tautologies intuitionnistes et classiques dans le système complet, et μ_k représente la fraction limite d'un ensemble par rapport à l'ensemble de toutes les expressions du système propositionnel complet. Il en est de même pour μ_k^- et μ_k^+. En dehors de ce chapitre, toutes ces notations sont relatives au système de l'implication.

Théorème 4 *Dans le système propositionnel complet, nous avons*

$$\lim_{k\to\infty} \frac{\mu^-(Int_k)}{\mu(Cl_k)} = \lim_{k\to\infty} \frac{\mu^+(Int_k)}{\mu(Cl_k)} = \frac{5}{8}.$$

Ainsi, alors qu'asymptotiquement la plupart des tautologies étaient intuitionnistes dans les système de l'implication, seulement $5/8^e$ des tautologies sont intuitionnistes dans le système propositionnel complet.

Ce chapitre va établir une démonstration de ce résultat mais aussi expliquer pour quelle raison la limite ne vaut plus 1. Ceci nous permettra, dans l'extension 9.4.1 d'enrichir au fur et à mesure l'ensemble des connecteurs pour passer de $\{\rightarrow\}$ à $\{\rightarrow, \wedge, \vee\}$ et de conclure quant à la valeur des limites.

Commençons par établir quelques généralités relatives à ce nouveau système : Les expressions sont obtenues via la grammaire suivante :

$$F = x_1 \mid \ldots \mid x_k \mid \perp \mid F \rightarrow F \mid F \wedge F \mid F \vee F.$$

La fonction génératrice énumérant toutes les expressions de ce système vérifie :

$$f_k(z) = (k+1)z + 3f_k(z)^2.$$

En effet, une expression est soit une feuille étiquetée par une variable ou la constante $\perp$, soit une expression dont la racine est étiquetée par l'un des trois connecteurs binaires et les fils

gauche et droit sont deux expressions. Résoudre cette équation (et choisir la solution adaptée $f_k(0) = 0$) donne :

$$f_k(z) = \frac{1 - \sqrt{1 - 12(k+1)z}}{6}.$$

Ainsi f_k possède une unique singularité $\rho = 1/(12(k+1))$.

9.1 Tautologies simples

Le but de cette partie est de présenter une famille de tautologies que nous appellerons simples. Le reste du chapitre prouvera qu'asymptotiquement par rapport à k, la plupart des tautologies sont simples.

Dans un arbre, un noeud (interne ou externe) est dit *positif* si le chemin joignant ce noeud à la racine ne passe par aucun noeud étiqueté par $\wedge$ et lorsqu'il passe par un noeud $\rightarrow$, alors le chemin continue dans le fils droit de ce noeud.

Un noeud est dit *négatif* s'il existe, sur le chemin le joignant à la racine, un noeud positif η étiqueté par $\rightarrow$ en lequel, le chemin se prolonge dans le fils gauche (de η) et s'arrête, ou ne traverse ensuite plus que des noeuds étiquetés par $\wedge$. La Figure 9.1 présente des noeuds positifs et négatifs. Le chemin de la racine à η sera appelé *préfixe positif* du chemin menant au noeud négatif.

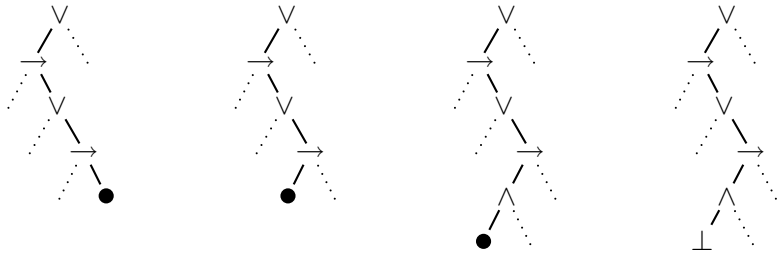

FIG. 9.1 – De gauche à droite : Un noeud positif, un noeud non positif, un noeud négatif, un noeud négatif étiqueté par $\perp$.

Nous remarquons qu'il suffit d'évaluer un noeud positif à 1 pour que toute l'expression s'évalue à 1. En effet le chemin de la racine au noeud positif ne passe que par des noeuds $\vee$ ou du côté droit des noeuds $\rightarrow$, donc tout l'arbre s'évalue à 1. De même, il suffit d'évaluer un noeud négatif à 0 pour que toute l'expression s'évalue à 1. Une récurrence sur la taille de l'arbre établit une démonstration plus formelle.

La Figure 9.2 représente une arbre dont les feuilles sont numérotées de 1 à 9. Nous remarquons que les feuilles 1 et 6 sont négatives et la feuille 7 est positive. Les autres feuilles ne sont ni positives ni négatives.

Soient $G_k^{\perp}$ les expressions contenant *une feuille négative étiquetée par* $\perp$, et G_k^r les expressions possédant *une feuille positive et une négative, toutes les deux étiquetées avec la même variable* x_i.

Lemme 33 *Les ensembles $G_k^{\perp}$ et G_k^r ne contiennent que des tautologies.*

Preuve: Grâce à la remarque précédente, ceci est clair pour $G_k^{\perp}$. En ce qui concerne G_k^r : Soit $A \in G_k^r$ avec les feuilles positive et négative étiquetées par x_i. En évaluant x_i à 1, la feuille positive entraîne le fait que tout l'arbre s'évalue à 1. Et en évaluant x_i à 0, la feuille négative entraîne le fait que tout l'arbre s'évalue à 1. □

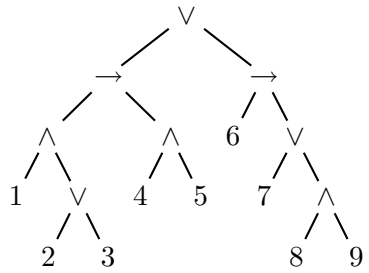

FIG. 9.2 – Exemple de feuilles négatives ou positives.

La théorie concernant les algèbres de Heyting s'applique à ce nouveau modèle (contenant deux connecteurs supplémentaires). Par conséquent le Fait 3 [p. 72] nous permet de caractériser les tautologies intuitionnistes.

Proposition 13 *Les expressions de $G_k^{\perp}$ sont des tautologies intuitionnistes.*

Preuve: Soit v une évaluation. Les expressions de $G_k^{\perp}$ ont une feuille négative étiquetée par $\perp$. Soit $A \in G_k^{\perp}$. Soit η le noeud $\rightarrow$ du chemin négatif où le chemin part à gauche. Notons A_η le sous-arbre enraciné en η. Nous avons $[[A_\eta]]_v = \mathbb{R}$ du fait de la particularité du chemin négatif à partir de η. Or le chemin de la racine de l'arbre jusqu'à η est un chemin positif, donc $[[A]]_v = \mathbb{R}$. L'arbre A est une tautologie intuitionniste. □

Nous notons que les expressions de G_k^r ne sont pas toutes intuitionnistes. En effet, $x_1 \vee (x_1 \rightarrow x_2)$ appartient à cet ensemble mais n'est pas intuitionniste – il suffit de prendre une évaluation v affectant l'ouvert $\mathbb{R} \setminus \{0\}$ à x_1 et l'ensemble vide à x_2. Toutefois $G_k^r \cap Int_k$ n'est pas vide – $x_1 \rightarrow x_1$ appartient également à G_k^r.

Pour une expression $A \in \mathcal{F}_k$, nous dirons qu'une feuille ν de A appartient à la *strate* s si le chemin allant de la racine à ν part exactement s fois à gauche d'un noeud $\rightarrow$. Par ailleurs une feuille ν appartient à l'*enveloppe* s s'il existe une strate s' telle que $s' \leqslant s$ et ν appartient à la strate s'. Sur la Figure 9.3, nous avons étiqueté les feuilles par la valeur de leur strate.

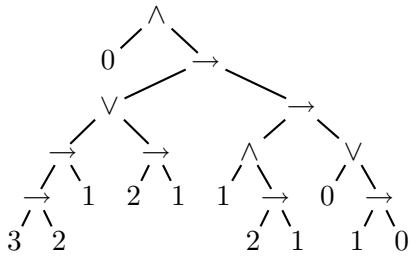

FIG. 9.3 – Feuilles étiquetées par la valeur de leurs strates.

Nous notons que toutes les feuilles positives sont dans la strate 0 et toutes les négatives dans la strate 1. Toutefois, toutes les feuilles de la strate 0 ne sont pas positives, et toutes celles de la strate 1 ne sont pas négatives.

Par la suite, nous appellerons une *répétition*, deux occurrences de la même variable ou une occurrence de la constante $\perp$. Ainsi, dans l'expression $(x_1 \vee x_2) \rightarrow (x_1 \rightarrow \perp)$, nous dirons qu'il y a 2 répétitions. Soit $\mathcal{H}$ l'ensemble des expressions ayant exactement une répétition dans l'enveloppe 1. Nous remarquons que $\mathcal{H} \setminus \{G_k^{\perp} \cup G_k^r)$ ne contient pas de tautologies – il

suffit d'évaluer à 0 les variables de la strate 0 et à 1 les autres pour obtenir une évaluation à 0 d'une expression de cet ensemble. Afin de définir précisément les expressions intuitionnistes de G_k^r, nous ne nous intéresserons qu'aux expressions qui sont des tautologies en raison de la répétition de la variable x_i et non pas en raison d'une autre partie de l'arbre.

Proposition 14 *Soit $A \in G_k^r \cap \mathcal{H}$ et x_i la variable répétée dans les noeuds positif et négatif. L'arbre A est une tautologie intuitionniste si et seulement si le préfixe positif du chemin menant à la feuille négative x_i, est inclus dans le chemin menant à la feuille positive x_i. (voir Figure 9.4).*

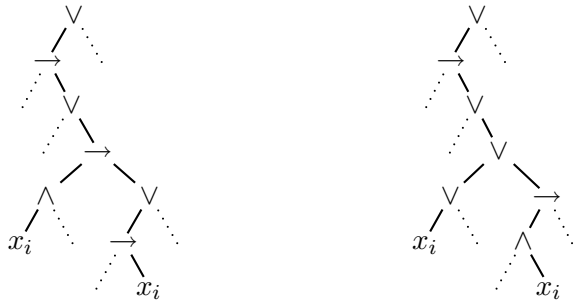

FIG. 9.4 – Des expressions de G_k^r : intuitionniste à gauche et non intuitionniste à droite.

Preuve: Soit $A \in G_k^r \cap \mathcal{H}$, il existe une unique variable x_i répétée dans un noeud positif et un négatif. Pour la première implication, raisonnons par l'absurde. Supposons que A soit intuitionniste et que le préfixe positif du chemin menant à la feuille négative x_i, ne soit pas inclus dans le chemin menant à la feuille positive x_i. Nous avons donc une expression ressemblant à l'expression de la Figure 9.4. Soit v une évaluation associant l'ouvert $\mathbb{R} \setminus \{0\}$ à x_i, $\mathbb{R}$ aux variables de la strate 1 et $\emptyset$ aux autres variables. Une récurrence sur la structure d'un arbre ne contenant aucune répétition dans la strate 1 et la remarque suivante permettent de prouver que A s'interprète à $\mathbb{R}\setminus\{0\}$, ce qui est contradictoire. Nous remarquons que l'arbre enraciné en η (le dernier noeud positif du chemin négatif) s'interprète à l'ensemble vide. En remontant le chemin négatif on tombe sur le noeud $\vee$ qui sépare le chemin négatif du positif. L'arbre enraciné en ce noeud s'interprète à $\mathbb{R} \setminus \{0\}$.

Pour démontrer la réciproque, supposons que le préfixe positif du chemin menant à la feuille négative x_i, soit inclus dans le chemin menant à la feuille positive x_i (voir l'expression gauche de la Figure 9.4). Soit η le noeud étiqueté par $\rightarrow$ où le chemin négatif se sépare du positif. Appelons B et C respectivement ses fils gauche et droit. Soit v une évaluation. Dans B, il y a un chemin ne contenant que des $\wedge$ (ou alors aucun noeud interne) qui débouche sur x_i. Ainsi $[[B]]_v \subset [[x_i]]_v$. Dans C, il y a un chemin positif pour rejoindre x_i, donc $[[C]]_v \supset [[x_i]]_v$. Le fait que η soit étiqueté par $\rightarrow$ implique que l'arbre enraciné en η s'interprète à $\mathbb{R}$. Or le chemin de la racine globale à η est un chemin positif, donc A s'interprète à $\mathbb{R}$, i.e. A est une tautologie simple. □

Maintenant que nous discernons les tautologies intuitionnistes dans l'ensemble des tautologies simples, procédons à l'énumération de ces tautologies. Soit $\phi^{\wedge}(y,z)$ la fonction génératrice énumérant tous les arbres, z marquant la taille et y les feuilles telles que le chemin les reliant à la racine de l'arbre ne contient que des noeuds $\wedge$ – ou est vide.

$$\phi^{\wedge}(y,z) = (k+1)yz + \phi^{\wedge}(y,z)^2 + 2f_k(z)^2.$$

En effet, soit un arbre est une feuille, soit il possède une racine $\wedge$, soit une racine $\rightarrow$ ou $\vee$. Du fait que $\phi^{\wedge}(0,0) = 0$, on en déduit :

$$\phi^{\wedge}(y,z) = \frac{1 - \sqrt{1 - 8f_k(z)^2 - 4(k+1)yz}}{2}.$$

Soit $\phi^{<0}(y,z)$ la fonction génératrice énumérant tous les arbres, z marquant la taille et y les feuilles négatives.

$$\phi^{<0}(y,z) = (k+1)z + \phi^{\wedge}(y,z)\phi^{<0}(y,z) + \phi^{<0}(y,z)^2 + f_k(z)^2.$$

L'équation est obtenue en analysant la racine des expressions. Soit l'expression est une feuille (non négative), soit le connecteur de la racine est $\rightarrow$, soit il s'agit de $\vee$ soit enfin le connecteur est $\wedge$.

