LMV
Laboratoire de mathématiques de Versailles - CNRS UMR 8100

UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES

Thèse
présentée pour obtenir le titre de
docteur de l'université de Versailles St-Quentin

Spécialité :
Mathématiques - Informatique

# Arbres booléens aléatoires et urnes de Pólya : approches combinatoire et probabiliste.

Cécile Mailler

Soutenue le 17 Octobre 2013 devant le jury composé de :

| | | |
|---|---|---|
| M. Dominique Barth | Université de Versailles St-Quentin | Examinateur |
| M. Philippe Chassaing | Université de Lorraine | Examinateur |
| Mme Brigitte Chauvin | Université de Versailles St-Quentin | Directrice |
| M. Julien Clément | Université de Caen | Examinateur |
| Mme Danièle Gardy | Université de Versailles St-Quentin | Directrice |
| M. Conrado Martínez | Université de Catalogne | Rapporteur |
| M. Alain Rouault | Université de Versailles St-Quentin | Examinateur |

Après avis des rapporteurs :

| | |
|---|---|
| M. Svante Janson | Université d'Uppsalla |
| M. Conrado Martínez | Université de Catalogne |

# Remerciements

Cette thèse, que vous allez peut-être lire, ou parcourir, est le fruit de trois années de ma vie. Ces trois années ont été très enrichissantes pour moi, bien entendu scientifiquement, mathématiquement, mais aussi humainement. Vous tous que j'ai croisés dans mon bureau, au laboratoire, en conférence, en groupe de travail, lors de mes activités d'enseignement, c'est grâce à vous que les mathématiques sont encore plus belles.

Je voudrais remercier mes rapporteurs, Svante Janson et Conrado Martínez d'avoir accepté de relire ma thèse, en français, et de l'avoir fait avec minutie et bienveillance. Merci aux membres du jury d'avoir accepté d'être présents le 17 Octobre 2013 malgré leurs nombreuses autres occupations : c'est un honneur de soutenir ces travaux devant Dominique Barth, Brigitte Chauvin, Philippe Chassaing, Julien Clément, Danièle Gardy, Conrado Martínez et Alain Rouault.

Il y a maintenant trois ans et demi, Brigitte Chauvin et Danièle Gardy m'ont fait confiance et ont accepté d'encadrer mon stage de M2, puis ma thèse. C'était un bonheur de travailler, d'écrire des articles, de partir en conférence avec vous. Protégée par vous des turbulences extérieures, ma thèse s'est déroulée comme dans un cocon. Pourtant, vous avez su ne pas m'enfermer entre vous deux, vous m'avez appris à voler de mes propres ailes : monter des collaborations, participer à des colloques, rencontrer d'autres chercheurs, trouver un postdoc, le tout sans être jamais loin si un conseil était nécessaire. Vous étiez les directrices de thèse qu'il me fallait : pédagogues, compréhensives, rassurantes, amicales, complémentaires, patientes, généreuses, chaleureuses, amusantes, enthousiastes... je vous remercie.

Ma thèse s'est déroulée au sein du Laboratoire de Mathématiques de Versailles où je me suis très vite sentie chez moi. Je remercie notamment Yvan Martel, directeur du laboratoire à mon arrivée, pour son enthousiasme envers les nouveaux étudiants en thèse : nous étions accueillis comme des rois ! Merci à Catherine Donati-Martin, qui lui a succédé, et grâce à qui la fin de ma thèse fut tout aussi agréable. Merci à Laure Frèrejean, la gestionnaire du laboratoire, pour toute son aide, son efficacité, et son amitié au cours de ces trois années : grâce à elle, je pouvais faire des mathématiques l'esprit libre. Merci aussi à Nadège Arnaud, la bibliothécaire du laboratoire, qui a supporté, toujours avec le sourire, les retards récurrents de mes emprunts et ma conduite automobile sportive entre l'Inria Rocquencourt et Versailles. Je pourrais ainsi citer tous les membres du laboratoire pour les bons moments passés à Versailles : l'équipe de probabilités et statistiques pour les séminaires hebdomadaires et les déjeuners qui s'en suivent au Picardie, le petit groupe de thésards pour les séminaires et les apéritifs à la Pirogue, sans compter les quelques fous rires dans le bureau quand Aurélien parle tout seul.

Je n'étais jamais très loin du laboratoire d'informatique, et j'aimerais remercier Chantal Ducoin, gestionnaire de l'ANR Boole, pour son efficacité et sa bonne humeur ainsi que les thésards adeptes

d'Abalone (on leur pardonne) du troisième étage pour les pauses cafés et autres pique-niques au château.

Durant ces trois années, j'ai effectué 64 heures d'enseignement par an. J'aimerais tout d'abord remercier les directeurs des départements de mathématiques et d'informatique de l'époque, Otared Kavian et Franck Quessette, qui ont accepté de me laisser enseigner un an en mathématique puis deux ans en informatique, et ce sans hésitation. Je leur en suis très reconnaissante. J'aimerais tout particulièrement remercier Franck Quessette pour l'attention qu'il m'a portée dans l'attribution des enseignements en informatique. Je suis très reconnaissante envers les différentes personnes avec qui j'ai pu enseigner, toujours avec plaisir : Sandrine Vial en algorithmique, Franck Quessette en simulation et en programmation, Pierre Coucheney et Yann Strozecki en simulation, Brigitte Chauvin et Oleksiy Khorunzhiy en probabilités. J'aimerais par ailleurs remercier Simon Clavière et Luca de Feo pour leurs remplacements réguliers et infaillibles lors de mes absences. Enfin, et surtout, merci à Liliane Roger, secrétaire du département de mathématiques pour son efficacité, sa constance, sa gentillesse, et sa conversation toujours agréable.

Ces trois années se sont aussi déroulées à l'extérieur de Versailles. Il serait trop long de remercier tous les membres d'Aléa pour leur accueil chaleureux parmi eux et pour la conférence au CIRM qui est, chaque année, l'un des points d'orgue de mon année scientifique : merci à tous. Merci aussi aux membres de l'ANR Boole pour les rencontres bi-annuelles toujours aussi enrichissantes, et, parmi eux, merci à Jérémie Lumbroso et Basile Morcrette pour leur initiation à la combinatoire analytique en petit comité ; merci aux membres du groupe de travail de l'Inria Rocquencourt et à Virginie Colette pour leur accueil sans réserves dans les pré-fabriqués ; ainsi qu'aux membres d'AofA que je suis ravie d'avoir pu rencontrer et écouter à Vienne, Będlewo, Montréal ou Minorque. Je voudrais particulièrement remercier mes différents collaborateurs : Antoine Genitrini pour son amitié rassurante, pour nos nombreuses discussions scientifiques et pour les bons moments passés à Vienne, à Jussieu, et ailleurs ; Bernhard Gittenberger et Veronika Kraus pour les séances de travail dans la bonne humeur entre Versailles et Vienne et pour les dîners au Valmont ou au Powidl ; Nicolas Pouyanne pour son enthousiasme scientifique sans bornes ; et Nicolas Broutin pour ce tout récent projet dans lequel je mets beaucoup d'espoir et d'entrain.

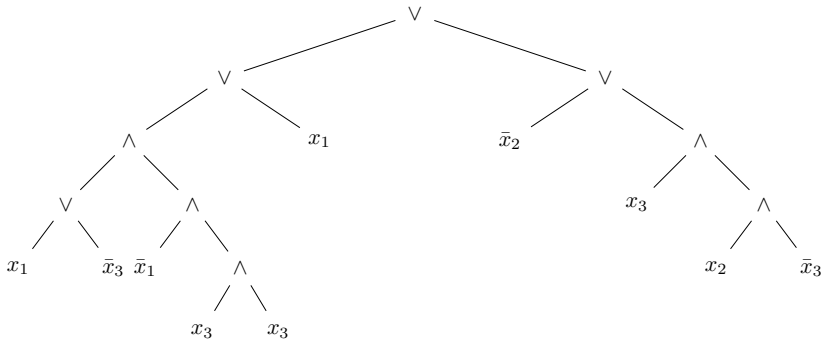
Enfin, je voudrais remercier mes amis grâce à qui j'ai pu profiter pleinement des week-ends et des vacances : merci à Alice, Victor et Mél pour les week-ends couture, rollers et jeux de rôle ; merci à Chantal et Olivier pour les sorties culturelles et musicales, et pour les mille services qu'ils m'ont rendus depuis maintenant sept ans ; merci à Antoine et Jules pour le tarot, la belote et les lendemains de nouvel an ; et merci à Gilbert, mon fidèle co-pupitre à l'orchestre du campus d'Orsay, pour son amitié indéfectible, malgré mon inconstance musicale. Merci aussi à ma famille : mes cousins François et Laurent et leurs parents auquels je ne rends pas assez visite à Tours, Constance ou La Rochelle, mon frère Sylvain et sa jolie petite famille... et surtout mes parents qui, je l'espère, n'ont pas besoin de ces remerciements écrits pour savoir que je leur suis reconnaissante au delà des mots pour tout ce qu'ils m'offrent depuis 26 ans.

Julien, merci de m'avoir supportée dans mes périodes de doute, merci d'avoir enduré mes nombreuses absences, merci de toujours me redonner confiance en moi, merci de jouer avec moi, de vivre avec moi, merci d'être là, toujours.

# Avant-propos

Ce mémoire présente les résultats obtenus pendant ma thèse concernant deux objets aléatoires : arbres booléens aléatoires et urnes de Pólya. Ces deux objets sont suffisamment distincts pour être présentés dans deux parties différentes, mais nous verrons à l'occasion comment un modèle d'urne peut être utile dans l'étude des arbres booléens, ou, inversement comment représenter une urne de Pólya par une forêt, c'est à dire un ensemble d'arbres.