Proposition 15 *La fraction limite de l'ensemble $G_k^{\perp}$ vaut*

$$\mu_k(G_k^{\perp}) = \frac{5}{8k} + O\left(\frac{1}{k^2}\right).$$

Preuve: Soit G^1 l'ensemble des tautologies ayant au moins une occurrence de $\perp$ parmi ses feuilles négatives. Soit $\phi_1(z)$ la fonction génératrice

$$\phi_1(z) = \frac{1}{k+1}\left(y\frac{\partial \phi^{<0}}{\partial y}(y,z)\right)_{|y=1}.$$

Afin de pointer une feuille dans un arbre, il suffit de la marquer via une variable supplémentaire dans la fonction génératrice énumérant les arbres, puis d'effectuer une dérivation suivant cette variable dans l'expression obtenue, et enfin d'y substituer 1 (la méthode est détaillée dans le livre de Flajolet et Sedgewick [FS96, chap. 3]). Nous remarquons que $\phi_1(z)$ n'énumère pas exactement G^1 car les tautologies ayant plusieurs occurrences de $\perp$ en strate 1 sont comptées plusieurs fois. Soit G^2 l'ensemble des tautologies ayant au moins deux occurrences de $\perp$ parmi ses feuilles négatives. Soit $\phi_2(z)$ la fonction génératrice :

$$\phi_2(z) = \frac{1}{(k+1)^2}\left(y^2\frac{\partial^2 \phi^{<0}}{\partial y^2}(y,z)\right)_{|y=1}.$$

A nouveau, les tautologies ayant plus de deux occurrences de $\perp$ dans des feuilles négatives sont comptées plusieurs fois. Cependant ce qui nous intéresse ici c'est une borne supérieure de la fraction limite des expressions avec au moins deux répétitions de $\perp$ dans les feuilles négatives. Bien que nous comptions plusieurs fois des arbres, la borne que nous obtenons est suffisamment petite, donc nous pouvons nous en contenter.

En calculant les fractions limites des ensembles énumérés par ces fonctions génératrices, on obtient :

$$\mu_k(G^1) = \frac{5}{8k} + O\left(\frac{1}{k^2}\right),$$

$$\mu_k(G^2) = O\left(\frac{1}{k^2}\right).$$

Puisque $G^1 \setminus G^2 \subset G_k^{\perp} \subset G^1$, on en conclut l'énoncé de la proposition. □

Soit $\phi^{>0}(x,z)$ la fonction génératrice énumérant tous les arbres, z marquant la taille et x les feuilles positives.

$$\phi^{>0}(x,z) = (k+1)xz + f_k(z)\phi^{>0}(x,z) + \phi^{>0}(x,z)^2 + f_k(z)^2.$$

L'équation est obtenue en analysant la racine des expressions. Soit l'expression est une feuille (positive), soit le connecteur de la racine est $\rightarrow$, soit il s'agit de $\vee$ soit enfin le connecteur est $\wedge$.

Soit $\phi^{>0;<0}(x,y,z)$ la fonction génératrice énumérant tous les arbres, z marquant la taille, y les feuilles négatives et x les feuilles positives.

$$\phi^{>0;<0}(x,y,z) = (k+1)xz + \phi^{\wedge}(y,z)\phi^{>0;<0}(x,y,z) + \phi^{>0;<0}(x,y,z)^2 + f_k(z)^2.$$

De nouveau l'équation est obtenue en analysant la racine des expressions.

Proposition 16 *Les fractions limites des ensembles $G_k^r \cap \mathcal{H}$ et $(G_k^r \cap \mathcal{H}) \cap Int_k$ valent*

$$\mu_k(G_k^r \cap \mathcal{H}) = \frac{11}{8k} + O\left(\frac{1}{k^2}\right),$$

$$\mu_k((G_k^r \cap \mathcal{H}) \cap Int_k) = \frac{5}{8k} + O\left(\frac{1}{k^2}\right).$$

Preuve: Soit G^3 l'ensemble des tautologies ayant au moins deux occurrences de la même variable, l'une parmi ses feuilles négatives, l'autre parmi ses feuilles positives. Soit $\phi_3(z)$ la fonction génératrice suivante :

$$\phi_3(z) = \frac{k}{(k+1)^2}\left(xy\frac{\partial^2\phi^{<0}}{\partial x \partial y}(x,y,z)\right)_{|x=1;y=1}.$$

Nous remarquons que $\phi_3(z)$ n'énumère pas exactement G^3 car les tautologies ayant plusieurs répétitions sont comptées plusieurs fois.

Soit G^4 l'ensemble des tautologies ayant au moins deux répétitions parmi ses feuilles positives et négatives. Soit $\phi_4(z)$ la fonction génératrice :

$$\phi_4(z) = \frac{k^2}{(k+1)^4}\left(x^2y^2\frac{\partial^4\phi^{<0}}{\partial x^2 \partial y^2}(y,z)\right)_{|x=1;y=1}.$$

A nouveau, les tautologies ayant plus de deux répétitions sont comptées plusieurs fois. En calculant les fractions limites des ensemble énumérés par ces fonctions génératrices, on obtient :

$$\mu_k(G^3) = \frac{11}{8k} + O\left(\frac{1}{k^2}\right),$$

$$\mu_k(G^4) = O\left(\frac{1}{k^2}\right).$$

Puisque $G^3 \setminus G^4 \subset G_k^r \subset G^3$, on en conclut :

$$\mu_k(G_k^r \cap \mathcal{H}) = \frac{11}{8k} + O\left(\frac{1}{k^2}\right).$$

Intéressons-nous désormais aux tautologies intuitionnistes de cet ensemble G_k^r. Nous avons observé que les expressions intuitionnistes de cet ensemble ont un noeud η positif et étiqueté par $\rightarrow$. Son fils gauche contient une feuille négative (pour l'arbre global), i.e. un chemin ne passant que par des $\wedge$ et débouchant sur l'une des occurrences de la variable répétée. Son fils droit contient un chemin positif jusqu'à la seconde occurrence de la variable répétée. Soit $\phi_{int}^{>0;<0}(x,y,z)$ la fonction génératrice énumérant tous les arbres intuitionnistes de ce type, z marquant la taille, y les feuilles négatives et x les feuilles positives.

$$\phi_{int}^{>0;<0}(x,y,z) = \frac{1}{(k+1)z}\left(\frac{\partial\phi^p}{\partial x}(x,z)\right)_{|x=1}\phi^{\wedge}(y,z)\phi^p(x,z).$$

Soit G^5 l'ensemble des tautologies de $(G_k^r \cap \mathcal{H}) \cap Int_k$ ayant au moins deux occurrences de la même variable, l'une parmi ses feuilles négatives, l'autre parmi ses feuilles positives. Soit $\phi_5(z)$ la fonction génératrice suivante :

$$\phi_5(z) = \frac{k}{(k+1)^2}\left(xy\frac{\partial^2\phi_{int}^{>0;<0}}{\partial x\partial y}(x,y,z)\right)_{|x=1;y=1}.$$

Nous remarquons que $\phi_5(z)$ n'énumère pas exactement G^5 car les tautologies ayant plusieurs répétitions sont comptées plusieurs fois.

Soit G^6 l'ensemble des tautologies ayant au moins deux répétitions parmi ses feuilles positives et négatives. Soit $\phi_6(z)$ la fonction génératrice :

$$\phi_6(z) = \frac{k^2}{(k+1)^4}\left(x^2y^2\frac{\partial^4\phi_{int}^{>0;<0}}{\partial x^2\partial y^2}(y,z)\right)_{|x=1;y=1}.$$

A nouveau, les tautologies ayant plus de deux répétitions sont comptées plusieurs fois. En calculant les fractions limites des ensembles énumérés par ces fonctions génératrices, on obtient :

$$\mu_k(G^5) = \frac{5}{8k} + O\left(\frac{1}{k^2}\right),$$

$$\mu_k(G^6) = O\left(\frac{1}{k^2}\right).$$

Puisque $G^5 \setminus G^6 \subset (G_k^r \cap \mathcal{H}) \cap Int_k \subset G^5$, on en conclut :

$$\mu_k((G_k^r \cap \mathcal{H}) \cap Int_k) = \frac{5}{8k} + O\left(\frac{1}{k^2}\right).$$

□

9.2 Autres tautologies

Dans cette partie nous allons déterminer quelle structure ont les autres tautologies et par la suite nous calculerons la fraction limite des expressions ayant cette structure. Nous nous rendrons compte qu'elle vaut $O(1/k^2)$. Par conséquent, les autres tautologies sont en quantité négligeable par rapport aux tautologies simples.

Divisons tout d'abord l'ensemble des tautologies par rapport à leur nombre de répétitions dans telle ou telle enveloppe.

(a) E_a : l'ensemble des tautologies avec zéro répétition dans l'enveloppe 1
(b) $(G_k^{\perp} \cup G_k^r) \cap \mathcal{H}$
(c) E_c : l'ensemble des tautologies avec une répétition dans l'enveloppe 1 mais n'appartenant pas à $G_k^{\perp} \cup G_k^r$
(d) E_d : l'ensemble des tautologies avec plusieurs répétitions dans l'enveloppe 1.

Nous allons remarquer qu'il ne s'agit pas tout à fait d'une partition car certains ensembles sont vides. Puis nous allons évaluer les fractions limites des autres ensembles. La fraction limite de l'ensemble donné en (b) a été traité dans la partie précédente.

Lemme 34 *Soit A une expression ne contenant aucune répétition dans l'enveloppe 1. Soit v une évaluation affectant la valeur 0 aux variables de la strate 0 de A et 1 à celles de la strate 1. Soit w une évaluation affectant la valeur 1 aux variables de la strate 0. L'arbre A s'évalue à 0 pour v et à 1 pour w.*

Preuve: Soient $A \in \mathcal{F}_k$ ne contenant aucune répétition dans l'enveloppe 1, et v et w des évaluations vérifiant les conditions suivantes : v affecte la valeur 0 aux variables de la strate 0 de A et 1 à celles de la strate 1 et w affecte la valeur 1 aux variables de la strate 0. Démontrons le résultat par récurrence sur la taille de A. Supposons que A soit de taille 1, i.e. A est une feuille. L'arbre s'évalue à 0 pour v et à 1 pour w. Supposons le résultat vrai pour tous les arbres de taille inférieure strictement à n. Soit A un arbre de taille n. Notons respectivement G et D ses fils gauche et droit. Si le connecteur dans la racine de A est $\vee$ ou $\wedge$, alors l'ensemble des noeuds de la strate 0 de A correspond exactement à l'union disjointe de l'ensemble dessnoeuds de la strate 0 de G et de l'ensemble des noeuds de la strate 0 de D. Il en est de même pour ceux de la strate 1 de A. Or G et D s'évaluent tous les deux à 0 pour v et à 1 pour w. Par conséquent, il en est de même pour A. Supposons que la racine de A soit étiquetée par $\rightarrow$. Les noeuds de la strate 0 de D sont sur la strate 0 de A – et il en est de même pour ceux de la strate 1 de D. Par contre, les noeuds de la strate 0 de G appartiennent à la strate 1 de A. Ainsi, pour v, qui évalue à 0 la strate 0 de A et à 1 la strate 1, le sous-arbre D s'évalue à 0 et le sous-arbre G à 1. Donc pour v l'arbre global A s'évalue à 0. Pour w, la strate 0 de A est évaluée à 1, donc D s'évalue à 1 et il en est donc de même pour A. Ainsi, la proposition est toujours vraie au rang n. □

En conséquence l'ensemble E_a est vide.