FIGURE 1 – *Un arbre booléen.*

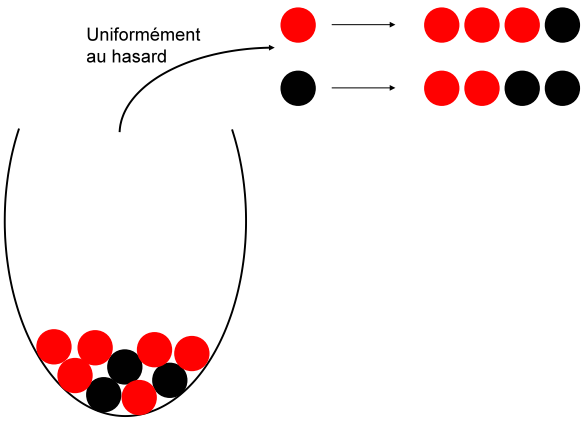



Ces deux thèmes ont aussi en commun les méthodes utilisées dans la littérature pour les étudier. Dans la littérature, tout comme dans la première partie de ce mémoire, les arbres booléens aléatoires sont étudiés principalement via des méthodes de combinatoire analytique. Un modèle étudié dans ce mémoire, celui de l'*arbre bourgeonnant*, inspiré de l'arbre binaire de recherche aléatoire (cf. Chapitre 5), donnera cependant lieu à une étude par plongement en temps continu (ou *poissonisation*). L'étude de ce modèle sera aussi l'occasion de deux modélisations par modèles d'urnes : un problème d'allocations, ou d'anniversaires, et un problème d'urne de Pólya.

Les urnes de Pólya, quant à elles, ont historiquement été étudiées par combinatoire, et, depuis les travaux d'Athreya, par plongement en temps continu. Plus récemment, les travaux de Flajolet et ses co-auteurs ont ouvert la voie à une approche des urnes de Pólya par combinatoire analytique. Dans la Partie II de ce mémoire, consacrée aux urnes de Pólya, nous ne développerons pas d'approche par combinatoire analytique, mais il est amusant de remarquer que de l'urne de Pólya utilisée dans le Chapitre 5 dans le cadre de l'*arbre bourgeonnant* est étudiée par combinatoire analytique. Dans la partie consacrée aux urnes de Pólya, nous utiliserons et couplerons deux approches : le plongement en temps continu et l'étude de la structure arborescente de l'urne.

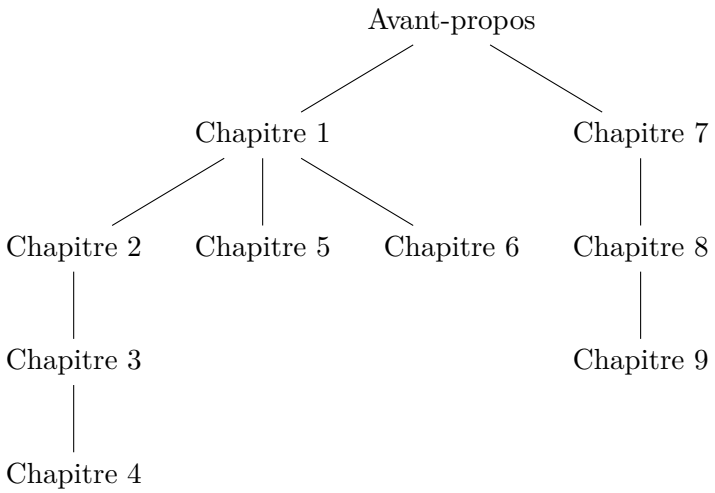
Ainsi, plongement en temps continu, combinatoire analytique et modèles d'arbres sont des approches que nous croiserons tout au long de ce mémoire. Si la Partie I traite d'arbres booléens aléatoires et la partie II d'urnes de Pólya, nous aurons l'agréable surprise de rencontrer des urnes dans la Partie I et des arbres, voire des forêts, dans la Partie II.

FIGURE 2 – *Une urne de Pólya.*




Ces deux thèmes sont aussi parents via leurs liens avec l'informatique fondamentale. La Partie I s'inscrit dans le cadre de la logique quantitative, en lien avec la logique intuitionniste et le lambda-calcul. Les arbres aléatoires étudiés seront par ailleurs issus de modèles classiques en informatique (arbres de Catalan, arbre binaire de recherche) ou en probabilités (arbres de Catalan, arbres de Galton-Watson). Modéliser et étudier des structures de données comme les arbres 2-3 ou les arbres $m$-aires de recherche via des urnes de Pólya (Partie II) est une approche standard en informatique fondamentale. Et c'est en vue d'obtenir des résultats sur ces structures de données qu'une partie de la théorie des urnes de Pólya s'est développée.

Les Parties I et II sont indépendantes. Les Chapitres 2, 3 et 4 sont conçus pour être lus dans cet ordre. Par contre, les Chapitres 5 et 6 sont indépendants. Enfin, je conseille au lecteur de lire les trois chapitres composant la Partie II dans l'ordre.



# Table des matières

# Première partie

# Arbres booléens et fonctions booléennes aléatoires

# Chapitre 1

# Introduction et préliminaires

## 1.1 Contexte

La première partie de ce mémoire est consacrée aux arbres booléens aléatoires. L'idée générale consiste à définir une distribution de probabilité sur l'ensemble des fonctions booléennes via une représentation arborescente. Nous nous attacherons à définir plusieurs distributions de probabilité sur l'ensemble des fonctions booléennes, nous les étudierons et les comparerons.

Considérons tout d'abord l'espace des fonctions booléennes à $k$ variables, où $k$ est un entier fixé. Cet ensemble est fini et a pour cardinal $2^{2^k}$. La distribution uniforme sur cet ensemble a été étudiée par Shannon [Sha49], qui a démontré que, presque sûrement quand $k$ tend vers $+\infty$, presque toute fonction booléenne est de *complexité* presque maximale : ce phénomène est appelé *effet Shannon*. Il y a deux types de complexité d'une fonction booléenne : la complexité en circuit ou la complexité en arbre (cf. [Weg05, Juk12] par exemple). La complexité en circuit d'une fonction booléenne est la taille du plus petit circuit logique qui la représente. C'est cette complexité que considère Shannon dans ses travaux. La complexité en arbre d'une fonction booléenne est la taille du plus petit arbre booléen représentant cette fonction : les résultats de Shannon sont généralisés à la complexité en arbre par Wegener [Weg87].

Paris et al. [PVW94] puis Lefmann et Savický [LS97] ont défini une nouvelle distribution sur l'ensemble des fonctions booléennes à $k$ variables via leur représentation par des arbres booléens. Dans ces travaux, un arbre booléen est un arbre binaire plan dont les nœuds internes sont étiquetés par des connecteurs logiques et dont les feuilles sont étiquetées par des littéraux, i.e. des variables $\{x_1, \dots, x_k\}$ et leurs négations $\{\bar{x}_1, \dots, \bar{x}_k\}$. Chaque arbre booléen est équivalent à une expression booléenne et représente donc une fonction booléenne à $k$ variables. La taille d'un arbre booléen sera le nombre de ses feuilles. Tirons uniformément au hasard un arbre booléen de taille $n$ : cet arbre calcule une fonction booléenne aléatoire $f_n$ dont nous notons la loi $\mu_{n,k}$. On s'intéresse à la limite, si elle existe, de cette suite de distributions quand $n$ tend vers $+\infty$. Ce modèle a été étudié par différents auteurs, dans deux systèmes logiques principaux : le système et/ou qui est un système complet au sens où toute fonction booléenne peut être représentée par un arbre de ce modèle, et le système de l'implication qui n'est pas complet mais est intéressant pour ses propriétés logiques. Il a été démontré par Lefmann et Savický [LS97], et Chauvin et al. [CFGG04] pour le modèle et/ou, et par Fournier et al [FGGG08] que cette suite de distributions tend vers une distribution limite $\mu_k$, appelée distribution des arbres de Catalan, quand la taille des arbres $n$ tend vers $+\infty$. Cette distribution limite a été étudiée par ces différents auteurs, dans deux systèmes logiques principaux : le système et/ou qui est un système complet au sens où toute fonction booléenne peut être représentée par un arbre de ce modèle, et le système de l'implication qui n'est pas complet mais qui est intéressant

pour ses propriétés logiques.

Il a été démontré dans les deux modèles (cf. Kozik [Koz08] pour le système et/ou et Fournier et al. [FGGG12] pour le système de l'implication) que la distribution des arbres de Catalan donne plus de poids aux fonctions de faible complexité, ce qui en fait donc, a priori, une distribution très différente de la distribution étudiée par Shannon. Genitrini et Gittenberger [GG10] ont montré que la distribution des arbres de Catalan n'exhibe pas d'effet Shannon, prouvant ainsi que son comportement est en effet très distinct de la distribution uniforme sur l'ensemble des fonctions booléennes à $k$ variables.

Au détour de ces travaux, d'autres résultats ont été démontrés : par exemple, Kozik et Genitrini [GK12] comparent logiques classiques et intuitionnistes. Une revue de littérature de Gardy [Gar06] résume les résultats antérieurs à 2006, ainsi que les méthodes utilisées : ces approches, depuis les travaux de Chauvin et al. [CFGG04], utilisent des outils de combinatoire analytique (auxquels une bonne introduction peut être lue dans le livre de Flajolet et Sedgewick [FS09]).

S'il est possible d'étendre cette étude à différents systèmes logiques (cf. Genitrini et Kozik [GK12]), il est aussi possible de modifier la distribution choisie sur l'ensemble des arbres booléens. Chauvin et al. définissent ainsi une distribution inspirée du processus critique de Galton-Watson que l'on étiquette ensuite aléatoirement de façon à obtenir un arbre booléen aléatoire. Cet arbre induit une distribution sur l'ensemble des fonctions booléennes à $k$ variables. Cette distribution a été étudiée dans le système de l'implication [FGGG12, GG10], aussi bien que dans le système et/ou [CFGG04], et il a été démontré que cette distribution a les même propriétés que la distribution des arbres de Catalan. Peut-on en déduire que toutes les distributions construites à partir d'arbres sur l'espace des fonctions booléennes à $k$ variables ont le même comportement ?