Lemme 35

$$\mu_k^+(E_c) = O\left(\frac{1}{k^2}\right).$$

Preuve: Commençons par démontrer que les expressions de E_c ont une seconde répétition dans l'enveloppe 2. Par définition, les expressions de E_c ont une répétition dans l'enveloppe 1 mais n'appartiennent pas à $G_k^{\perp} \cup G_k^r$. Soit $A \in E_c$. Soit une feuille de la strate 0 de A est étiquetée par $\perp$, mais puisqu'aucune autre répétition n'existe dans l'enveloppe 1, il suffit d'évaluer à 0 les variables de la strate 0 et à 1 celle de la strate 1 pour que l'arbre A s'évalue à $faux$ (voir Lemme 34, en remplaçant $\perp$ par une nouvelle variable qu'on évalue à 0) : impossible car A est une tautologie. Soit une feuille non négative de la strate 1 est étiquetée par $\perp$. Entre le noeud η (étiqueté par $\rightarrow$) où le chemin part sur la gauche et cette feuille, il existe un connecteur $\vee$ ou $\rightarrow$ (dans ce cas la suite du chemin part à droite). Dans le cas du connecteur $\vee$, le fait d'évaluer la strate 1 à 1 (sauf $\perp$) fera que le sous-arbre gauche de η

s'évaluera à 1 et donc l'arbre global s'évalue à 0 dès lors que les feuilles de la strate 0 sont à 0 : ce ne sont pas des tautologies. Dans le cas du connecteur $\rightarrow$, supposons qu'il n'y ait qu'une seule répétition dans l'enveloppe 2 (la même que celle de l'enveloppe 1), on peut alors évaluer les feuilles de la strate 2 à 0, celles de la strate 1 à 1 (sauf $\perp$) et celles de la strate 0 à 0, l'arbre global s'évalue dès lors à 0 : ce n'est pas une tautologie. Donc, il y a au moins deux répétitions dans l'enveloppe 2 – dont une répétition dans l'enveloppe 1. Enfin, il se peut que les deux occurrences de la même variable n'ont pas lieu simultanément dans une feuille négative et une positive. Si les deux occurrences appartiennent à la même strate, d'après le Lemme 34 de tels arbres ne sont pas des tautologies. Donc une occurrence appartient à la strate 0 et l'autre à la strate 1. De la même manière que dans le cas précédent (avec $\perp$ dans strate 1), on démontre qu'il existe une seconde répétition dans la l'enveloppe 2.

Enumérons l'ensemble des expressions contenant deux répétitions dans l'enveloppe 2. Cet ensemble contient E_c. Soit $\phi(t,u)$ la fonction génératrice énumérant les arbres construits avec les trois connecteurs $\{\rightarrow,\wedge,\vee\}$, tels que les fils gauches de $\rightarrow$ soient réduits à une feuille, marquée par t et les feuilles de la strate 0 par la variable u.

$$\phi(t,u) = u + t\phi(t,u) + 2\phi(t,u)^2.$$

En effet, une expression est soit une feuille de la strate 0, soit un arbre de racine $\rightarrow$ avec t son fils gauche et un fils droit quelconque, soit un arbre de racine $\wedge$ ou $\vee$ avec des fils quelconques. En choisissant la solution vérifiant la condition initiale $\phi(0,0) = 0$, on obtient :

$$\phi(t,u) = \frac{1-t-\sqrt{(1-t)^2-8u}}{4}.$$

Soit $\phi_{0,1,2}(z,u)$ la fonction génératrice énumérant les arbres tels que les feuilles de la strate 0, 1 et 2 sont marquées par u et les autres feuilles par z.

$$\phi_{0,1,2}(z,u) = \phi(\phi(\phi(f_k(z),u),u),u).$$

En effet, le fait d'itérer trois fois la fonction ϕ permet de marquer les feuilles des strates 0 à 2. Le fait de différencier quatre fois par rapport à u (et de multiplier par u^4) permet de compter de combien de manières on peut étiqueter les feuilles des strates 0 à 2 avec quatre feuilles marquées. Puis le fait de remplacer u par $(k+1)z$ étiquette les feuilles des strates 0 à 2. Enfin il faut diviser le tout par $(k+1)^2$ car les deux étiquettes marquées doivent être des répétitions des deux autres. Ici nous n'avons pas spécifié si les deux répétitions correspondaient à trois occurrences de la même variable, ou deux occurrences de $\perp$, ou deux occurrences de deux variables différentes. Cependant, dans les calculs, nous effectuons des majorations en multipliant par $(k+1)^2/(k+1)^4$ ce qui engendre une majoration de l'énumération de tous les cas.

$$\phi(z) = \frac{1}{(k+1)^2}\left(u^4\frac{\partial^4\phi_{0,1,2}}{\partial u^4}(z,u)\right)_{u=(k+1)z}.$$

Nous obtenons :

$$\mu_k^+(E_c) \leqslant \lim_{n\to\infty}\frac{[z^n]\phi(z)}{[z^n]f_k(z)} = O\left(\frac{1}{k^2}\right).$$

□

Lemme 36

$$\mu_k^+(E_d) = O\left(\frac{1}{k^2}\right).$$

La preuve se fait de la même manière que la seconde partie de la preuve du Lemme 35.

9.3 Synthèse

Nous sommes désormais en mesure de présenter une démonstration du Théorème 4 [p. 87].

Preuve: La preuve du théorème provient des Propositions 15 et 16 et des Lemmes 35 et 36. En effet,

$$\mu_k(G_k^{\perp}) + \mu_k((G_k^r \cap \mathcal{H}) \cap Int_k) \leqslant \mu_k^-(Int_k) \leqslant \mu_k^+(Int_k),$$

$$\mu_k^+(Int_k) \leqslant \mu_k(G_k^{\perp}) + \mu_k((G_k^r \cap \mathcal{H}) \cap Int_k) + \mu_k^+(E_c) + \mu_k^+(E_d).$$

Par ailleurs,

$$\mu_k(G_k^{\perp}) + \mu_k(G_k^r \cap \mathcal{H}) \leqslant \mu_k(Cl_k) \leqslant \mu_k(G_k^{\perp}) + \mu_k(G_k^r \cap \mathcal{H}) + \mu_k^+(E_c) + \mu_k^+(E_d).$$

Par conséquent on en déduit :

$$\lim_{k\to\infty} \frac{\mu_k^-(Int_k)}{\mu_k(Cl_k)} = \lim_{k\to\infty} \frac{\mu_k^+(Int_k)}{\mu_k(Cl_k)} = \frac{5}{8}.$$

□

9.4 Extensions

9.4.1 Systèmes logiques intermédiaires

Nous pouvons aisément démontrer que les deux familles de tautologies simples que nous avons exhibées restent les familles prépondérantes de tautologies lorsqu'on restreint l'ensemble de connecteurs ou lorsqu'on n'utilise plus la constante *Faux*. Bien entendu ces familles se restreignent aussi dans ces cas. Le cas extrême est obtenu avec la logique de l'implication, lorsqu'on n'utilise plus les connecteurs $\wedge$ ni $\vee$ et non plus la constante *Faux*. Dès lors, la famille $G_k^{\perp}$ est vide, puisque $\perp$ est banni. Par ailleurs, il y a toujours une unique feuille positive, il s'agit du but de l'arbre. Pour obtenir une feuille négative, il faut qu'une prémisse soit réduite à une feuille (du fait que le connecteur $\wedge$ n'est pas utilisable). Nous nous apercevons que notre ensemble G_k^r se restreint à la famille des tautologies simples définies dans le chapitre 5 [p. 23].

Ainsi, simplement en énumérant (modification aisée des fonctions génératrices) à nouveau les sous-ensembles des familles décrites dans ce chapitre, pour une logique donnée, nous retrouvons que la plupart des tautologies sont simples (k tend vers l'infini) et par conséquent, l'on peut à nouveau comparer les logiques classique et intuitionniste. Voilà une synthèse des résultats, qu'on obtient.

Tout d'abord, nous remarquons que dans les systèmes logiques ne contenant pas le connecteur $\vee$, il n'y a pas de tautologies simples qui soient non intuitionnistes et donc, dans tous ces cas, le rapport entre les fractions limites des tautologies intuitionnistes et les tautologies classiques tend vers 1 lorsque k tend vers l'infini. Lorsque $\vee$ fait partie des connecteurs, voilà un tableau récapitulatif du rapport entre ces fractions limites :

Remarquons que dans tous les systèmes nous avons :

$$lim_{k\to\infty} \frac{\mu_k^-(Int_k)}{\mu_k(Cl_k)} = lim_{k\to\infty} \frac{\mu_k^+(Int_k)}{\mu_k(Cl_k)}.$$

connecteurs	$lim_{k\to\infty}\ \mu_k^{\pm}(Int_k)/\mu_k(Cl_k)$
$\{\rightarrow\}$	1
$\{\rightarrow,\bot\}$	1
$\{\rightarrow,\wedge\}$	1
$\{\rightarrow,\wedge,\bot\}$	1
$\{\rightarrow,\vee\}$	3/13
$\{\rightarrow,\vee,\bot\}$	2/7
$\{\rightarrow,\vee\wedge\}$	5/11
$\{\rightarrow,\vee\wedge,\bot\}$	5/8

FIG. 9.5 – Différences quantitatives entre logiques intuitionniste et classique.

9.4.2 Modèle avec un nombre non borné de littéraux

Le lecteur se sera aperçu que nous avons pu obtenir des résultats exploitables dès lors que le nombre de variables devient grand. Ainsi, dans le modèle des grands arbres, nous calculons la fraction limite en étudiant les grands arbres (n tend vers l'infini) puis le résultat est analysé lorsque le nombre de variables devient grand (k tend vers l'infini). Les conditions ne sont pas réunies pour permuter directement les deux limites – il n'y a pas de convergence uniforme. Est-il imaginable de regarder le problème sous un nouvel angle, en considérant que nous travaillons avec un nombre infini (mais dénombrable) de variables et que dans ce contexte nous nous intéressions à la fraction limite ?

Nous remarquons immédiatement un problème résultant de la définition de la fraction limite. Celle-ci définit une proportion entre des ensembles d'arbres. Cependant, dans ce nouveau contexte, les ensembles d'arbres sont infinis et la fraction limite n'est donc plus définie. Par conséquent, nous introduisons la relation d'équivalence α suivante : *deux expressions sont α-équivalentes s'il existe une application injective r, appelée renommage, de l'ensemble des variables dans lui-même telle que la seconde expression est obtenue à partir de la première après l'avoir réétiquetée selon r*. Nous noterons $\mathcal{F}_\infty$ l'ensemble des formules à permutation des variables près. Les classes d'expressions à permutation des variables près sont désormais finies lorsque nous considérons des expressions de taille fixée.

Du fait que les tautologies restent des tautologies après renommage de leurs variables, cette nouvelle approche se justifie pour ces expressions. Nous noterons Cl_∞ et Int_∞ respectivement les classes des tautologies classiques et intuitionnistes. Pour cette approche, nous pouvons définir la suite suivante :

$$d_\infty(n) = \frac{Int_\infty(n)}{Cl_\infty(n)}.$$

Théorème 5 *Dans le modèle des expressions à permutation des variables près, nous avons*

$$\lim_{n\to\infty} d_\infty(n) = \frac{5}{8}.$$

Nous pouvons à nouveau donner l'interprétation informelle : *environ* $5/8^e$ *des tautologies classiques sont intuitionnistes.*

Preuve: Nous allons présenter les idées-clé de la preuve. Une preuve détaillée est écrite dans l'article écrit en collaboration avec Kozik [GK09]. Pour une expression $\tilde{\phi} \in \mathcal{F}_\infty$ de taille n (i.e. avec n feuilles), un étiquetage des feuilles de $\tilde{\phi}$ est une relation d'équivalence R sur l'ensemble

$\{0, 1, \ldots, n\}$ consistant en tous les couples de de numéros de feuilles qui sont étiquetées par le même symbole (variable ou $\bot$). Les couples de la forme $(0, j)$ ou $(j, 0)$ expriment le fait que la feuille j est étiquetée par $\bot$. Remarquons que la relation R ne dépend pas du représentant $\tilde{\phi}$ choisi dans la classe d'équivalence de $\tilde{\phi}$. En effet, la relation R contient l'information sur l'ensemble des feuilles ayant la même étiquette, mais ne dit rien sur l'étiquette elle-même sauf pour $\bot$ qui n'est pas concernée par la permutation de variables.

Nous pouvons désormais donner le cardinal de $\mathcal{F}_\infty(n)$:

$$\mathcal{F}_\infty(n) = C_{n-1}B(n+1),$$

où C_{n-1} correspond au $(n-1)^e$ nombre de Catalan et $B(n+1)$ correspond au nombre de relations d'équivalence sur l'ensemble $\{0, 1, \ldots, n\}$. Il s'agit du $(n+1)^e$ nombre de Bell – voir un des livres [GKP89] ou [FS09].

L'étude que nous avons réalisé dans le début du chapitre peut être entièrement adaptée à ce nouveau modèle. Pour cela, il suffit d'énumérer les arbres en deux temps. Tout d'abord on se concentre sur l'ensemble des arbres dont seules les feuilles ne sont pas étiquetées. Puis, il reste à compter le nombre de manières différentes d'étiqueter les feuilles. Dans cette partie, les nombres de Bell vont intervenir.