Fournier et al. (cf. [FGG09] ou [Gen09]) étudient la distribution induite par la distribution uniforme sur les arbres booléens équilibrés de hauteur $h$ (arbres dont toutes les feuilles sont à la même hauteur $h$). Lorsque $h$ tend vers $+\infty$, cette distribution converge vers la distribution *dégénérée* qui ne charge que les deux fonctions constantes `Vrai` et `Faux` dans le cas du modèle et/ou et vers la distribution qui ne charge que la fonction constante `Vrai` dans le cas du modèle de l'implication.

Pour résumer, presque toute fonction booléenne selon la distribution uniforme sur $\mathcal{F}_k$ est de complexité exponentielle, les distributions des arbres de Catalan et de Galton-Watson favorisent les fonctions de faible complexité, et la distribution des arbres équilibrés est dégénérée au sens où elle ne charge que les deux fonctions de complexité nulle. Peut-on comprendre pourquoi ces modèles ont des comportements différents ? Notamment, pour les modèles arborescents, quelles sont les propriétés de l'arbre booléen aléatoire qui décident du comportement de la distribution induite sur l'espace des fonctions booléennes ?

Le présent manuscrit s'attache à définir de nouvelles distributions induites par la représentation arborescente des fonctions booléennes, afin de mieux comprendre le lien entre les propriétés de la distribution définie sur l'ensemble des arbres booléens et celles de la distribution induite sur les fonctions booléennes. Après les études des arbres de Catalan et du processus binaire critique de Galton-Watson, il est naturel d'introduire une distribution induite par le processus de l'arbre binaire de recherche aléatoire (cf. Cormen et al. [CLR89] pour une définition de ce processus). Ce nouveau modèle d'arbre est, a priori, intéressant car l'arbre binaire de recherche aléatoire à $n$ feuilles est de hauteur d'ordre $\ln n$, alors que les arbres de Catalan sont de hauteur d'ordre $\sqrt{n}$. De plus, le niveau de saturation (i.e. la hauteur de la feuille la plus proche de la racine) de l'arbre binaire de recherche aléatoire est d'ordre moyen $\ln n$ alors que celui des arbres de Catalan est d'ordre $\Theta(1)$ (quand $n$ tend vers $+\infty$). Cela en fait donc un modèle très différent du modèle des arbres de Catalan. La forme de l'arbre binaire de recherche se rapproche plus de celle d'un arbre équilibré : cela implique-t-il que la distribution qu'il induit sur l'ensemble des fonctions booléennes est dégénérée ?

Il est intéressant de s'arrêter sur un autre modèle évoqué notamment dans Genitrini et al. [GKZ07,

GK12] : un modèle dans lequel les arbres booléens sont étiquetés sur un ensemble infini de variables et non sur l'ensemble fini $\{x_1, \ldots, x_k\}$. Dans le modèle classique des arbres de Catalan, on opère en effet une double limite ordonnée : tout d'abord, la taille des arbres tends vers $+\infty$ afin de définir la distribution des arbres de Catalan, puis le nombre de variables $k$ tend vers $+\infty$. Cette seconde limite sur $k$ est nécessaire pour pouvoir étudier le modèle : aucun résultat n'est actuellement connu *pour $k$ petit* dans la littérature. Cette double limite biaise peut-être la distribution induite sur les fonctions booléennes : les arbres sont grands mais étiquetés sur un ensemble petit de variables. La solution naturelle étudiée par Genitrini et al. [GKZ07, GK12] est d'étiqueter les arbres de Catalan sur un ensemble infini de variables, ce qui inverse donc les deux limites. Dans ces travaux, seule la fonction constante `Vrai` est étudiée, et seulement dans le modèle de l'implication. Que peut-on dire en toute généralité de la distribution ainsi induite sur l'ensemble des fonctions booléenne ? Se comporte-t-elle comme dans le cas classique des arbres de Catalan ?

Par ailleurs, comme souligné dans le survey de Gardy [Gar06], le modèle des arbres de Catalan est finalement assez peu naturel puisqu'il ne prend pas en compte les propriétés logiques des connecteurs, comme l'associativité et la commutativité des connecteurs ET et OU. Pour prendre en compte ces propriétés, il suffit de considérer des arbres non binaires (pour l'associativité) et non plans (pour la commutativité). Prendre en compte ces propriétés logiques change-t-il le comportement général de la distribution induite sur l'ensemble des fonctions booléennes ?

Cette Partie Arbres booléens aléatoires est séparée en deux sous-parties : la première se concentre sur les modèles d'arbres prenant en compte les propriétés logiques (associativité, commutativité) des connecteurs logiques en introduisant des modèles d'arbres booléens non-binaires et non-plans : la seconde sous-partie se recentre sur des arbres booléens binaires et plans en introduisant un nouveau modèle d'arbre aléatoire ou un nouvel étiquetage.

Les travaux exposés dans les Chapitres 2, 3 et 4 de la première sous-partie sont issus d'une collaboration avec Antoine Genitrini (LIP6, Paris), Bernhard Gittenberger (TU, Vienne) et Veronika Kraus (TU, Vienne).

Les Chapitres 2 et 3 concernent le système logique et/ou dans lequel associativité et commutativité des connecteurs sont prises en compte par l'introduction d'arbres non binaires et non plans. Dans le Chapitre 2, la taille d'un arbre est le nombre de ses feuilles : ce choix permet de comparer les résultats obtenus dans le modèle des arbres de Catalan où la taille se mesure en nombre de feuilles. Nous verrons que prendre en compte les propriétés logiques des connecteurs ne change pas fondamentalement le comportement induit sur l'ensemble des fonctions booléennes. Les méthodes utilisées dans ce chapitre sont similaires à celles utilisées dans le modèle classique : fonctions génératrices et théorème de Drmota-Lalley-Woods [Drm97, Lal93, Woo97] pour l'existence de la distibution asymptotique quand la taille $n$ des arbres tend vers $+\infty$, combinatoire analytique et *théorie des motifs* de Kozik pour montrer qu'un arbre typique calculant une fonction booléenne $f$ fixée est, asymptotiquement quand $k$ tend vers $+\infty$, un arbre minimal de $f$ dans lequel un grand arbre a été *greffé*, sans changer la fonction calculée par l'arbre minimal. Les résultats de ce Chapitre sont actuellement soumis au journal Random Structures and Algorithms (cf. [GGKM13]).

Mesurer la taille d'un arbre en terme de feuilles est cohérent lorsque l'on parle d'arbre binaire puisqu'un arbre binaire à $n$ feuilles a $n-1$ nœuds internes. Cette relation étroite entre nombre de feuilles et nombre de nœuds internes n'est plus vraie dans le cas d'arbres non-binaires, et le choix de la taille comme nombre de feuilles devient dès lors contestable : que se passe-t-il si la notion de taille est définie comme étant le nombre total de nœuds d'un arbre ? Il est à remarquer que changer la notion de taille des arbres change aussi la notion de complexité d'une fonction booléenne. Nous montrerons que ce changement de notion de taille ne semble pas changer profondément le comportement induit sur l'ensemble des fonctions booléennes. Ceci dit, les méthodes utilisées dans le Chapitre 2 ne s'appliquent plus, et le modèle s'avère assez contre-intuitif s'il est comparé au

modèle de taille usuel. Les travaux présentés dans le Chapitre 3 sont en cours, nous présentons ici quelques résultats ainsi qu'une conjecture.

Le Chapitre 4 concerne le système logique de l'implication : une formule booléenne de la forme $(A_{\sigma(1)} \to (A_{\sigma(2)} \to \dots (A_{\sigma(p)} \to \alpha)))$ calcule la même fonction booléenne, quel que soit l'ordre des *prémices* $A_1, \dots, A_p$, donc quelle que soit la permutation $\sigma$ de $\{1, \dots, p\}$. Pour prendre en compte cette propriété, nous représentons une telle formule par un arbre non-plan dont la racine est étiquetée par le littéral $\alpha$, et dont les sous-arbres de la racine représentent les prémices $A_1, \dots, A_p$. Tous les nœuds de ces arbres sont étiquetés par des littéraux, alors que dans le modèle binaire, chaque nœud interne est étiqueté par $\to$ : une information qui est finalement inutile puisque $\to$ est le seul connecteur logique autorisé. Nous définissons donc un modèle plus *élégant*, car sans information inutile. La taille d'un tel arbre booléen est le nombre total de ses nœuds, ce qui correspond au nombre de littéraux apparaissant dans la formule logique associée, et donc au nombre de feuilles dans un arbre binaire représentant cette formule. La comparaison entre ce nouveau modèle et le modèle binaire plan d'arbres implicatifs est donc justifiée. Nous montrons dans ce chapitre que le comportement global de la distribution induite sur l'ensemble des fonctions booléennes n'est pas modifié par la prise en compte de cette propriété logique du connecteur $\to$, et pour ce faire, nous généralisons les méthodes de combinatoire analytique développées dans [FGGG12]. Ces travaux sont publiés dans l'Electronic Journal of Combinatorics [GGKM12].

La seconde sous-partie, constituée des Chapitres 5 et 6, se recentre sur des arbres binaires et plans. Le Chapitre 5 regroupe un travail en collaboration avec Brigitte Chauvin (LMV, France) et Danièle Gardy (PRISM, France), publié à ANALCO 2011 et soumis, en version longue à Random Structures and Algorithms, et un travail en cours (cf. Section 5.5) en collaboration avec Nicolas Broutin (Inria, France). L'objet principal de ce chapitre est d'étudier la loi induite sur l'ensemble des fonctions booléennes par un arbre binaire de recherche aléatoire étiqueté uniformément au hasard de façon à être un arbre booléen. Nous montrons dans ce chapitre, via deux approches différentes (combinatoire analytique, et plongement en temps continu), que cette distribution est *dégénérée* au sens où elle ne charge que les deux fonctions constantes `Vrai` et `Faux` dans le modèle et/ou, et que la fonction `Vrai` dans le modèle de l'implication. Nous étudions en outre des modèles dans lequel l'étiquetage des nœuds internes est biaisé : la probabilité qu'un nœud interne soit étiqueté par ET n'est plus $\frac{1}{2}$ mais $q \in [0, 1]$ (et ce indépendamment des autres nœuds internes). Cette étude de modèles biaisés permet de remarquer que la distribution induite par l'arbre binaire de recherche aléatoire (renommé arbre bourgeonnant dans ce mémoire) a le même comportement que celle induite par les arbres équilibrés (cf. Fournier et al. [FGG09]). C'est cette remarque qui conduit à l'élaboration d'une conjecture plus générale reliant le niveau de saturation des arbres booléens aléatoires au comportement de la loi qu'ils induisent sur l'ensemble des fonctions booléenne : conjecture démontrée dans la Section 5.5.

Enfin, le Chapitre 6 s'intéresse aux arbres booléens étiquetés sur un nombre infini de variables. De façon à éviter qu'il existe un nombre infini d'arbres booléens de taille $n$ (ce qui annihilerait toute approche par combinatoire analytique), nous regroupons les arbres booléens et les fonctions booléennes dans des classes d'équivalence : sans entrer dans les détails, deux arbres seront équivalents s'ils sont égaux à re-numérotation près des variables qui les étiquettent. Dès lors, comme un arbre de taille $n$ ne peut contenir plus de $n$ variables différentes comme étiquettes de ses feuilles, étudier ce modèle revient à étudier un modèle dans lequel le nombre de variables $k$ est en réalité égal à $n$ : nous faisons tendre en même temps $k$ et $n$ vers l'infini. Forts de cette remarque, nous pouvons généraliser à $k = k(n)$ où $k(n)$ est une fonction croissante qui tend vers $+\infty$ quand $n$ tend vers $+\infty$. L'étude de ce nouveau modèle se fait par combinatoire analytique et nécessite une généralisation de la *théorie des motifs* de Kozik. Nous montrons un comportement similaire à celui du cas classique "$k$ fini" : les fonctions de faible complexité sont favorisées par cette nouvelle distribution. Nous

montrons l'existence d'un seuil de l'ordre de $\frac{n}{\ln n}$ à partir duquel *ajouter de nouvelles variables ne change pas le comportement de la distribution induite sur les fonctions booléennes.* Ce travail, en collaboration avec Antoine Genitrini (LIP6, Paris), est soumis à LATIN 2014.

La fin de ce chapitre introductif est consacré à une présentation détaillée de l'état de l'art concernant l'étude des distributions arborescentes sur l'ensemble des fonctions booléennes puis à une présentation très résumée de la combinatoire analytique.

## 1.2 Fonctions booléennes et arbres booléens

Comme signalé auparavant, ce mémoire s'inscrit à la suite de nombreux travaux sur les représentations arborescentes de fonctions booléennes. Cette partie a pour but de présenter un état de l'art de ces différents travaux. Elle permettra de poser quelques définitions utiles dans toute cette partie sur les arbres booléens.

**Définition 1.2.1**

*Une fonction booléenne est une fonction de $\{0,1\}^{\mathbb{N}}$ dans $\{0,1\}$. On note $\mathcal{F}_\infty$ l'ensemble des fonctions booléennes.*

Il est courant de confondre l'ensemble $\{0,1\}$ avec l'ensemble $\{\texttt{Faux}, \texttt{Vrai}\}$, et nous utiliserons dans la suite aussi bien `Faux` que 0 et `Vrai` que 1.

**Exemple :** Voici quelques exemples de fonctions booléennes :
- les deux fonctions constantes `Vrai` : $((x_i)_{i\geqslant 1} \mapsto 1)$ et `Faux` : $((x_i)_{i\geqslant 1} \mapsto 0)$,
- les fonctions littéral, $((x_i)_{i\geqslant 1} \mapsto x_\ell)$ et $((x_i)_{i\geqslant 1} \mapsto \bar{x}_\ell = 1 - x_\ell)$ pour tout $\ell \geqslant 1$,
- une fonction XOR , $((x_i)_{i\geqslant 1} \mapsto x_1 \text{XOR } x_2 = (x_1 \wedge \bar{x}_2) \vee (\bar{x}_1 \wedge x_2))$.

En pratique, nous aurons souvent à considérer l'ensemble des fonctions booléennes à $k$ variables, pour $k$ un entier positif :

**Définition 1.2.2**

*Une fonction booléenne à $k$ variables est une fonction de $\{0,1\}^k$ dans $\{0,1\}$. Nous noterons $\mathcal{F}_k$ l'ensemble des fonctions booléennes à $k$ variables.*

Bien entendu, toute fonction de $\mathcal{F}_k$ peut être vue comme fonction de $\mathcal{F}_\infty$, ce qui nous amène à définir la notion de variable essentielle pour une fonction booléenne $f$.

**Définition 1.2.3**

*Pour tout $i \geqslant 1$, la variable booléenne $x_i$ ($i \geqslant 1$) est une* **variable essentielle** *de la fonction booléenne $f$ si et seulement si $f_{|x_i=0} \not\equiv f_{|x_i=1}$ (où $f_{|x_i=0}$ est la restriction de $f$ au sous-espace de $\{0,1\}^{\mathbb{N}}$ défini par l'équation $x_i = 0$). On notera $E(f)$ le nombre de variables essentielles de $f$.*

**Exemple :**
- Les deux fonctions constantes `Vrai` et `Faux` n'admettent pas de variables essentielles.
- Pour tout $\ell \geqslant 1$, la fonction littéral $((x_i)_{i\geqslant 1} \mapsto x_\ell)$ a une unique variable essentielle : $x_\ell$.
- La fonction XOR admet deux variables essentielles.

Avant d'introduire la notion d'arbre booléen, définissons quelques mots de vocabulaire concernant les arbres. Dans ce mémoire, les arbres considérés seront tous enracinés. Nous appellerons **nœuds** les sommets de l'arbre, le **degré** d'un nœud sera le nombre d'arêtes auxquelles il appartient, un nœud de degré 1 est une **feuille** de l'arbre et les nœuds qui ne sont pas des feuilles sont

des **nœuds internes**. Étant donné un nœud $\nu$ de l'arbre, il existe un unique chemin $\gamma_\nu$ reliant $\nu$ à la racine. Le nombre d'arêtes composant ce chemin est la **hauteur** de $\nu$ dans l'arbre, ou sa **génération** : la racine est de hauteur zéro. Les nœuds appartenant à $\gamma_\nu$ sont les **ancêtres** de $\nu$, l'ancêtre de $\nu$ relié à $\nu$ par une arête est son **parent**, et réciproquement, $\nu$ est un **enfant** de son parent. Les autres enfants du parent de $\nu$ sont ses **frères**. Tous les nœuds dont $\nu$ est un ancêtre sont appelés **descendants** de $\nu$, et l'**arité** de $\nu$ sera le nombre de ses enfants. La racine d'un arbre n'a pas de parent, ni de frère, et tous les nœuds de l'arbre sauf elle-même sont ses descendants.

**Définition 1.2.4**

*Un* **arbre booléen** *est un arbre enraciné dans lequel tout nœud interne a au moins deux descendants, dont les nœuds internes sont étiquetés par des connecteurs logiques, et dont les feuilles sont étiquetées par des littéraux.*

Un arbre booléen est naturellement équivalent à une expression booléenne, et représente donc une fonction booléenne. Bien entendu, cette correspondance entre arbres booléens et fonctions booléennes n'est pas bijective puisque plusieurs arbres peuvent représenter une même fonction booléenne. L'objet de ce mémoire est de définir des distributions de probabilité sur l'ensemble des fonctions booléennes via leur représentation arborescente. Dans la suite, nous pourrons avoir besoin de la fonction suivante qui formalise cette correspondance entre arbres booléens et fonctions booléennes :

**Définition 1.2.5**

*La fonction $\Phi$ a pour espace de départ l'ensemble des arbres booléens $\mathcal{B}$, et pour espace d'arrivée celui des fonctions booléennes $\mathcal{F}_\infty$, et est définie par : pour tout arbre $t \in \mathcal{B}$,*

$$\Phi(t) = f \text{ si et seulement si } t \text{ représente } f.$$

Le choix des connecteurs logiques et des littéraux autorisés pour l'étiquetage d'un arbre définit un **système logique**. Dans ce mémoire, nous nous intéresserons principalement à deux systèmes logiques : le système et/ou et le système de l'implication. Le système et/ou est largement étudié dans la littérature car c'est un système complet : toute fonction booléenne peut être exprimée dans ce système logique. Le système de l'implication n'est pas complet (la fonction constante `Faux` ne peut être exprimée dans ce système), mais c'est un système simple qui est utile pour ses applications en logique, comme peut en attester le papier de Genitrini et Kozik [GK12] dans lequel logiques classique et intuitionniste sont quantitativement comparées.