Or d'après le résultat de Moser et Wyman [MW55] concernant le comportement asymptotique de ces nombres :

$$\frac{B(n-2)}{B(n)} = o\left(\frac{B(n-1)}{B(n)}\right),$$

nous pouvons démontrer à nouveau que seules les tautologies simples ont une fraction limite significative pour la fraction limite des tautologies. □

Quatrième partie

Arbres équilibrés

Chapitre 10

Construction d'arbres équilibrés

En considérant spécifiquement la forme des arbres, une des plus régulières qui soit est l'arbre équilibré, que nous définissons par l'arbre binaire complet ayant toutes ses feuilles au même niveau, i.e. à la même profondeur. Nous notons immédiatement que, pour une taille donnée, le nombre de tels arbres est très restreint par rapport à l'ensemble des arbres. En effet, pour une taille n donnée, il y a exactement un arbre équilibré (sans étiquette), lorsque n est une puissance de 2, contre C_{n-1} arbres binaires complets au total (et aucun arbre équilibré si n n'est pas une puissance de 2). Ainsi nous nous attendons à obtenir des résultats bien différents pour l'étude de la fraction limite des arbres équilibrés par rapport à celle des arbres sans forme particulière. La classe d'arbres obtenue en ne conservant que les arbres équilibrés a été étudiée en détails sous deux noms distincts : "amplification probabiliste" et "processus de croissance". L'initiateur de ce champ de recherche a été Valiant en 1984 [Val84]. Son but était de construire une expression de petite taille calculant la fonction *majorité* sur k variables Booléennes. Celle-ci vaut 1 pour une affectation a, si et seulement si au moins la moitié des composantes de a valent 1. Evidemment, le problème est intéressant lorsque le nombre de variables k devient grand. Au fil des années la méthode de Valiant a été élargie afin d'obtenir le même genre de résultats pour d'autres fonctions. Boppana [Bop85] généralise la méthode de Valiant. Il obtient des petites expressions pour chacune des *fonctions seuil* : généralisation de *majorité*.

De façon plus formelle, les expressions Booléennes sont construites avec un ensemble $\mathcal{C}$ de connecteurs, avec k variables $x_1, \ldots, x_k$ et de plus elles sont équilibrées. Un expression équilibrée a toutes ses feuilles à la même profondeur dans sa représentation sous forme d'arbre.

Soient $k > 0$, μ_0 une distribution sur $H_0 = \{x_1, \ldots, x_k\}$ et p une distribution sur $\mathcal{C}$. Pour tout $n \geqslant 1$ posons $H_n = \bigcup_{c \in \mathcal{C}} \{h_1\ c\ h_2 \mid h_1, h_2 \in H_{n-1}\}$ et μ_n une distribution de probabilité sur H_n vérifiant : $\mu_n(h_1\ c\ h_2) = p(c)\mu_{n-1}(h_1)\mu_{n-1}(h_2)$. Notons que H_n contient l'ensemble des arbres équilibrés de profondeur n dont les noeuds internes sont étiquetés de manière indépendante par un des connecteurs de $\mathcal{C}$ selon la distribution p.

La suite (μ_n) induit une suite (π_n) de distributions de probabilité sur $\mathcal{B}_k$ de la manière suivante (ces suites dépendent de manière implicite de k) : Pour tout $f \in \mathcal{B}_k$,

$$\pi_n(f) = \sum_{\{h \in H_n \mid h \text{ calcule } f\}} \mu_n(h).$$

Le but de l'étude est de démontrer l'existence d'une distribution limite pour (π_n) et de la caractériser. Les résultats que nous avons évoqués en introduction du chapitre ne considèrent que les arbres obtenus à partir d'un unique connecteur – $\mathcal{C}$ est un singleton.

Les travaux dans ce domaine peuvent se classer en deux types. Le premier intérêt de l'étude d'arbres équilibrés dans un système donné est l'obtention de petites expressions calculant une fonction ou une classe de fonctions.

Ainsi, Valiant [Val84] a construit de petites expressions pour la fonction *majorité*. Son système est basé sur le connecteur quaternaire $C(x, y, z, t) = (x \vee y) \wedge (z \vee t)$ et un ensemble H_0 contenant un nombre pair k de littéraux positifs et la constante 0. De plus, la distribution μ_0 pondère fortement la constante et partage le complément de manière uniforme sur les littéraux. Muni de ce système, l'auteur a démontré que la suite de distributions (π_n) admet une distribution limite concentrée sur la fonction *majorité(k)* et de plus, parmi les formules de tailles $O(k^{5,3})$ au moins la moitié d'entre elles calculent cette fonction. Afin d'obtenir son résultat, Valiant a utilisé des outils d'analyse réelle.

Boppana [Bop85] a obtenu de petites expressions pour chacune des fonctions *s-seuil* à k variables. Pour ce faire, il systématise la méthode de Valiant puis l'améliore pour le cas des fonctions *seuil*. Par ailleurs, il prouve que le résultat sur la vitesse de convergence de Valiant est optimal par rapport au système de départ. Il nomme la méthode : *Amplification probabiliste*.

Plus tard, Gupta et Mahajan [GM97] ont obtenu une meilleure amplification de la fonction *majorité* et des fonctions *seuil* que celle des deux auteurs originels, en exhibant un système convergeant plus rapidement vers sa distribution limite.

Enfin Servedio [Ser04] a étendu encore ces résultats en exhibant des systèmes permettant de construire les fonctions *seuil linéaires*. Au lieu que ce soit la somme des composantes d'une affectation qui détermine la valeur de la fonction évaluée en ce point, il s'agit d'une somme pondérée dont les poids sont fixés avec le système.

L'autre intérêt d'utiliser l'amplification réside en l'étude d'une classe de systèmes dont le connecteur vérifie certaines conditions. Ainsi, Savický [Sav90] démontre qu'un système dont l'ensemble d'expressions initial contient les $2k$ littéraux et les deux constantes 1 et 0, muni d'une distribution initiale μ_0 uniforme et d'un connecteur non linéaire et équilibré, admet une distribution limite uniforme sur l'ensemble de toutes les fonctions Booléennes à k variables. Son résultat utilise comme principal outil la transformée de Fourier.

Plus tard, Brodsky et Pippenger [BP05] ont étendu la classification selon que le connecteur soit linéaire, auto-dual ou monotone. On dit qu'une fonction Booléenne est *linéaire* si elle vérifie $f(x_1, \ldots, x_r) = x_1 \oplus \cdots \oplus x_r$ ou $x_1 \oplus \cdots \oplus x_r \oplus 1$. Une fonction est dite *auto-duale* si $f(x_1, \ldots x_k) = \overline{f}(\overline{x}_1, \ldots, \overline{x}_k)$. Enfin, une fonction *monotone* est telle que pour deux affectations $a = (a_1, \ldots, a_k)$ et $b = (b_1, \ldots, b_k)$ vérifiant pour tout $i \in \{1, \ldots, k\}$, $a_i \leqslant b_i$ alors $f(a) \leqslant f(b)$. Un connecteur possède l'une des trois propriétés si la fonction qu'il calcule la possède. Pour ces classes de connecteurs, Brodsky et Pippenger démontrent l'existence ou non d'une distribution limite d'une amplification et calculent également la vitesse de convergence de la suite de distributions (lorsqu'elle converge). Les démonstrations sont aussi basées sur la transformée de Fourier.

Les résultats qu'ils ont prouvés (ainsi que les autres travaux antérieurs) ne dépendent pas de l'arité des connecteurs. Par exemple, il est immédiat que le connecteur $\vee$ fait partie de la classe monotone, et il en est de même pour le connecteur de Valiant.

Nous remarquons que les résultats concernant un connecteur permettent d'obtenir les résultats concernant le dual du connecteur. En effet, pour une expression F calculant une fonction f, le fait de remplacer ses connecteurs par leur dual donne une expression F' calculant f', la fonction duale de f, définit de la manière suivante : $f'(x_1, \ldots, x_k) = \overline{f}(\overline{x}_1, \ldots, \overline{x}_k)$.

Toutefois, en se restreignant spécifiquement aux connecteurs binaires, nous remarquons

qu'un seul couple (connecteur, dual) ne fait partie d'aucune des classes de Brodsky et Pippenger. En effet, le connecteur implication ($\rightarrow$) ou son dual ($\cdot \wedge \bar{\cdot}$) ne sont ni linéaires, ni auto-duaux, ni monotones. Voila une synthèse des distributions limites obtenue en construisant des arbres équilibrés avec un connecteur binaire :

connecteur	propriétés du connecteur	distribution limite	preuve
$(x,y) \rightarrow x$ $(x,y) \rightarrow y$ $(x,y) \rightarrow 1$	linéaires	uniforme sur les	Brodsky et Pippenger [BP05]
$(x,y) \rightarrow \bar{x}$ $(x,y) \rightarrow \bar{y}$ $(x,y) \rightarrow 0$	connecteurs duaux des précédents	fonctions expressibles	
$(x,y) \rightarrow x \oplus y$	linéaire	uniforme sur les	[BP05]
$(x,y) \rightarrow \bar{x} \oplus y$	conn. dual	fct. expressibles	
$(x,y) \rightarrow x \vee y$	non linéaire, non équilibré, monotone	concentrée sur	[BP05]
$(x,y) \rightarrow \bar{x} \wedge \bar{y}$	conn. dual	une fonction seuil	
$(x,y) \rightarrow x \wedge y$	non linéaire, non équilibré, monotone	concentrée sur	[BP05]
$(x,y) \rightarrow \bar{x} \vee \bar{y}$	conn. dual	une fonction seuil	
$(x,y) \rightarrow \bar{x} \vee y$ $(x,y) \rightarrow x \vee \bar{y}$	non linéaires, non équilibrés, non monotones	?	
$(x,y) \rightarrow \bar{x} \wedge y$ $(x,y) \rightarrow x \wedge \bar{y}$	conn. duaux des précédents	?	

FIG. 10.1 – Tableau présentant les distributions limites d'arbres équilibrés.

Nous allons dans un premier temps étudier une amplification basée sur le connecteur binaire implication, le seul binaire non encore étudié. Puis nous analyserons le comportement d'une amplification dont l'ensemble de connecteurs comporte deux éléments aléatoires : pour le moment, toutes les études concernant l'amplification n'ont pris en compte qu'un unique connecteur. Nous baserons cette partie sur les deux connecteurs binaires *et* et *ou*.

Chapitre 11

Système de l'implication

Soit $k > 1$ – le cas $k = 1$ est trivial. Le système que nous étudions est basé sur $H_0 = \{x_1, \ldots, x_k\}$ avec μ_0 uniforme sur H_0. Dans cette section, nous considérons que l'ensemble de connecteurs $\mathcal{C}$ est réduit à $\{\rightarrow\}$. La méthode de Valiant permet d'obtenir aisément le résultat. Nous appellerons une affectation $a = (a_1, \ldots, a_k)$ des k variables, un *point* a de $\{0,1\}^k$.

Le *poids* d'un point est le réel : $\omega(a) = \mu_0(x_1).a_1 + \ldots + \mu_0(x_k).a_k$. Il est clair que $\omega(a) \in [0,1]$.

Prouvons le lemme d'analyse réelle suivant :

Lemme 37 *Soient (u_n) et (v_n) les deux suites récurrentes suivantes :*

$$\left\{ \begin{array}{l} u_0 \in]0,1[, \\ u_{n+1} = 1 - u_n + u_n^2. \end{array} \right. \qquad \left\{ \begin{array}{l} v_0 \in]0,1[, \\ v_{n+1} = v_n - v_n^2. \end{array} \right.$$

Elles vérifient :

$$1 - u_n \sim \frac{1}{n} \qquad et \qquad v_n \sim \frac{1}{n}.$$

Preuve: Commençons par étudier (v_n). L'équation de récurrence et la valeur initiale prouvent la décroissance de la suite. Une récurrence sur n permet de prouver que pour tout $n \in \mathbb{N}$, $v_n \in]0,1[$. La suite (réelle) étant bornée, elle converge vers ℓ satisfaisant : $\ell^2 = 0$. On en conclut que (v_n) admet pour limite 0. Pour tout n, $v_n > 0$ donc posons $w_n = 1/v_n$. L'équation de récurrence sur v_n se traduit par $w_{n+1} = w_n + 1 + 1/(w_n - 1)$, par ailleurs $w_0 \in]0,1[$. Par récurrence nous obtenons $n \leqslant w_n \leqslant n + \ln(n-2) + w_2$. Nous en concluons donc $v_n \sim 1/n$. Passons à la suite (u_n). En posant $u'_n = 1 - u_n$, nous obtenons une suite (u'_n) telle que $u'_0 \in]0,1[$ et vérifiant la même équation de récurrence que la suite (v_n). Par conséquent, $(1 - u_n)$ converge vers 0 et est équivalente à $1/n$. □

Proposition 17 *Soit $k \geqslant 1$ et μ_0 une distribution uniforme sur $H_0 = \{x_1, \ldots, x_k\}$. La suite (π_n) admet une distribution limite concentrée sur la fonction* Vrai.