**Définition 1.2.6 (Système logique - et/ou)**

*Dans le système logique et/ou, les connecteurs logiques sont le connecteur $\wedge$ (ET) et le connecteur $\vee$ (OU), et les littéraux autorisés sont les littéraux positifs $(x_i)_{i\geqslant 1}$ et négatifs $(\bar{x}_i)_{i\geqslant 1}$.*

**Définition 1.2.7 (Système logique - implication)**

*Dans le système logique de l'implication, le seul connecteur logique est celui de l'implication $\rightarrow$, et seuls les littéraux positifs $(x_i)_{i\geqslant 1}$ sont autorisés.*

Dans toute la suite du mémoire, nous aurons besoin d'une notion de **taille** pour les arbres booléens. De manière générale, la taille d'un arbre booléen sera son nombre de feuilles, et donc le nombre de littéraux qui apparaissent dans l'arbre. Dans la pratique, selon le modèle étudié, nous aurons besoin de modifier cette notion de taille et la taille pourra être le nombre de nœuds internes, ou le nombre total de nœuds. Pour tout arbre $t \in \mathcal{B}$, la taille de $t$ sera notée $|t|$. Une fois qu'une notion de taille est définie, elle induit une notion de complexité pour une fonction booléenne :

**Définition 1.2.8**

*Soit $f$ une fonction booléenne de $\mathcal{F}_\infty \backslash \{\texttt{Vrai}, \texttt{Faux}\}$, la* **complexité** *de $f$, notée $L(f)$ est la taille des plus petits arbres qui représentent $f$.*

Il est important de noter que la complexité d'une fonction booléenne $f$ dépend du système logique choisi, ainsi que de la notion de taille choisie. Dans le modèle et/ou, on posera par définition $L(\texttt{Vrai}) = L(\texttt{Faux}) = 0$. Dans le modèle de l'implication, nous poserons $L(\texttt{Vrai}) = 0$ et $L(\texttt{Faux}) = +\infty$, car $f$ n'est pas expressible dans ce système logique. De manière générale, toue fonction non expressible dans le système logique choisi sera de complexité $+\infty$.

Dans ses travaux sur les circuits booléens [RS42], Shannon a établi le résultat suivant, dont nous énonçons ici une version concernant les expressions booléennes (cf. [Weg87] et [FS09, page 77]). Plaçons-nous sur l'ensemble $\mathcal{F}_k$ des fonctions booléennes à $k$ variables, considérons le système logique et/ou et fixons la taille d'un arbre comme étant le nombre de ses feuilles :

**Théorème 1.2.9 (Riordan et Shannon [RS42])**

*La complexité maximale d'une fonction de $\mathcal{F}_k$ est égale à $\frac{2^k}{\log_2 k}$. Pour toute constante $\varepsilon > 0$, si l'on considère la distribution uniforme sur $\mathcal{F}_k$, alors, asymptotiquement quand $k$ tend vers $+\infty$, presque toute fonction de $\mathcal{F}_k$ a une complexité plus grande que $\frac{(1-\varepsilon)2^k}{\log_2 k}$.*

Lorsque l'on définit une nouvelle distribution sur $\mathcal{F}_k$, on dira que cette distribution exhibe un *effet Shannon* si asymptotiquement quand $k$ tend vers $+\infty$, presque toute fonction booléenne de $\mathcal{F}_k$, selon cette distribution, est de complexité exponentielle en $k$. Si, a contrario, on peut exhiber une sous-famille $\mathcal{S}_k$ de $\mathcal{F}_k$ dont la probabilité ne tend pas vers 0 quand $k$ tend vers $+\infty$, et telle que toute fonction de $\mathcal{S}_k$ a une complexité sous-exponentielle, on dira que cette distribution n'exhibe pas d'effet Shannon.

## 1.3 Arbres de Catalan

Dans les travaux originels de Paris et al. [PVW94], Lefmann et Savický [LS97], et Chauvin et al. [CFGG04], on se place dans le système logique et/ou, et on se restreint à considérer des arbres binaires, plans, étiquetés par des variables de l'ensemble $\{x_1, \ldots, x_k\}$ et leurs négations. Pour tout $n$ fixé, il y a un nombre fini d'arbres booléens de taille $n$ : nous pouvons donc considérer la distribution uniforme sur cet ensemble. Dans cette partie, nous rappelons la définition précise de ce modèle, dit des arbres de Catalan, et quelques résultats et méthodes de preuves.

Dans toute cette partie, la taille d'un arbre sera le nombre de ses feuilles.

### 1.3.1 Arbres et/ou

Notons $\mathcal{T}_{n,k}$ l'ensemble des arbres et/ou de taille $n$ étiquetés avec $k$ variables et leurs négations, et $T_{n,k}$ son cardinal : comme chaque nœud interne d'un tel arbre peut avoir deux étiquettes différentes et chaque feuille peut en avoir $2k$ différentes, nous avons

$$T_{n,k} = 2^{n-1}\texttt{Cat}_{n-1}(2k)^n,$$

où $\texttt{Cat}_n$ est le $n^{\text{ième}}$ nombre de Catalan : pour tout $n \geqslant 0$,

$$\texttt{Cat}_n = \frac{1}{n+1}\binom{2n}{n}. \tag{1.1}$$

Pour toute fonction $f \in \mathcal{F}_k$, notons $T_{n,k}(f)$ le nombre d'arbres de $\mathcal{T}_{n,k}$ calculant $f$, puis définissons la quantité suivante :

$$\mu_{n,k}(f) = \frac{T_{n,k}(f)}{T_{n,k}}.$$

On s'intéresse à la limite de ces suites quand $n$ tend vers $+\infty$, c'est à dire quand la taille des arbres considérés tend vers $+\infty$. Chauvin et al. démontrent que cette limite existe bien, et même que la valeur limite est strictement positive pour toute fonction booléenne de $\mathcal{F}_k$. Pour ce faire, les auteurs introduisent les fonctions génératrices suivantes : pour toute fonction $f \in \mathcal{F}_k$,

$$\phi_f(z) = \sum_{n \geqslant 1} T_{n,k}(f) z^n.$$

Ces $2^{2^k}$ fonctions génératrices vérifient le système suivant : pour toute fonction $f \in \mathcal{F}_k$,

$$\phi_f(z) = z \mathbb{1}_{f \text{ lit}} + \sum_{g \wedge h \equiv f} \phi_g(z)\phi_h(z) + \sum_{g \vee h \equiv f} \phi_g(z)\phi_h(z),$$

où $\mathbb{1}_{f \text{ lit}} = 1$ si $f$ est l'une des $2k$ fonctions littéral, et $\mathbb{1}_{f \text{ lit}} = 0$ sinon. Ce système d'équations vérifie les hypothèses du théorème de Drmota [Drm97], Lalley [Lal93] et Woods [Woo97]. Nous faisons référence au livre de Flajolet et Sedgewick [FS09, page 489] pour un énoncé unifié de ce théorème, qui nous assure ici de la convergence de chaque suite $(\mu_{n,k}(f))_{f \in \mathcal{F}_k}$ vers un réel strictement positif que nous noterons $\mu_k(f)$. La distribution $\mu_k$ ainsi définie sur $\mathcal{F}_k$ est appelée **distribution des arbres de Catalan**.

Cette distribution est l'objet d'un article de Kozik [Koz08] où le théorème suivant est démontré :

**Théorème 1.3.1**

*Dans le modèle et/ou, pour toute fonction booléenne $f$ fixée (i.e. telle que le nombre de ses variables essentielles $E(f)$ ne dépend pas de $k$), il existe une constante $\lambda_f > 0$ telle que, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(f) \sim \lambda_f \left(\frac{1}{k}\right)^{L(f)+1}.$$

Autrement dit, asymptotiquement quand $k$ tend vers $+\infty$, la distribution des arbres de Catalan donne plus de poids au fonctions de petite complexité. Pour montrer ce résultat, Kozik introduit une *théorie des motifs* qui lui permet de sélectionner dans un arbre un sous-ensemble de feuilles *cruciales* selon l'usage que l'on veut en faire. Cette théorie sera utilisée et généralisée tout au long du mémoire : en voici une brève description.

**Définition 1.3.2**

*Un* **motif** *est un arbre plan dont les nœuds internes sont étiquetés par des connecteurs $\wedge$ ou $\vee$, et dont les feuilles sont étiquetées par $\bullet$ ou par $\square$. Les feuilles étiquetées par $\bullet$ seront appelées* **feuilles de motifs** *et celles étiquetées par $\square$ seront appelées* **emplacements**.

*Un ensemble de motifs sera appelé* **langage de motifs**.

Nous pouvons par exemple définir des langages de motifs par induction : $N = \bullet | N \vee N | N \wedge \square$ (cf. Figure 1.1). Cela signifie qu'un motif de $N$ est soit une feuille de motif, soit une racine étiquetée par $\wedge$ dont le premier sous-arbre est dans $N$ et dont le second fils est un emplacement, soit une racine étiquetée par $\vee$ et dont les deux sous-arbres sont dans $N$.

FIGURE 1.1 – *Un arbre du langage de motifs $N$.*

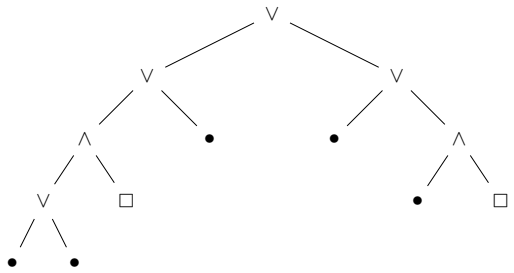
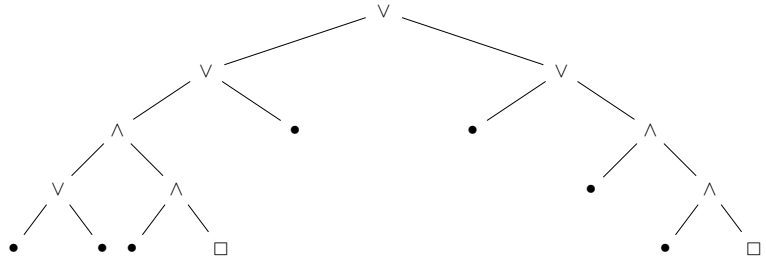


FIGURE 1.2 – *Un arbre du langage de motifs $N[N]$.*



Pour tout langage de motifs $L$, pour toute famille d'arbres $\mathcal{T}$, on note $L[\mathcal{T}]$ l'ensemble des arbres obtenus en greffant dans chaque emplacement d'un motif de $L$ un arbre de $\mathcal{T}$. On dira que $L$ est **non ambigu** si pour toute famille d'arbres $\mathcal{T}$, tout arbre de $L[\mathcal{T}]$ ne peut être obtenu qu'à partir d'un unique motif de $L$. On remarquera par exemple que $N$ est non ambigu. Étant donnés deux langages de motifs $L$ et $M$, on notera $L[M]$ le langage de motifs obtenu en greffant des arbres de $M$ dans les emplacements des arbres de $L$. On peut par ailleurs définir la fonction génératrice du langage de motifs $L$, où $x$ marque les feuilles de motifs, et $y$ marque les emplacements :

$$\ell(x,y) = \sum_{m\geqslant 0}\sum_{q\geqslant 0} \ell_{m,q} x^m y^q,$$

où $\ell_{m,q}$ est le nombre d'arbres dans $L$ ayant $m$ feuilles de motif et $q$ emplacements. Par la méthode symbolique, nous pouvons déduire que la fonction génératrice du langage de motifs $N$ vérifie :

$$n(x,y) = x + n(x,y)^2 + yn(x,y),$$

et donc

$$n(x,y) = \frac{1}{2}(1 - y - \sqrt{(1-y)^2 - 4x}).$$

Nous dirons que le langage de motifs $L$ est **sous-critique** pour la famille d'arbres $\mathcal{T}$ si et seulement si, la série génératrice $T(z)$ de la famille $\mathcal{T}$ a une singularité dominante en $\rho$ de type racine-carrée (cf. Sous-section 1.5), et s'il existe un $\varepsilon > 0$ tel que $\ell(x,y)$ est analytique sur l'ensemble $\{(x,y), |x| < \rho + \varepsilon, |y| < T(\rho) + \varepsilon\}$.