Preuve: Pour une affectation $a \in \{0,1\}^k$ des variables, définissons la suite (u_n) telle que :

$$u_n = \mathbb{P}_{f \sim \pi_n}[f(a)] = 1.$$

Par définition de π_n, u_n est la probabilité qu'une expression $F \in H_n$, choisie aléatoirement selon μ_n, calcule 1 pour l'affectation a. Notons $a = (a_1, \ldots, a_k)$. Le terme initial vérifie :

$u_0 = \sum_{i=1}^{k} a_i.\mu_0(x_i) = \omega(a)$. En étudiant la racine et les deux fils d'une expression $F \sim \mu_{n+1}$ (F choisie selon la distribution μ_{n+1}), on obtient une équation de récurrence :

$$u_{n+1} = 1 - u_n + u_n^2.$$

En effet, afin que l'expression F calcule 1 pour l'affectation a, soit son fils gauche calcule 0 (et le fils droit est quelconque), soit ses deux fils calculent 1. Si $a \in \{(0,\ldots,0),(1,\ldots,1)\}$ alors pour tout $n \geqslant 1$, on a $u_n = 1$. Sinon, $u_0 = \omega(a) \in]0,1[$ et d'après le Lemme 37, la suite (u_n) converge aussi vers 1. Ceci étant vrai pour tout point a, la limite π de (π_n) existe et est concentrée sur la fonction *Vrai*. □

Intéressons-nous désormais à la vitesse de convergence vers la distribution limite. Nous allons l'étudier par rapport au nombre d'itérations pour un ensemble fixe de variables et des valeurs initiales p et μ_0 données.

Soit

$$\|\pi_n - \pi\| = \max_{f \in \mathcal{B}_k} |\pi_n(f) - \pi(f)|.$$

Si $\|\pi_n - \pi\| = \Theta(1/n)$, nous dirons que la convergence est *logarithmique*.

Proposition 18 *La vitesse de convergence de (π_n) est logarithmique.*

Preuve: Soit $f \in \mathcal{B}_k \setminus \{Vrai\}$ une fonction constructible dans le système – si f n'est calculée par aucune expression du système, alors $\pi_n(f) = 0$ pour tout n. Soit $a \in \{0,1\}^k$ tel que $f(a) = 0$. Le point a est distinct de $(1,\ldots,1)$ car toute fonction du système calcule 1 en le point $(1,\ldots,1)$. Définissons la suite (v_n) telle que : pour tout $n \in \mathbb{N}$,

$$v_n = \mathbb{P}_{f\sim\pi_n}[f(a) = 0].$$

On a $0 \leqslant \pi_n(f) \leqslant v_n$, puisque $f(a) = 0$. Si $a = (0,\ldots,0)$ alors pour tout n on a $v_n = 0$. Sinon, $\omega(a) \in]0,1[$ et pour tout $n \geqslant 1$, on a $v_{n+1} = v_n - v_n^2$. D'après le Lemme 37, on en conclut :

$$0 \leqslant \pi_n(f) \leqslant \frac{1}{n}.$$

Soit $g = Vrai$, et $a \in \{0,1\}^k \setminus \{(0,\ldots,0),(1,\ldots,1)\}$. Nous avons $\pi_n(g) = 1 - \sum_{f\neq g} \pi_n(f)$. Or le nombre de fonctions différentes de g est borné indépendamment de n, donc

$$1 - \pi_n(g) = |\pi(g) - \pi_n(g)| \leqslant \frac{|\mathcal{B}_k|}{n}.$$

Ainsi, nous obtenons $||\pi - \pi_n|| = O(1/n)$.

Soit $g = Vrai$ et $a \in \{0,1\}^k \setminus \{(0,\ldots,0),(1,\ldots,1)\}$. Du fait que $g(a) = 1$, on a :

$$\pi_n(g) \leqslant \mathbb{P}_{f\sim\pi_n}[f(a) = 1] = 1 - \mathbb{P}_{f\sim\pi_n}[f(a) = 0].$$

Définissons la suite : $v_0 = \omega(a)$ et

$$v_n = \mathbb{P}_{f\sim\pi_n}[f(a) = 0].$$

Or $\omega(a) \in]0,1[$ et

$$v_n \leqslant 1 - \pi_n(g) = \pi(g) - \pi_n(g).$$

Or $v_n \leqslant 1/n$, ce qui nous permet de conclure $||\pi - \pi_n|| = \Omega(1/n)$. □

Nous remarquons que la distribution limite est concentrée sur une unique fonction alors que, dans le cas non équilibré, le support de la distribution limite contenait toutes les fonctions du système. Ce résultat est intéressant mais pas surprenant du fait que nous avions remarqué que pour une taille donnée, le nombre d'arbres équilibrés est très faible par rapport au nombre d'arbres quelconques. On comprend donc bien que les deux distributions limites peuvent très différentes.

Nous sommes désormais en mesure de présenter le tableau 10.1 [p. 105] complété. Quel que soit le connecteur binaire utilisé, nous savons caractériser la distribution limite engendrée par une amplification basée sur ce connecteur.

connecteur	propriétés du connecteur	distribution limite	preuve
$(x,y) \to x$ $(x,y) \to y$ $(x,y) \to 1$	linéaires	uniforme sur les	Brodsky et Pippenger [BP05]
$(x,y) \to \bar{x}$ $(x,y) \to \bar{y}$ $(x,y) \to 0$	connecteurs duaux des précédents	fonctions expressibles	
$(x,y) \to x \oplus y$	linéaire	uniforme sur les	[BP05]
$(x,y) \to \bar{x} \oplus y$	conn. dual	fct. expressibles	
$(x,y) \to x \vee y$	non linéaire, non équilibré, monotone	concentrée sur	[BP05]
$(x,y) \to \bar{x} \wedge \bar{y}$	conn. dual	une fonction seuil	
$(x,y) \to x \wedge y$	non linéaire, non équilibré, monotone	concentrée sur	[BP05]
$(x,y) \to \bar{x} \vee \bar{y}$	conn. dual	une fonction seuil	
$(x,y) \to \bar{x} \vee y$ $(x,y) \to x \vee \bar{y}$	non linéaires, non équilibrés, non monotones	concentrée sur la fonction *Vrai*	chapitre 11
$(x,y) \to \bar{x} \wedge y$ $(x,y) \to x \wedge \bar{y}$	conn. duaux des précédents	concentrée sur la fonction *Faux*	

FIG. 11.1 – Tableau présentant les distributions limites d'arbres équilibrés.

Maintenant que le comportement des systèmes construits avec un connecteur binaire est connu, nous allons envisager dans le chapitre suivant le cas de plusieurs connecteurs aléatoires.

En considérant l'ensemble de connecteurs Et/Ou *aléatoires, y-a-t-il une différence flagrante entre les distributions obtenues via l'amplification et via le modèle des grands arbres (cf. [CFGG04, Koz08]) ?*

Chapitre 12

Arbres *Et/Ou* équilibrés

Considérons désormais les arbres *Et/Ou* équilibrés. Rappelons simplement que dans le cas non équilibré, toutes les fonctions ont une probabilité non nulle. Afin de construire nos arbres, dans un premier temps, nous étiquetterons indépendamment les feuilles avec un des k littéraux positifs, puis nous verrons que le modèle autorisant les littéraux négatifs est une petite extension de ce modèle. De plus, les littéraux seront choisis d'après une distribution de probabilité μ_0. Dans un second temps, au lieu de considérer les arbres dont les connecteurs sont choisis de manière uniforme entre *Et* et *Ou*, notre modèle consistera à choisir les connecteurs indépendamment et d'après une distribution de Bernoulli. Il en sera de même pour les littéraux – et en plus nous autoriserons les littéraux négatifs.

12.1 Arbres *Et/Ou* équilibrés : distribution limite

Soient $k > 0$, μ_0 une distribution sur $H_0 = \{x_1, \ldots, x_k\}$ et l'ensemble de connecteurs $\mathcal{C} = \{\wedge, \vee\}$ avec $p \in [0,1]$ la probabilité associée à $\wedge$. Pour tout $n \geqslant 1$ posons $H_n = \{h_1 \wedge h_2, h_1 \vee h_2 \mid h_1, h_2 \in H_{n-1}\}$ et μ_n une distribution de probabilité sur H_n vérifiant : $\mu_n(h_1 \wedge h_2) = p\mu_{n-1}(h_1)\mu_{n-1}(h_2)$, et $\mu_n(h_1 \vee h_2) = (1-p)\mu_{n-1}(h_1)\mu_{n-1}(h_2)$. Notons que si μ_0 est uniforme et $p = 1/2$, alors $\pi_n(f)$ est la proportion d'arbres de H_n qui calculent f.

Avant d'énoncer le résultat, donnons quelques définitions. Rappelons tout d'abord qu'une affectation $a = (a_1, \ldots, a_k)$ des k variables, est nommé un *point* a de $\{0,1\}^k$. Le *poids* d'un point est le réel : $\omega(a) = \mu_0(x_1).a_1 + \ldots + \mu_0(x_k).a_k$. Il est clair que $\omega(a) \in [0,1]$. Par ailleurs nous noterons $\prec$ l'ordre usuel partiel strict sur $\{0,1\}^k$; ainsi on a, $(a_1, \ldots, a_k) \prec (b_1, \ldots, b_k)$, si pour tout i, $a_i \leqslant b_i$ et $a \neq b$. Etant donné deux points $a, b \in \{0,1\}^k$, la plus grande borne inférieure à $\{a, b\}$ est notée par $inf(a, b)$. Enfin, étant donnés $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ et $\theta \in \mathbb{R}$, la *fonction seuil linéaire* $T_{\alpha,\theta}$ est la fonction Booléenne sur $\{0,1\}^k$ définie par :

$$T_{\alpha,\theta}(a) = 1 \;\Leftrightarrow\; \alpha_1.a_1 + \ldots + \alpha_k.a_k \geqslant \theta.$$

Nous sommes désormais prêts à énoncer notre résultat.

Théorème 6 *Soient $k \geqslant 1$, $p \in [0,1]$, et μ_0 une distribution de probabilité sur $\{x_1, \ldots, x_k\}$ telle que pour tout i, $\mu_0(x_i) > 0$. La suite de distributions (π_n) admet la distribution limite suivante :*

- *Supposons $p > 1/2$. Le support de la distribution limite est restreint à la fonction $x_1 \wedge \ldots \wedge x_k$. De façon duale, si on a $p < 1/2$, alors le support de la distribution limite est restreint à la fonction $x_1 \vee \ldots \vee x_k$.*

– *Si $\wedge$ et $\vee$ ont la même probabilité ($p = 1/2$), alors la distribution limite de probabilité est concentrée sur l'ensemble des fonctions seuil linéaire de la forme $\{T_{\mu_0,\theta} \mid \theta \in \mathbb{R}\}$ de la manière suivante : elle est égale à la loi de $T_{\mu_0,\theta}$ où θ est une variable aléatoire uniforme sur $[0,1]$. Voici une manière équivalente de décrire π : Soient $\theta_0 = 0 < \theta_1 < \ldots < \theta_s = 1$ les différents poids de tous les points de $\{0,1\}^k$, pour $i \in \{1,\ldots,s\}$, $\pi(T_{\mu_0,\theta_i}) = \theta_i - \theta_{i-1}$.*

La distribution limite, dans le cas $p = 1/2$, a une interprétation géométrique naturelle. Soient $h_0, h_1, \ldots, h_s$ les différents hyperplans affines réels, normaux à μ_0 et intersectant le cube unité $\{0,1\}^k$, et tels que h_{i+1} soit juste au dessus de h_i. Remarquons que chaque hyperplan ne contenant pas $(0,\ldots,0)$ définit une fonction linéaire seuil de la manière suivante : les points de $\{0,1\}^k$ appartenant au demi-espace ouvert contenant $(0,\ldots,0)$ évaluent la fonction à 0, tandis que les autres l'évaluent à 1. La distribution limite π est concentrée sur les fonctions définies par les hyperplans $h_1,\ldots,h_s$; de plus, la probabilité de la fonction définie par h_i est proportionnelle à la distance euclidienne $\mathrm{d}(h_i, h_{i-1})$ entre h_i et h_{i-1} – i.e. elle est égale à $\mathrm{d}(h_i, h_{i-1})/\mathrm{d}(h_0, h_s)$. Ceci est illustré par un exemple avec deux variables dans la Figure 12.1.

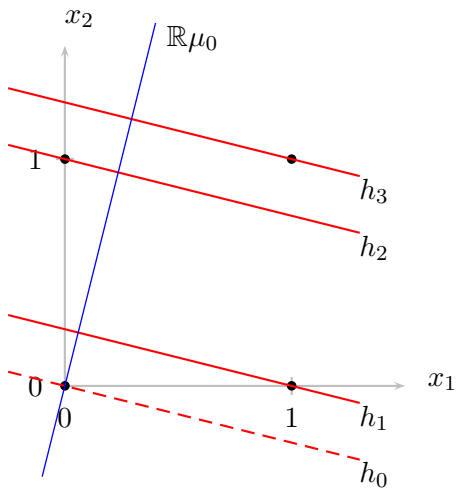

FIG. 12.1 – Les hyperplans définissant les fonctions linéaires seuil pour la distribution limite de (π_n) pour les paramètres $p = \frac{1}{2}$, $\mu_0(x_1) = \frac{1}{5}$ et $\mu_0(x_2) = \frac{4}{5}$.