Après ces définitions, nous pouvons introduire la notion de *restriction*, notion qui mène à un résultat de Kozik qui sera crucial dans tout ce mémoire.

**Définition 1.3.3**

*Nous appellerons* **squelette** *un arbre booléen dont les feuilles ne sont pas étiquetées. Notons $\mathcal{S}$ la famille des arbres et/ou binaires plans à feuilles non étiquetées.*

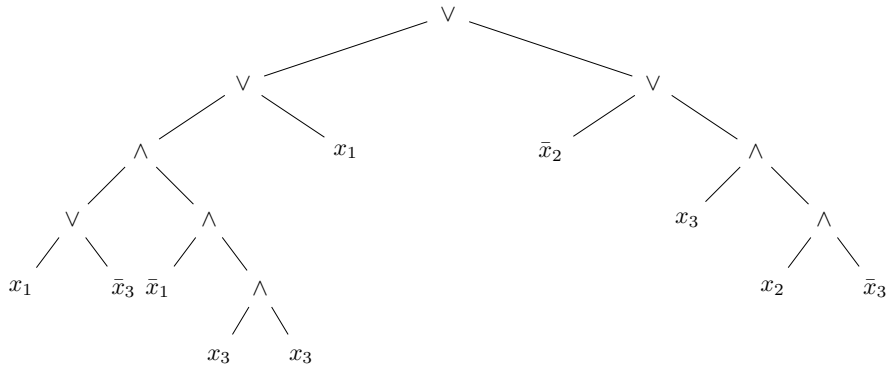
Soit $L$ un langage de motifs tel que $L[\mathcal{S}] = \mathcal{S}$ : tout arbre booléen est un motif de $L$ dans lequel ont été greffés des squelettes de $\mathcal{S}$ et dont les feuilles ont ensuite toutes été étiquetées par des littéraux. On appellera $L$-feuilles de motifs d'un arbre booléen les feuilles $\bullet$ de l'arbre de motif duquel il est issu.

**Définition 1.3.4**

*Soit $t$ un arbre et/ou et soit $\Gamma$ un sous-ensemble de $\{x_1, \ldots, x_k\}$. On dira que $t$ a $q$* $L$**-répétitions** *si $q$ est égal au nombre de ses $L$-feuilles de motifs moins le nombres de variables différentes qui les étiquettent. On dira que $t$ a $q$* $(L, \Gamma)$**-restrictions** *si $q$ est égal au nombre de $L$-répétitions plus le nombre de variables de $\Gamma$ distinctes qui apparaissent comme étiquettes des $L$-feuilles de motif de $t$.*

Par exemple, l'arbre de la Figure 1.3 est construit à partir du motif de $N$ de la Figure 1.1 : il a cinq $N$-feuilles de motif et admet 2 $N$-répétitions et 4 $(N, \{x_1, x_2\})$-restrictions. Il est aussi construit à partir du motif de $N[N]$ de la Figure 1.2, et admet 7 $N[N]$-feuilles de motifs, 5 $N[N]$ répétitions et 9 $(N[N], \{x_1, x_2\})$-restrictions.

FIGURE 1.3 – *L'arbre ci-dessous calcule la fonction booléenne $x_1 \vee \bar{x}_2$.*




Kozik a établi le lemme suivant, que nous généraliserons dans la suite à différents modèles :

**Lemme 1.3.5 (Kozik [Koz08])**

*Soit $L$ un langage de motifs non-ambigu, sous-critique par rapport à $\mathcal{S}$ (l'ensemble des squelettes, cf. Définition 1.3.3) et tel que $L[\mathcal{S}] = \mathcal{S}$, soit $\Gamma \subset \{x_1, \ldots, x_k\}$ un ensemble dont le cardinal ne dépend pas de $k$. Soit $T_{n,k}^{[p]}$ (resp. $T_{n,k}^{[\geqslant p]}$) le nombre d'arbres booléens de taille $n$ ayant exactement (resp. au moins) $p$ $(L, \Gamma)$-restrictions. Alors,*

$$\frac{T_{n,k}^{[p]}}{T_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right) \quad et \quad \frac{T_{n,k}^{[\geqslant p]}}{T_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right).$$

Kozik utilise ce lemme de manière centrale dans la preuve du Théorème 1.3.1 : il démontre que, asymptotiquement quand $k$ tend vers $+\infty$, *un arbre typique calculant $f$ est un arbre minimal de $f$*

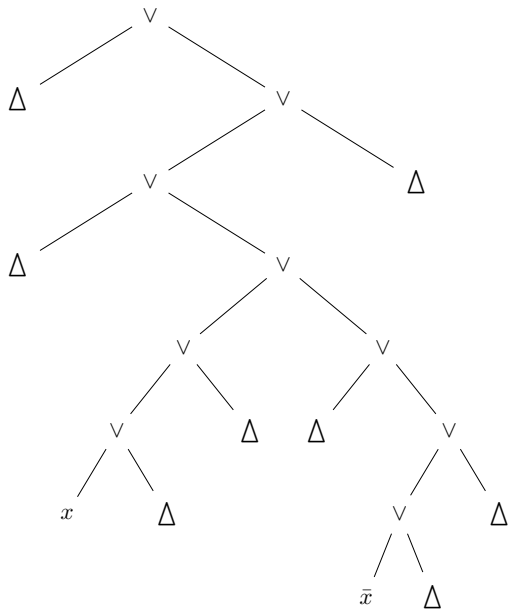
*dans lequel a été greffé un grand sous-arbre qui ne change pas la fonction globale calculée.* Nous reprendrons les idées de cette preuve dans le Chapitre 2, lorsque nous généraliserons le Théorème 1.3.1 à des arbres non binaires et non plans. Nous généraliserons aussi le lemme de Kozik 1.3.5 à un modèle dans lequel le nombre de variables $k$ est une fonction de la taille $n$ des arbres considérés (cf. Chapitre 6). Ces généralisations permettront de comprendre la preuve de ce lemme, ainsi que son utilisation.

La première étape de la preuve du Théorème 1.3.1 consiste à étudier le cas $f \equiv \texttt{Vrai}$. Il s'agit de montrer que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute **tautologie**, i.e. presque tout arbre calculant `Vrai`, est une *tautologie simple*, au sens suivant :

**Définition 1.3.6 (cf. Figure 1.4)**

*Dans le système logique et/ou, une* **tautologie simple** *est un arbre dans lequel il existe deux feuilles reliées à la racine par deux chemins de $\vee$ et étiquetées par une variable et sa négation. On note $\mathcal{ST}$ l'ensemble des tautologies simples et $ST_{n,k}$ le nombre de tautologies simples de taille $n$.*

FIGURE 1.4 – *Un exemple de tautologie simple dans le système et/ou : les $\bigwedge$ peuvent être remplacés par n'importe quel arbre et/ou.*




**Théorème 1.3.7 ([Koz08])**

*Dans le modèle et/ou, asymptotiquement quand $k$ tend vers $+\infty$,*

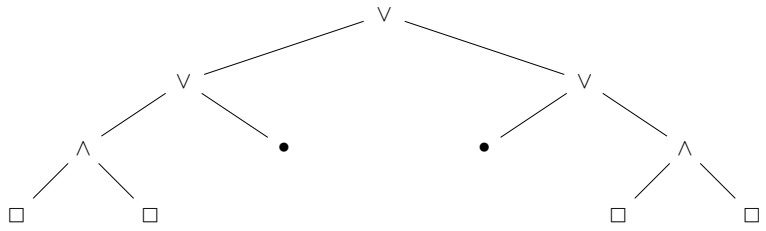
$$\mu_k(\texttt{Vrai}) \sim \lim_{n\to+\infty} \frac{ST_{n,k}}{T_{n,k}},$$

*autrement dit,* presque toute tautologie est simple.

Ce théorème est crucial car il réduit l'étude des tautologies à celle des tautologies simples, qui sont plus faciles à énumérer. Ce théorème se démontre à l'aide du langage de motifs $N = \bullet | N \vee N | N \wedge \square$ qui a la propriété suivante : si l'on peut assigner toutes les $N$-feuilles de motifs

d'un arbre et/ou à `Faux`, alors l'arbre total calcule la fonction `Faux` pour cette affectation partielle des variables. C'est cette propriété qui permet de démontrer le Théorème 1.3.7. Pour compter les tautologies simples, Kozik propose d'utiliser le langage de motifs $S = \bullet | S \vee S | \square \wedge \square$ : une tautologie simple est un arbre et/ou qui a deux $S$-feuilles de motifs étiquetées par une variable et sa négation. Par exemple, l'arbre de la Figure 1.3 est construit à partir du motif de la Figure 1.5 et a deux $S$-feuilles de motifs. Ces deux feuilles n'étant pas étiquetées par une variable et sa négation, ce n'est pas une tautologie simple.

FIGURE 1.5 – *L'arbre de la Figure 1.3 est construit à partir de ce motif de $S$. Il a donc deux $S$-feuilles de motif.*



Nous détaillerons ces preuves dans le Chapitre 2, lorsque nous les généraliserons à un modèle d'arbres non binaires, ou dans le Chapitre 6, lorsque nous les généraliserons au cas où $k$ dépend de $n$.

### 1.3.2 Arbres de l'implication

Dans le cadre du système logique de l'implication (cf. Définition 1.2.7), des résultats similaires ont été démontrés, notamment dans Fournier et al. ([FGGG12]) et dans la thèse de Genitrini. Le nombre d'arbres booléens de taille $n$ dans ce système logique est donné par

$$T_{n,k} = \mathtt{Cat}_{n-1} k^n.$$

Via des fonctions génératrices et le théorème de Drmota-Lalley-Woods, on peut démontrer que

$$\mu_{n,k}(f) = \frac{T_{n,k}(f)}{T_{n,k}}$$

converge quand $n$ tend vers $+\infty$ vers un réel $\mu_k(f)$ strictement positif pour toute fonction $f$ expressible dans le système de l'implication. Notons qu'une fonction booléenne est expressible dans le système de l'implication si, et seulement si, il existe un littéral $\alpha$ et une fonction booléenne $g$ tels que $f = \alpha \vee g$. Fournier et al. ont montré le théorème suivant :

**Théorème 1.3.8**

*Dans le modèle de l'implication, pour toute fonction booléenne $f$ fixée, il existe une constante $\lambda_f > 0$ telle que, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(f) \sim \frac{\lambda_f}{k^{L(f)+1}}.$$

Notons que si $f$ n'est pas expressible dans le système de l'implication, on peut poser $L(f) = +\infty$ afin que le théorème s'applique.

Pour prouver ce résultat, Fournier et al., tout comme dans le cas et/ou, montrent qu'*un arbre typique calculant $f$ est un arbre minimal de $f$ dans lequel a été greffé un grand sous-arbre qui ne change pas la fonction globale calculée*. Pour ce faire, ils utilisent une méthode alternative à celle des motifs de Kozik, laquelle ne semble pas pouvoir s'adapter au système de l'implication. Nous réutiliserons ces méthodes dans le Chapitre 4 dans lequel nous généraliserons ce résultat à un modèle d'arbres non binaires et non planaires qui prend en compte les propriétés logiques du connecteur $\rightarrow$.

Le Théorème 1.3.8 indique que la distribution des arbres de Catalan donne plus de poids aux fonctions de petite complexité. Cependant, ce résultat n'est pas suffisant pour contredire l'effet Shannon (cf. Théorème 1.2.9). Dans un article plus récent, Genitrini et Gittenberger ([GG10]) montrent que la distribution des arbres de Catalan n'exhibe pas l'effet Shannon :

**Théorème 1.3.9 ([GG10])**

*Soit $R_k = \frac{9\pi k^2}{16}$. Asymptotiquement quand $k$ tend vers $+\infty$, la probabilité des fonctions de complexité au plus $R_k$ est plus grande que $\frac{9}{64}$. La distribution des arbres de Catalan n'exhibe donc pas d'effet Shannon.*

Un tel résultat n'est pas démontré pour le système et/ou, même s'il est communément admis. Il semble donc que la distribution des arbres de Catalan est fondamentalement différente de la distribution uniforme sur $\mathcal{F}_k$. Il est donc naturel de se demander si ce comportement est typique des distributions issues de la représentation arborescente des fonctions booléennes.

Dans le système de l'implication, c'est aussi l'étude des tautologies qui est fondamentale. Tout comme dans le cas et/ou, on montre que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple. Ce résultat, en plus d'être fondamental pour la suite de l'étude de la distribution des arbres de Catalan dans le système de l'implication, est intéressant en lui-même puisqu'il permet d'affirmer que logique classique et intuitionniste sont, asymptotiquement quand $k$ tend vers $+\infty$, équivalentes. En effet, il est connu que les tautologies simples sont intuitionnistes, et le résultat suivant, montré dans [FGGZ07] nous assure donc que presque toute tautologie est intuitionniste.

**Définition 1.3.10 (cf. Figure 1.6)**

*Dans le modèle de l'implication, toute expression booléenne s'écrit sous la forme $A_1 \rightarrow (A_2 \rightarrow \ldots (A_p \rightarrow \alpha))$, où les $(A_i)_{i=1,\ldots,p}$ sont des expressions booléennes. Les sous-arbres représentant $A_1, \ldots, A_p$ sont appelés* **prémisses** *de l'arbre booléen, et $\alpha$ est son* **but**. *Une* **tautologie simple** *est un arbre booléen dont au moins l'une des prémisses est réduite à une feuille étiquetée par le littéral $\alpha$.*

*On notera $\mathcal{ST}$ l'ensemble des tautologies simples et $ST_{n,k}$ l'ensemble des tautologies simples de taille $n$ étiquetées sur $k$ variables.*

**Théorème 1.3.11**

*Dans le système de l'implication, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\texttt{Vrai}) \sim \lim_{n\to+\infty} \frac{ST_{n,k}}{T_{n,k}}.$$

Tout comme dans le système et/ou, ce résultat permet de ramener l'étude des tautologies à l'étude d'une sous-famille plus facilement énumérable. La démonstration de ce résultat sera reprise dans le Chapitre 4, et généralisée pour un modèle d'arbres non binaires et non planaires.

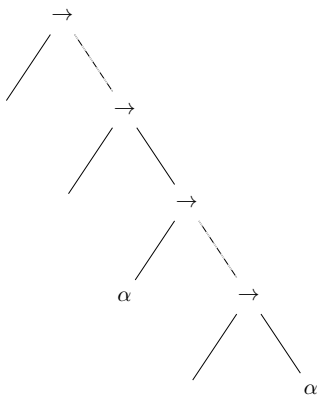



FIGURE 1.6 – *Une tautologie simple dans le système de l'implication.*

## 1.4 Arbre de Galton-Watson

Parallèlement à l'étude de la distribution des arbres de Catalan, s'est développée l'étude d'un modèle introduit par Chauvin et al. ([CFGG04]) inspiré du processus de Galton-Watson. Nous considérons dans cette partie des arbres booléens binaires plans, étiquetés par des variables de l'ensemble $\{x_1, \ldots, x_k\}$.

Considérons un processus critique de Galton-Watson binaire : à l'étape 0, on commence avec une feuille-racine, à l'étape $n$, les feuilles de la $n^{\text{ième}}$ génération meurent sans descendance avec probabilité $\frac{1}{2}$ ou donnent naissance à deux feuilles de $(n+1)^{\text{ième}}$ génération avec probabilité $\frac{1}{2}$, et ce indépendamment les unes des autres. Il est connu que ce processus termine presque sûrement : c'est donc un moyen de générer aléatoirement un arbre de taille presque sûrement finie. Il ne nous reste plus qu'à étiqueter cet arbre **uniformément au hasard**, c'est à dire comme suit : tout nœud interne est étiqueté en choisissant uniformément au hasard parmi les connecteurs proposés par le système logique choisi, chaque feuille est étiquetée en choisissant uniformément au hasard parmi les littéraux proposés par le système logique choisi, l'étiquetage de chaque nœud étant indépendant de tous les autres. Ce procédé définit un arbre booléen aléatoire de loi notée $\Pi_k$ qui induit via $\Phi$ (cf. Définition 1.2.5) une loi sur l'ensemble des fonctions booléennes à $k$ variables $\mathcal{F}_k$ que nous appellerons **loi de Galton-Watson** et que nous noterons $\pi_k$. Cette loi a été étudiée dans les deux systèmes logiques évoqués plus haut : le système et/ou et le système de l'implication. Nous résumons ici les résultats obtenus.

**Théorème 1.4.1 ([FGGG12])**

*Dans le système de l'implication, pour toute fonction booléenne fixée, asymptotiquement quand $k$ tend vers $+\infty$, il existe une constante $\lambda_f > 0$ telle que*

$$\pi_k(f) \sim \frac{\lambda_f}{k^{L(f)}}.$$

Ce théorème n'est démontré que pour le système de l'implication, mais son équivalent dans le système et/ou est communément admis. La démonstration de ce théorème est plus immédiate que dans le cas de la distribution des arbres de Catalan car, asymptotiquement quand $k$ tend vers $+\infty$, *un arbre typique calculant $f$ est un arbre minimal de $f$*. Ce théorème est très comparable à celui obtenu pour la distribution des arbres de Catalan (Théorème 1.3.8). Cette similarité est peut-être due au fait qu'un arbre de Galton-Watson critique conditionné à être de taille $n$ est distribué uniformément

parmi les arbres de Catalan de taille $n$. Les deux modèles induisent des distributions proches sur $\mathcal{F}_k$, et, tout comme la distribution des arbres de Catalan, la distribution de Galton-Watson dans le système de l'implication n'exhibe pas d'effet Shannon (cf. Théorème 1.2.9) :

**Théorème 1.4.2 ([GG10])**

*On dit qu'une fonction booléenne $f$ est* **read-once** *si $L(f) = E(f)$ et on note $\mathcal{R}$ l'ensemble des fonctions read-once. Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\pi_k(\mathcal{R}) \to 1.$$

*Par conséquent,comme les fonction read-once sont de complexité au plus linéaire en $k$, la distribution de Galton-Watson n'exhibe pas d'effet Shannon.*

Dans le modèle de Galton-Watson, tout comme dans celui des arbres de Catalan, les tautologies font l'objet d'une étude poussée, et nous retrouvons un comportement similaire à celui observé dans le cas des arbres de Catalan : asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple :

**Théorème 1.4.3**

*[[GG10]] Dans le système de l'implication, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\pi_k(\texttt{Vrai}) \sim \Pi_k(\mathcal{ST}),$$

*où, pour rappel, $\Pi_k(\mathcal{ST})$ est la probabilité que l'arbre de Galton-Watson étiqueté uniformément au hasard sur $k$ variables soit une tautologie simple.*

Ainsi, au vu de la littérature sur le sujet, il semble que les distributions issues de la représentation arborescente des fonctions booléennes aient toutes un comportement similaire : elle donnent plus de poids aux fonctions de faible complexité, et n'exhibent pas d'effet Shannon. Dans ce mémoire, nous généraliserons ces résultats à d'autres modèles d'arbres, en considérant des arbres non binaires, non planaires (Chapitres 2, 3 et 4), en changeant la loi de l'arbre sous-jacent (Chapitre 5), et en autorisant le nombre de variables $k$ à dépendre de la taille des arbres $n$ (Chapitre 6).