Le cas uniforme est une conséquence directe du Théorème 6 :

Corollaire 5 *Pour $p = 1/2$ et μ_0 uniforme sur $H_0 = \{x_1,\ldots,x_k\}$, la suite (π_n) a une distribution limite qui est uniforme sur les k fonctions seuil $x_1 + \ldots + x_k \geqslant i$ pour $i \in \{1,\ldots,k\}$.*

Une extension facile du Théorème 6 consiste à considérer les symboles étiquetant des feuilles et permettant de calculer toutes les fonctions Booléennes. En autorisant les littéraux négatifs, la suite de probabilités induites sur les fonctions Booléennes par les arbres *Et/Ou* équilibrés de profondeur croissante, admet également une distribution limite, qui est obtenue en faisant des remplacements de variables, puis des fusions des fonctions identiques dans la distribution limite décrite par le théorème. Voilà un cas particulier :

Corollaire 6 *Pour $p = 1/2$, et μ_0 uniforme sur $H_0 = \{x_1, \overline{x}_1, \ldots, x_k, \overline{x}_k\}$, la suite π_n a une distribution limite qui est uniforme sur les deux fonctions constantes* Vrai *et* Faux.

Preuve: Considérons π'_0 uniforme sur $H'_0 = \{x_1, y_1, \ldots, x_k, y_k\}$. En utilisant le Corollaire 5, on conclut que la suite (π'_n) induite sur les fonctions Booléennes admet une distribution limite uniforme sur les $2k$ fonctions seuil : $x_1 + y_1 + \ldots + x_k + y_k \geqslant i$ pour $i \in \{1, \ldots, 2k\}$. En remplaçant y_i par $\overline{x}_i$ dans ces fonctions, on obtient une distribution limite de la suite (π_n) basée sur la distribution initiale μ_0 sur H_0. Cette distribution limite est uniforme sur les $2k$ fonctions $x_1, \overline{x}_1 + \ldots + x_k + \overline{x}_k \geqslant i$, $i \in \{1, \ldots, 2k\}$. Cependant, on remarque que $x_j + \overline{x}_j$ vaut toujours 1, par conséquent les fonctions précédentes sont exactement les deux fonctions constantes *Vrai* – pour $i \in \{1, \ldots, k\}$) et *Faux* (pour $i \in \{k+1, \ldots, 2k\}$. □

Il est intéressant de remarquer que la distribution limite est concentrée sur un petit nombre de fonctions (au plus 2^k). En particulier ce nombre de fonctions est 2 (resp. k) dans le cas uniforme avec les littéraux positifs et négatifs (resp. uniquement les littéraux positifs). Ce qui est à comparer avec le modèle non équilibré d'arbres qui pondère toutes les fonctions (resp. toutes les fonctions monotones).

12.2 Etude de la convergence vers une distribution limite

Dans toute cette partie nous supposons que la distribution μ_0 sur $H_0 = \{x_1, \ldots, x_k\}$ satisfait $\mu_0(x_i) > 0$ pour tout $i \in \{1, \ldots, k\}$.

Remarquons que pour une expression F calculant une fonction f, en substituant le connecteur $\wedge$ par $\vee$ et inversement, nous obtenons une expression F' calculant le dual f'de f, défini par $f'(x_1, \ldots, x_k) = \overline{f}(\overline{x}_1, \ldots, \overline{x}_k)$. Par conséquent, les cas p et $1 - p$ sont duaux ; ainsi nous ne considérerons que le cas $p \geqslant 1/2$ dans les preuves.

Nous allons nous intéresser tout d'abord une distribution non uniforme sur les connecteurs : Proposition 19, puis le cas d'une distribution uniforme sur les connecteurs : Proposition 20. Le Théorème 6 sera une conséquence de ces deux propositions et de la remarque précédente.

Lemme 38 *Soit $p \in [0, 1]$. Pour un point $a \in \{0, 1\}^k$,définissons la suite suivante :*

$$u_n = \mathbb{P}_{f \sim \pi_n} [f(a) = 1] . \tag{12.1}$$

La suite (u_n) vérifie :

$$\begin{cases} u_0 = \omega(a) \\ u_{n+1} = (2p-1)u_n^2 + 2(1-p)u_n. \end{cases}$$

En particulier :
- *Pour $p > 1/2$ et $a \neq (1, \ldots, 1)$, $u_n \to 0$;*
- *Pour $p = 1/2$, la suite (u_n) est constante égale à $\omega(a)$.*

Preuve: Suivant la définition de π_n, u_n est la probabilité qu'une expression $F \in H_n$, choisie aléatoirement selon μ_n, s'évalue à 1 au point a. Supposons $a = (a_1, \ldots, a_k)$ avec la distribution initiale μ_0 alors $u_0 = \sum_{i=1}^k a_i.\mu_0(x_i) = \omega(a)$. L'équation de récurrence sur (u_n) est obtenue en étudiant l'étiquette de la racine d'une expression $F \sim \mu_{n+1}$, (i.e. $F \in H_{n+1}$ choisie aléatoirement selon la distribution μ_{n+1}) : F s'évalue à 1 en a si et seulement si la racine de

F est étiquetée par $\wedge$ et ses deux fils s'évaluent à 1 pour a ; ou si la racine est étiquetée par $\vee$ et les deux fils ne s'évaluent pas simultanément à 1 en a. Nous obtenons l'équation suivante.

$$u_{n+1} = p\ u_n^2 + (1-p)\ (1-(1-u_n)^2).$$

Une simplification de celle-ci donne l'équation de récurrence énoncée dans le lemme.

Supposons désormais que $p > 1/2$ et $a \neq (1,\ldots,1)$. D'après l'hypothèse sur μ_0,nous savons que $u_0 = \omega(a) < 1$. L'étude du graphe de la fonction réelle $x \mapsto (2p-1)x^2 + 2(1-p)x$ montre que $u_n \to\ 0$. Le cas $p = 1/2$ est immédiat. □

Considérons tout d'abord une distribution non uniforme sur les connecteurs.

Proposition 19 *Soit $p > 1/2$, la suite (π_n) a une limite concentrée sur la fonction $x_1 \wedge \ldots \wedge x_k$.*

Preuve: Notons que toutes les expressions, que nous construisons, calculent des fonctions monotones non constantes. Ainsi, chaque expression construite dans le système s'évalue à 1 au point $(1,\ldots,1)$. Soit $a \in \{0,1\}^k \setminus (1,\ldots,1)$. D'après le lemme 38, nous savons que

$$\mathbb{P}_{f\sim\pi_n}[f(a) = 1] \to\ 0.$$

Ainsi, toute fonction f distincte de $x_1 \wedge \ldots \wedge x_k$ satisfait $\pi_n(f) \to\ 0$. En conclusion, $\pi_n(x_1 \wedge \ldots \wedge x_k) \to\ 1$. □

Considérons désormais une distribution uniforme sur les connecteurs.

Lemme 39 *Soient $p = 1/2$, et $a, b \in \{0,1\}^k$ deux points distincts tels que $\omega(a) \leqslant \omega(b)$. Soit*

$$v_n = \mathbb{P}_{f\sim\pi_n}[f(a) = 1 \text{ et } f(b) = 0]. \tag{12.2}$$

La suite (v_n) vérifie :

$$\begin{cases} v_0 = \omega(a) - \omega(\inf\{a,b\}) \\ v_{n+1} = v_n(1 - \omega(b) + \omega(a) - v_n). \end{cases}$$

En particulier, $v_n \to\ 0$. Par ailleurs, pour tout n :

$$\mathbb{P}_{f\sim\pi_n}[f(a) = 0 \text{ et } f(b) = 1] -\ \mathbb{P}_{f\sim\pi_n}[f(a) = 1 \text{ et } f(b) = 0] =\ \omega(b) - \omega(a).$$

Preuve: Soient $\alpha, \beta \in \{0,1\}$, et

$$v_n^{(\alpha,\beta)} = \mathbb{P}_{f\sim\pi_n}[f(a) = \alpha \text{ et } f(b) = \beta].$$

Remarquons que $v_n^{(\alpha,\beta)}$ est la probabilité qu'une expression $F \in H_n$, choisie aléatoirement selon μ_n, s'évalue à α au point a et à β au point b. Pour $a = (a_1,\ldots,a_k)$ et $b = (b_1,\ldots,b_k)$, on a

$$\begin{aligned} v_0^{(1,0)} &= \sum_{i=1}^{k} a_i(1-b_i)\ \mu_0(x_i) \\ &= \sum_{i=1}^{k} a_i\ \mu_0(x_i) - \sum_{i=1}^{k} a_i b_i\ \mu_0(x_i) \\ &= \omega(a) - \omega(\inf\{a,b\}). \end{aligned}$$

Les équations de récurrence des quatre suites $v_n^{(\alpha,\beta)}$ sont obtenues en étudiant le connecteur de la racine (soit $\wedge$ soit $\vee$) de l'expression $F \sim \mu_{n+1}$. Pour $v_{n+1}^{(0,0)}$, on obtient :

$$\begin{aligned} v_{n+1}^{(0,0)} &= \frac{1}{2}(v_n^{(0,0)})^2 \\ &+\frac{1}{2}\left((v_n^{(0,0)})^2 + 2v_n^{(0,0)}v_n^{(0,1)} + 2v_n^{(0,0)}v_n^{(1,0)}\right) \\ &+\frac{1}{2}\left(2v_n^{(0,0)}v_n^{(1,1)} + 2v_n^{(0,1)}v_n^{(1,0)}\right). \end{aligned}$$

D'une manière similaire, on obtient les trois autres équations. Par ailleurs, en utilisant le fait que $v_n^{(0,0)} + v_n^{(0,1)} + v_n^{(1,0)} + v_n^{(1,1)} = 1$ pour tout n, le système se simplifie comme suit :

$$\left\{\begin{array}{l} v_{n+1}^{(0,0)} = v_n^{(0,0)} + v_n^{(1,0)}v_n^{(0,1)} \\ v_{n+1}^{(1,1)} = v_n^{(1,1)} + v_n^{(1,0)}v_n^{(0,1)} \\ v_{n+1}^{(0,1)} = v_n^{(0,1)} - v_n^{(1,0)}v_n^{(0,1)} \\ v_{n+1}^{(1,0)} = v_n^{(1,0)} - v_n^{(1,0)}v_n^{(0,1)} \end{array}\right. \tag{12.3}$$

Bien entendu, $v_n^{(\alpha,\beta)} \in [0,1]$ pour tout n. En conséquence, d'après les équations (12.3), les suites $(v_n^{(0,0)})$ et $(v_n^{(1,1)})$ sont croissantes, du fait qu'elles sont bornées par 1, elles convergent. De manière similaire, les deux suites $(v_n^{(0,1)})$ et $(v_n^{(1,0)})$ convergent puisqu'elles sont décroissantes et bornées par 0. Pour $\alpha, \beta \in \{0,1\}$, soient $\ell^{(\alpha,\beta)} = \lim v_n^{(\alpha,\beta)}$. En faisant tendre n vers l'infini dans les équations (12.3), on obtient : $\ell^{(1,0)} \cdot \ell^{(0,1)} = 0$.

Désormais, en soustrayant les deux dernières équations du système (12.3), nous notons que la suite $(v_n^{(0,1)} - v_n^{(1,0)})$ est constante ; elle est égale à $v_0^{(0,1)} - v_0^{(1,0)} = (\omega(b) - \omega(\inf\{a,b\})) - (\omega(a) - \omega(\inf\{a,b\})) = \omega(b) - \omega(a)$. Il s'en suit que $\ell^{(0,1)} - \ell^{(1,0)} = \omega(b) - \omega(a) \geqslant 0$. Or rappelons que $\ell^{(1,0)}.\ell^{(0,1)} = 0$; nous obtenons donc $\ell^{(1,0)} = 0$, i.e. $v_n^{(1,0)} \to 0$. □

Proposition 20 *Pour $p = 1/2$, la suite (π_n) admet une distribution limite, qui est la loi de $T_{\theta,U}$, où U est uniforme dans $[0,1]$.*

Preuve: Soient deux points distincts $a, b \in \{0,1\}^k$ tels que $\omega(a) \leqslant \omega(b)$. D'après le Lemme 39, nous savons que

$$\mathbb{P}_{f\sim\pi_n}[f(a) = 1 \text{ et } f(b) = 0] \to 0.$$

Par conséquent, toute fonction f satisfait $\pi_n(f) \to 0$ si elle ne remplit pas la condition suivante :

$$\text{Pour tout } a, b \in \{0,1\}^k,\ \omega(a) \leqslant w(b) \Rightarrow\ f(a) \leqslant f(b).$$

Remarquons que les fonctions satisfaisant la condition précédente sont exactement les fonctions seuil linéaire de la forme $T_{\mu_0,\theta}$ pour $\theta \in \mathbb{R}$.

Soient $\theta_0 = 0 < \theta_1 < \theta_2 < \ldots < \theta_s = 1$ les différents poids de tous les points de $\{0,1\}^k$. Pour $0 \leqslant i \leqslant s$, soit $a_i \in \{0,1\}^k$ un point de poids θ_i. Soit $j \in \{0,\ldots,s\}$; d'après le Lemme 38, nous savons que

$$\mathbb{P}_{f\sim\pi_n}[f(a_j) = 1] = \omega(a_j) = \theta_j,$$

pour tout n. Ainsi il s'en suit que

$$\pi_n(T_{\mu_0,\theta_0}) + \ldots + \pi_n(T_{\mu_0,\theta_j}) \to \theta_j.$$

Par récurrence sur j, pour tout $j \in \{1,\ldots,s\}$, nous avons $\pi_n(T_{\mu_0,\theta_j}) \to \theta_j - \theta_{j-1}$. Ceci termine la démonstration puisque $\sum_{j=1}^{s}(\theta_j - \theta_{j-1}) = 1$. □

12.3 Analyse de la vitesse de convergence

Rappelons que nous étudions la vitesse de convergence vers la distribution limite par rapport au nombre d'itérations pour un ensemble de variables fixe et des valeurs initiales p et μ_0 fixés.

Soit

$$\|\pi_n - \pi\| = \max_{f \in \mathcal{B}_k} |\pi_n(f) - \pi(f)|.$$

Si $\|\pi_n - \pi\| = 2^{-\Theta(n)}$, nous dirons que la convergence est *linéaire*, et elle sera dite *logarithmique* si $\|\pi_n - \pi\| = \Theta(1/n)$. Le système étudié ne présente pas de vitesse plus rapide que linéaire dans les cas non triviaux.

A nouveau nous supposons que μ_0 satisfait $\mu_0(x_i) > 0$ pour tout $i \in \{1, \ldots, k\}$. Remarquons que dans le cas $k = 1$, toutes les expressions construites calculent toujours la même fonction x_1. Donc supposons que nous soyons dans le cas $k > 1$ pour le reste de l'étude. Le cas $p \in \{0, 1\}$ correspond correspond à l'utilisation d'un unique connecteur ; on montre aisément que $\|\pi_n - \pi\| = 2^{-\Theta(2^n)}$ dans ce cas. Enfin, rappelons que nous ne traiterons que le cas $p \geqslant 1/2$ par dualité. Nous utiliserons le fait suivant, un corollaire du théorème des accroissements finis dont la démonstration se trouve dans tout livre d'analyse réelle :

Fait 4 *Soit $f : [a, b] \to [a, b]$ une fonction réelle de classe C^2. Supposons qu'il existe $c < 1$ tel que $|f'(x)| < c$ pour tout $x \in [a, b]$. Soit $x_0 \in [a, b]$. La suite (x_n) définie par $x_{n+1} = f(x_n)$ converge vers l'unique point fixe ℓ de f. En outre, si $f'(\ell) \neq 0$ et $x_0 \neq \ell$, alors il existe $\lambda \neq 0$ tel que $x_n - \ell \sim \lambda f'(\ell)^n$.*

Dans un premier temps, nous considérons une distribution non uniforme sur les connecteurs.

Lemme 40 *Soient $1/2 < p < 1$ et $a \in \{0, 1\}^k \setminus \{(0, \ldots, 0), (1, \ldots, 1)\}$. La suite (u_n) définie par (12.1) satisfait $u_n = \Theta\left((2 - 2p)^n\right)$.*

Preuve: D'après les hypothèses sur μ_0, nous avons $0 < \omega(a) < 1$. Le résultat est obtenu à partir du Fait 4, en utilisant l'équation de récurrence énoncée dans le Lemme 38. □

Proposition 21 *Pour $1/2 < p < 1$, la vitesse de convergence de (π_n) est linéaire.*

Preuve: Pour les fonctions constantes f, nous savons déjà que $\pi_n(f) = 0$ pour tout n. Soit f une fonction non constante et différente de $x_1 \wedge \ldots \wedge x_k$. D'après le Théorème 6, $\pi(f) = 0$. Soit $a \neq (1, \ldots, 1)$ tel que $f(a) = 1$. Bien évidemment, $0 \leqslant \pi_n(f) \leqslant \mathbb{P}_{f \sim \pi_n}[f(a) = 1]$ puisque $f(a) = 1$. Le Lemme 40 prouve que

$$\mathbb{P}_{f \sim \pi_n}[f(a) = 1] = O((2 - 2p)^n).$$

Il s'en suit : $|\pi_n(f) - \pi(f)| = 2^{-O(n)}$.

Intéressons-nous désormais à la fonction g calculée par $x_1 \wedge \ldots \wedge x_k$. Bien sûr, $\pi_n(g) = 1 - \sum_{f \neq g} \pi_n(f)$. Du fait qu'il n'y ait qu'un nombre fini de fonctions et indépendant de n, il existe une constante $C > 0$ (indépendante de n et f) telle que $\pi_n(f) \leqslant 2^{-C.n}$ pour tout $f \neq g$. En utilisant la première partie de la preuve, nous obtenons $1 - \pi_n(g) \leqslant |\mathcal{B}_k| \cdot 2^{-C.n}$.

Nous savons du Théorème 6 que $\pi(g) = 1$. Ceci implique $|\pi_n(g) - \pi(g)| = 2^{-O(n)}$. Ainsi, en utilisant tout cela, nous montrons que $\|\pi_n - \pi\| = 2^{-O(n)}$.

Soit $a \in \{0,1\}^k \setminus \{(0,\dots,0),(1,\dots,1)\}$ – il existe puisque $k > 1$. D'après l'hypothèse sur μ_0, nous avons $0 < \omega(a) < 1$. Du fait que $g(a) = 0$, nous avons $\pi_n(g) \leqslant \mathbb{P}_{f\sim\pi_n}[f(a) = 0] = 1 - \mathbb{P}_{f\sim\pi_n}[f(a) = 1]$. Rappelons désormais le Théorème 6 qui prouve $\pi(g) = 1$. Ainsi nous avons obtenu

$$\mathbb{P}_{f\sim\pi_n}[f(a) = 1] \leqslant \pi(g) - \pi_n(g).$$

D'après le Lemme 40, nous savons que $\mathbb{P}_{f\sim\pi_n}[f(a) = 1] = 2^{-\Omega(n)}$. Par conséquent nous avons démontré que $|\pi_n(g) - \pi(g)| = 2^{-\Omega(n)}$. La borne inférieure de $\|\pi_n - \pi\|$ est donc prouvée. □

Intéressons-nous désormais au cas d'une distribution uniforme sur les connecteurs.

Lemme 41 *Soient $p = 1/2$ et $a, b \in \{0,1\}^k$ des points distincts tels que $\omega(a) \leqslant \omega(b)$. Le comportement asymptotique de (v_n), définie par (12.2), est le suivant :*

- *Si $a \prec b$, alors (v_n) est constante, égale à 0 ;*
- *Si $\omega(a) = \omega(b)$, alors $v_n \sim 1/n$;*
- *Dans les autres cas, $v_n = \Theta((1 - \omega(b) + \omega(a))^n)$.*

Preuve: Si $a \prec b$, bien évidemment $v_n = 0$ pour tout n, car notre système ne construit que des fonctions monotones.

Supposons $\omega(a) = \omega(b)$. D'après le Lemme 39, on a $v_0 = \omega(a) - \omega(\inf\{a,b\})$. Puisque $a \neq b$ et $\omega(a) \leqslant \omega(b)$, nous avons $a \neq (1,\dots,1)$; d'après l'hypothèse sur μ_0 il s'en suit que $\omega(a) < 1$. Ainsi $v_0 < 1$. Du fait que $a \not\prec b$, nous avons $\inf\{a,b\} \prec a$. L'hypothèse sur μ_0 donne $\omega(\inf\{a,b\}) < \omega(a)$. En conséquence, $0 < v_0 < 1$. Le Lemme 39 donne l'équation $v_{n+1} = v_n(1 - v_n)$. D'après le Lemme 37, $v_n \sim 1/n$.

Intéressons-nous désormais au dernier cas : supposons $a \not\prec b$ et $\omega(a) < \omega(b)$. Le terme initial vérifie $v_0 = \omega(a) - \omega(\inf\{a,b\}) \in]0, 1 - \omega(b) + \omega(a)[$. D'après l'équation de récurrence sur (v_n) obtenue dans le Lemme 39 et le Fait 4, nous concluons que $v_n = \Theta((1-\omega(b)+\omega(a))^n)$. □

Proposition 22 *Pour $p = 1/2$, la vitesse de convergence de (π_n) est linéaire si tous les points de $\{0,1\}^k$ ont des poids distincts ; autrement elle est logarithmique.*

Preuve: Soit f une fonction Booléenne n'appartenant pas au support de π. Si f est non-monotone ou constante, alors $\pi_n(f) = 0$ pour tout n. Supposons pour la suite que ce n'est pas le cas. Alors il existe deux points $a, b \in \{0,1\}^k$ tels que $\omega(a) \leqslant \omega(b)$, $f(a) = 1$ et $f(b) = 0$. Evidemment,

$$0 \leqslant \pi_n(f) \leqslant \mathbb{P}_{g\sim\pi_n}[g(a) = 1 \text{ et } g(b) = 0],$$

puisque $f(a) = 1$ et $f(b) = 0$. Du fait du comportement asymptotique de $\mathbb{P}_{g\sim\pi_n}[g(a) = 1 \text{ et } g(b) = 0]$ prouvé dans le Lemme 41, nous obtenons $\pi_n(f) = O(1/n)$ si $\omega(a) = \omega(b)$, et $\pi_n(f) = O((1 - \omega(b) + \omega(a))^n)$ si $\omega(a) < \omega(b)$; i.e. $\pi_n(f) = 2^{-O(n)}$ dans ce dernier cas.

Désormais, soit f une fonction du support de π. Soient $a \in \{0,1\}^k$ un point de poids maximal dans $f^{-1}(0)$, et $b \in \{0,1\}^k$ un point de poids minimal dans $f^{-1}(1)$. D'après le Théorème 6, $f = T_{\mu_0,\omega(b)}$ et $\pi(f) = \omega(b) - \omega(a)$.

Soit $\mathcal{F}^{(0,1)}$ l'ensemble des fonctions s'évaluant à 0 en a et à 1 en b, et symétriquement, $\mathcal{F}^{(1,0)}$ l'ensemble des fonctions s'évaluant à 1 en a et à 0 en b, Il est clair que $\pi_n(\mathcal{F}^{(0,1)}) = \pi_n(f) + \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\})$. D'après le Lemme 39, pour tout n on a :

$$\pi_n(\mathcal{F}^{(0,1)}) - \pi_n(\mathcal{F}^{(1,0)}) = \omega(b) - \omega(a).$$

Or $\pi(f) = \omega(b) - \omega(a)$. Donc nous obtenons

$$\pi_n(f) - \pi(f) = \pi_n(\mathcal{F}^{(1,0)}) - \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\}).$$

Remarquons qu'aucune fonction de $\mathcal{F}^{(1,0)}$ n'appartient au support de π, car $\omega(a) < \omega(b)$. De la même manière, aucune fonction de $\mathcal{F}^{(0,1)} \setminus \{f\}$ n'appartient au support de π – puisque f est l'unique fonction seuil linéaire de la forme $T_{\mu_0,\cdot}$ telle que $f(a) = 0$ and $f(b) = 1$. D'après les bornes du premier paragraphe, nous avons $|\pi_n(f) - \pi(f)| = 2^{-O(n)}$ si tous les points ont des poids différents ; sinon nous avons $|\pi_n(f) - \pi(f)| = O(1/n)$. Finalement nous avons démontré que $\|\pi_n - \pi\| = 2^{-O(n)}$ lorsque tous les points ont des poids distincts ; et $\|\pi_n - \pi\| = O(1/n)$ dans le cas contraire.

Démontrons maintenant une borne inférieure de $\|\pi_n - \pi\|$. Supposons tout d'abord que tous les points ont des poids distincts. Puisque $k > 1$, il existe $a, b \in \{0,1\}^k$, tels que $0 < \omega(a) < \omega(b) < 1$, et $\{c \in \{0,1\}^k \mid \omega(a) < \omega(c) < \omega(b)\} = \emptyset$. Soit $\mathcal{F}^{(0,1)}$ l'ensemble des fonctions s'évaluant à 0 en a et à 1 en b, et $\mathcal{F}^{(1,0)}$ l'ensemble des fonctions s'évaluant à 1 en a et à 0 en b. Soit $f = T_{\mu 0,\omega(b)}$. Nous avons $\pi_n(\mathcal{F}^{(0,1)}) = \pi_n(f) + \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\})$. En réutilisant une des méthodes précédente : nous pouvons écrire

$$\begin{aligned}
\pi_n(\mathcal{F}^{(1,0)}) &= (\pi_n(f) - \pi(f)) + \pi_n(\mathcal{F}^{(0,1)} \setminus \{f\}) \\
&= \sum_{g \in \mathcal{F}^{(0,1)}} (\pi_n(g) - \pi(g)) \\
&\leqslant \sum_{g \in \mathcal{F}^{(0,1)}} |\pi_n(g) - \pi(g)|.
\end{aligned}$$

En conséquence $\|\pi_n - \pi\| \geqslant \pi_n(\mathcal{F}^{(1,0)})/|\mathcal{B}_k|$ pour tout n. D'après le comportement asymptotique de la suite $(\pi_n(\mathcal{F}^{(1,0)}))$ donné dans le Lemme 39 (dans le cas où tous les points ont des poids distincts), nous avons $\|\pi_n - \pi\| = 2^{-\Omega(n)}$.

Supposons désormais qu'il existe des points $a, b \in \{0,1\}^k$ de même poids. A nouveau soit $\mathcal{F}^{(1,0)}$ l'ensemble des fonctions s'évaluant à 0 en a et à 1 en b. Il est clair que

$$\pi_n(\mathcal{F}^{(1,0)}) = \sum_{f \in \mathcal{F}^{(1,0)}} \pi_n(f) \leqslant \sum_{f \in \mathcal{F}^{(1,0)}} |\pi_n(f) - \pi(f)|$$

puisque $\pi(f) = 0$ pour tout $f \in \mathcal{F}^{(1,0)}$ – toutes les fonctions du support de π sont constantes sur l'ensemble des points de même poids. Puisque le Lemme 39 prouve que $\pi_n(\mathcal{F}^{(1,0)}) \sim 1/n$ dans ce cas, nous avons démontré que $\|\pi_n - \pi\| = \Omega(1/n)$. □

Cinquième partie

Conclusion

Chapitre 13

Synthèse et perspectives

A travers cette thèse, j'ai étudié les trois points suivants. Tout d'abord, j'ai analysé deux distributions de probabilité sur les fonctions Booléennes, dans le système de l'implication. J'ai exhibé la structure de la plupart des expressions représentant une fonction Booléenne donnée – quand le nombre de variables tend vers l'infini. Dans ce contexte, ce résultat a permis de donner le terme principal du développement asymptotique de la probabilité de la fonction et j'ai mis en évidence le lien entre la probabilité et la complexité de la fonction Booléenne. Puis je me suis intéressé plus en détail à la fonction Booléenne *Vrai* qui caractérise la puissance de la logique de l'implication. L'analyse plus fine de la structure des tautologies (expressions représentant la fonction *Vrai*) m'a permis de comparer de façon quantitative la logique classique de l'implication et la logique intuitionniste de l'implication, plus faible. Cette comparaison a mis en évidence certaines propriétés d'une classe de tautologies. J'ai démontré, pour le système propositionnel complet, que ces propriétés sont conservées et par conséquent j'ai pu comparer quantitativement les deux logiques dans ce système. Enfin je me suis intéressé à la classe des expressions Booléennes équilibrées. L'intérêt pour cet ensemble d'expressions réside dans le fait qu'il permet la construction d'expressions de petites tailles représentant une fonction donnée. Le résultat concernant la distribution de probabilité sur les fonctions Booléennes engendrée par les expressions équilibrées dans le système de l'implication était le seul manquant pour les systèmes utilisant un unique connecteur logique binaire. Pour finir, puisque l'étude des expressions équilibrées construites avec un connecteur binaire aléatoire est exhaustive, je me suis concentré sur le système *Et/Ou* et j'ai démontré quelle est la distribution de probabilité engendrées sur les fonctions Booléennes.

Intéressons-nous aux perspectives possibles pour chacun des trois thèmes. Dans le système de l'implication et pour le modèle du processus de branchement, les fonctions *read-once* de complexité fixée nous permettent d'obtenir une partie significative de la probabilité totale. Ce n'est pas le cas pour le modèle des grands arbres. Afin de déterminer quels ensembles de fonctions sont les plus probables dans ce modèle, il nous faudra étudier en détail les probabilités des fonctions dont la complexité est grande, i.e. est dépendante du nombre de variables k. Des progrès dans cette direction nous permettraient de conclure quant à l'effet Shannon dans le modèle des grands arbres, mais également d'obtenir la complexité moyenne d'une fonction aléatoire pour chacun des deux modèles. Par ailleurs, les récents résultats de Kozik [Koz08] relatifs aux expressions *Et/Ou* aléatoires montrent que les arguments que nous avons mis en oeuvre dans le système de l'implication restent valables dans son système. Ceci suggère que certains systèmes propositionnels partagent le même comportement. Pourrions-

nous déterminer quels sont ces systèmes de connecteurs ?

Pour la fonction *Vrai*, nous avons pu transposer nos résultats concernant le système de l'implication au système propositionnel complet. Ceci provient du fait que plus que les connecteurs utilisés, la structure d'arbre est importante dans le cas des expressions Booléennes. Une question se pose : ces résultats sont-ils adaptables au système auquel on ajoute les connecteurs existentiel et universel ? Cependant ces ajouts nécessitent la notion de variable libre et liée dans l'arbre et compliquent en particulier les énumérations. Toutefois des progrès dans l'énumération des *lambda-termes* [DGK+09] (arbres ayant des variables liés) pourraient permettre de progresser dans cette direction.

En ce qui concerne les expressions équilibrées, l'étude que nous avons proposée établit les premiers résultats dans le cas où l'on utilise aléatoirement deux connecteurs. Une perspective serait de généraliser les résultats à tout ensemble de connecteurs. Un pas dans cette direction consisterait à s'intéresser aux ensembles de connecteurs monotones, pour lesquels les démonstrations devraient pouvoir s'étendre. Cependant, on n'a pas encore de résultat général pour un unique connecteur quelconque, donc le cas de plusieurs connecteurs quelconques n'est pas sans doute pas le plus facile à aborder. Par ailleurs, nous nous sommes aperçus que la vitesse de convergence était lente. Y aurait-il un système dans lequel nous obtiendrions une distribution sur les fonctions seuil linéaire via d'autres connecteurs et tel que la convergence serait plus rapide ? Enfin, pourrions-nous généraliser nos expressions *Et/Ou* Booléennes aléatoires au cas des arbres min-max sur les nombres réels ?

Index

Bibliographie

[AKS83] M. Ajtai, , J. Komlós, and E. Szemerédi. An O(n log n) sorting network. In *STOC '83 : Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 1–9, New York, NY, USA, 1983. ACM.

[AN72] K. Athreya and P. Ney. *Branching Processes*. Springer, 1972.

[Bop85] R. B. Boppana. Amplification of Probabilistic Boolean Formulas. In *Proceedings of the 26th IEEE Symposium on Foundations of Computer Science*, pages 20–29, 1985.

[BP05] A. Brodsky and N. Pippenger. The Boolean functions computed by random Boolean formulas or how to grow the right function. *Random Structures and Algorithms*, 27 :490–519, 2005.

[CFGG04] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger. And/Or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5) :475–497, July-September 2004.

[DFLS04] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13(4–5) :577–625, 2004. Special issue on Analysis of Algorithms.

[DGK+09] R. David, K. Grygiel, J. Kozik, C. Raffalli, G. Theyssier, and M. Zaionc. Some properties of random lambda terms. *arXiv :0903.5505*, 2009.

[DNR04] R. David, K. Nour, and C. Raffalli. *Introduction à la logique : Théorie de la démonstration, 2^e édition*. Dunod, 2004.

[Drm97] M. Drmota. Systems of functional equations. *Random Structures and Algorithms*, 10(1-2) :103–124, 1997.

[DZ97] M. Dubiner and U. Zwick. Amplification by read-once formulas. *SIAM Journal on Computing*, 26(1) :15–38, 1997.

[FGG09] H. Fournier, D. Gardy, and A. Genitrini. Balanced and/or trees and linear threshold functions. In *5th SIAM Workshop on Analytic and Combinatorics (ANALCO)*, pages 51–57, New York, USA, January 2009.

[FGGG08] H. Fournier, D. Gardy, A. Genitrini, and B. Gittenberger. Complexity and limiting ratio of Boolean functions over implication. In *33rd International Symposium on Mathematical Foundations of Computer Science (MFCS'08)*, pages 347–362, Torun, Pologne, August 2008.

[FGGZ07] H. Fournier, D. Gardy, A. Genitrini, and M. Zaionc. Classical and intuitionnistic logic are asymptotically identical. In Springer-Verlag, editor, *Annual Conference on Computer Science Logic (CSL'07)*, pages 177–193, Lausanne, Suisse, 2007.

[FS96] P. Flajolet and R. Sedgewick. *An introduction to the Analysis of Algorithms*. Addison-Wesley, 1996.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[Gar06] D. Gardy. Random Boolean expressions. In *Colloquium on Computational Logic and Applications*, volume AF, pages 1–36. DMTCS Proceedings, 2006.

[GK09] A. Genitrini and J. Kozik. Quantitative comparison of intuitionistic and classical logics - full propositional system. In *International Symposium on Logical Foundations of Computer Science (LFCS'09)*, Florida, USA, January 2009.

[GKM08] A. Genitrini, J. Kozik, and G. Matecki. On the density and the structure of the peirce-like formulae. In *5th Colloquium on Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probabilities*, Blaubeuren, Allemagne, September 2008.

[GKP89] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics : a foundation for computer science*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[GKZ07] A. Genitrini, J. Kozik, and M. Zaionc. Intuitionistic vs. classical tautologies, quantitative comparison. In *TYPES 07*, pages 100–109, Cividale del Friuli, Italie, 2007.

[GM97] A. Gupta and S. Mahajan. Using amplification to compute *majority* with small majority gates. *Computational Complexity*, 6(1) :46–63, 1997.

[Har65] M. A. Harrisson. *Introduction to Switching and Automata Theory*. McGraw Hill, 1965.

[Koz08] J. Kozik. Subcritical pattern languages for And/Or trees. In *Fifth Colloquium on Mathematics and Computer Science*, Blaubeuren, Germany, september 2008. DMTCS Proceedings.

[KZ04] Z. Kostrzycka and M. Zaionc. Statistics of intuitionnistic versus classical logic. *Studia Logica*, 76(3) :307–328, 2004.

[Lal93] S. P. Lalley. Finite range random walk on free groups and homogeneous trees. *The Annals of Probability*, 21(4) :2087–2130, 1993.

[LS97] H. Lefmann and P. Savický. Some typical properties of large And/Or Boolean formulas. *Random Structures and Algorithms*, 10 :337–351, 1997.

[Mat05] G. Matecki. Asymptotic density for equivalence. *Electronic Notes in Theoretical Computer Science*, 140 :81–91, 2005.

[Mon04] D. Villa Monteiro. Etude des fonctions Booléennes et représentation par arbres binaires. Thèse de Master, 2004.

[MTZ00] M. Moczurad, J. Tyszkiewicz, and M. Zaionc. Statistical properties of simple types. *Mathematical Structures in Computer Science*, 10(5) :575–594, 2000.

[MW55] L. Moser and M. Wyman. An asymptotic formula for the bell numbers. *Transactions of the Royal Society of Canada*, XLIX, 1955.

[PSS08] C. Pivoteau, B. Salvy, and M. Soria. Combinatorial Newton iteration to compute Boltzmann oracle. In *Fifth Colloquium on Mathematics and Computer Science*, Blaubeuren, Germany, september 2008. DMTCS Proceedings.

[PVW94] J. B. Paris, A. Vencovská, and G. M. Wilmers. A natural prior probability distribution derived from the propositional calculus. *Annals of Pure and Applied Logic*, 70 :243–285, 1994.

[RS63] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics*. PWN Warsaw, 1963.

[Sav90] P. Savický. Random Boolean formulas representing any Boolean function with asymptotically equal probability. *Discrete Mathematics*, 83 :95–103, 1990.

[Ser04] R. A. Servedio. Monotone Boolean formulas can approximate monotone linear threshold functions. *Discrete Applied Mathematics*, 142(1–3) :181–187, 2004.

[Sha49] C. E. Shannon. The synthesis of two-terminal switching circuits. *Bell Systems Technical J.*, 28 :59–98, 1949.

[Sta80] R. Statman. On the existence of closed terms in the typed lambda calculus I. *Combinatory Logic, lambda calculus and formalism*, pages 511–534, 1980.

[SU98] M. H. Sørensen and P. Urzyczyn. *Lectures on the Curry-Howard isomorphism*, volume 149. Studies in Logic and the Foundations of Mathematics, 1998.

[TD88] A. S. Troelstra and D. Van Dalen. *Constructivism in Mathematics : an Introduction*. Studies in Logic and the Foundations of Mathematics. North-Holland, 1988.

[Val84] L. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5 :363–366, 1984.

[vD86] D. van Dalen. Intuitionistic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic : Volume III : Alternatives to Classical Logic*, pages 225–339. Reidel, Dordrecht, 1986.

[Weg87] I. Wegener. *The complexity of Boolean functions*. John Wiley & Sons, Inc., New York, NY, USA, 1987.

[Woo97] A. R. Woods. Coloring rules for finite trees, and probabilities of monadic second order sentences. *Random Structures Algorithms*, 10(4) :453–485, 1997.

[Woo05] A. Woods. On the probability of absolute truth for And/Or formulas. *Bulletin of Symbolic Logic*, 12(3), 2005.

[Zai05] M. Zaionc. On the asymptotic density of tautologies in logic of implication and negation. *Reports on Mathematical Logic*, 39, 2005.