## 1.5 Combinatoire analytique

Dans la lignée des travaux de Chauvin et al. [CFGG04] et Kozik [Koz08], nous utiliserons des méthodes de combinatoire analytique afin d'étudier les arbres booléens. L'objet de cette section est de présenter les bases de ce domaine, ainsi que quelques résultats qui seront utilisés tour au long de ce mémoire. Une introduction plus exhaustive à la combinatoire analytique peut être lue dans le livre de Flajolet et Sedgewick [FS09] : seuls les idées et résultats cruciaux pour la suite sont très rapidement présentés ici.

Étant donnée une suite $(s_n)_{n\geqslant 0}$, il s'agit de définir la série formelle dont les coefficients sont donnés par les éléments de la suite, série que l'on appellera **fonction génératrice** de la suite $(s_n)_{n\geqslant 0}$ :

$$S(z) = \sum_{n\geqslant 0} s_n z^n.$$

Il sera d'usage de noter $[z^n]S(n)$ le $n^{\text{ème}}$ coefficient de $S(z)$, i.e. $s_n$.

Une famille combinatoire peut être décrite par une **spécification**. Par exemple, un arbre de Catalan (i.e. un arbre binaire plan) est soit une feuille, soit un nœud interne dont les deux sous-arbres sont des arbres de Catalan. Si l'on note $\mathcal{T}$ la famille des arbres booléens de Catalan, et si

l'on définit la taille de ces arbres comme étant le nombre de feuilles,

$$\mathcal{T} = \mathcal{L} + \{\wedge, \vee\} \times \mathcal{T} \times \mathcal{T},$$

où $\mathcal{L}$ représente l'ensemble des atomes de taille 1, i.e. l'ensemble des $2k$ feuilles. Dès lors, comme la série génératrice de $\mathcal{L}$ est $2kz$, et celle de $\{\wedge, \vee\}$ est 2, la méthode symbolique nous permet de déduire directement

$$T(z) = 2kz + 2T(z)^2,$$

et, comme $T(0) = 0$, nous en déduisons

$$T(z) = \frac{1}{4}\left(1 - \sqrt{1 - 16kz}\right). \tag{1.2}$$

Le comportement asymptotique de toute suite peut être déduit de celui de sa série génératrice au voisinage de sa (ou plutôt de ses) singularité dominante, i.e. sa singularité de plus petit module, via un lemme de transfert. Il est à noter que si tous les coefficients d'une série génératrice sont positifs, alors sa (ou plutôt une de ses) singularité dominante est réelle positive. Le lemme de transfert que nos utiliserons largement au cours de ce mémoire est le théorème de transfert de Flajolet-Odlyszko [FO90], dont un énoncé peut être lu dans [FS09, page 406]. Ce théorème requiert une hypothèse d'analyticité de la série génératrice étudiée appelée $\Delta$-analyticité, qui assure, entre autres que la série génératrice admet une unique singularité dominante.

Nous dirons que la singularité dominante $\sigma$ de la fonction génératrice $S(z)$ est de **type racine carrée** si, et seulement si, il existe deux fonctions $g(z)$ et $h(z)$, analytiques au voisinage de $\sigma$, telles que, $h(\sigma) \neq 0$ et, pour tout $z$ au voisinage de $\sigma$,

$$S(z) = g(z) + h(z)\sqrt{1 - \frac{z}{\sigma}} + o\left(\sqrt{1 - \frac{z}{\sigma}}\right).$$

Comme nous travaillons sur des arbres enracinés, la plupart des singularités dominantes que nous rencontrerons dans ce cadre seront de type racine-carrée. Par exemple, la série génératrice des arbres de Catalan a une singularité dominante $\sigma = \frac{1}{16k}$ de type racine-carrée (cf. Équation (1.2)). Quitte à vérifier l'hypothèse de $\Delta$-analyticité, nous pouvons en déduire, via le lemme de transfert de Flajolet-Odlyszko qu'il existe une constante $c$ strictement positive telle que, asymptotiquement quand $n$ tend vers $+\infty$,

$$T_{n,k} \sim cn^{-3/2}\left(\frac{1}{16k}\right)^n.$$

Le lemme de transfert de Flajolet-Odlyszko permet de démontrer le lemme suivant, que nous utiliserons largement au cours de ce mémoire :

**Lemme 1.5.1**

*Étant données deux séries génératrices $S(z)$ et $R(z)$, de même singularité dominante et ayant le même comportement au voisinage de leur singularité dominante (on dira que les deux singularités sont de même type), alors, si $R$ et $S$ sont $\Delta$-analytiques au voisinage de leur singularité dominante,*

$$\lim_{n\to+\infty} \frac{[z^n]R(z)}{[z^n]S(z)} = \lim_{z\to\sigma} \frac{R'(z)}{S'(z)}.$$

Ce résultat sera très utile pour calculer ce que l'on appelle des proportions limites de sous-familles d'arbres. Soit $\mathcal{S}$ une famille d'arbres de série génératrice $S(z)$, et soit $\mathcal{R}$ une sous-famille de $\mathcal{S}$ de

série génératrice $R(z)$. Nous appellerons **proportion limite** (ou **fraction limite**) de la famille $\mathcal{R}$ la limite quand $n$ tend vers $+\infty$ de la proportion d'arbres de $\mathcal{R}$ parmi les arbres de $\mathcal{S}$ de taille $n$, si elle existe. Et nous noterons

$$\mu(\mathcal{R}) = \lim_{n \to +\infty} \frac{[z^n]R(z)}{[z^n]S(z)}.$$

Cette notation ne mentionne pas la famille $\mathcal{S}$, car le choix de celle-ci sera en général non-ambigu.

# Arbres booléens généraux

# Chapitre 2

# Arbres et/ou généraux

## 2.1 Introduction

L'objectif de ce chapitre est de définir et étudier la distribution induite sur l'ensemble des fonctions booléennes à $k$ variables $\mathcal{F}_k$ par la distribution des arbres de Catalan généralisée à des arbres et/ou non-binaires et non-planaires. Cette généralisation a pour but de prendre en compte les propriétés logiques des connecteurs $\wedge$ et $\vee$, c'est à dire leur associativité et leur commutativité. Nous allons étudier trois nouveaux modèles : le modèle des arbres non-binaires plans, appelés arbres associatifs, le modèle des arbres non-plans binaires, appelés arbres commutatifs, et le modèle des arbres généraux, associatifs et commutatifs, i.e. non-binaires non-plans. Cette étude s'inspire d'idées introduites dans le survey de Gardy [Gar06], les méthodes utilisées seront la combinatoire analytique, et notamment le comptage de Pólya pour les arbres non planaires (cf. Pólya et Reed [PR87]).

Dans ces trois nouveaux modèles, tout comme dans le modèle des arbres Catalan binaires plans présenté en introduction (cf. Section 1.3.1), la taille d'un arbre sera le nombre de ses feuilles. Garder la même notion de taille dans les quatre modèles permet de rendre les comparaisons possibles entre les différents résultats que nous obtiendrons. Dans chaque modèle, nous considérons la loi uniforme sur l'ensemble des arbres et/ou de taille $n$, et montrons que la suite de distributions induites sur $\mathcal{F}_k$ converge quand $n$ tend vers $+\infty$ vers une distribution asymptotique notée $\mu_k$. Nous étudions ensuite chacune des trois nouvelles distributions ainsi définies et la comparons à la distribution des arbres de Catalan (cf. Section 1.3.1).

Via des méthodes d'analyse de singularités de fonctions génératrices, nous montrons que prendre en compte l'associativité et la commutativité des connecteurs $\wedge$ et $\vee$ ne change pas le comportement global de la distribution induite sur $\mathcal{F}_k$, même si une analyse précise des constantes mises en jeu nous assure que les quatre distributions étudiées sont deux à deux distinctes (cf. Tableau 2.10).

La Section 2.2 définit les trois nouvelles distributions étudiées dans ce chapitre et montre, via le théorème de Drmota-Lalley-Woods [Drm97, Lal93, Woo97] la convergence vers une distribution asymptotique quand la taille $n$ des arbres tend vers $+\infty$ : cette partie est donc l'occasion de définir les objets étudiés et les différentes fonctions génératrices qui seront utiles dans la suite du chapitre. La Section 2.3 se concentre sur la fonction constante `Vrai` et étudie donc la fraction limite des arbres tautologiques. Nous verrons tout au long de cette partie Arbres booléens aléatoires que l'étude des tautologies est, non seulement une étude du cas particulier simple des fonctions de complexité 0, mais aussi la première étape de l'étude de la probabilité d'une fonction booléenne générale. Nous montrerons au passage, que, dans les trois nouveaux modèles, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est *simple*. Cette étude des tautologies nécessite la généralisation de la théorie des motifs et du lemme de Kozik 1.3.5 à nos modèles d'arbres non-binaires et non-plans.

Si généraliser cette théorie à des arbres non-binaires est assez direct, la généralisation à des arbres non-plans l'est moins car les motifs sont par essence plans. La Section 2.4 concerne l'étude des $2k$ fonctions de complexité 1 : cette étude est anecdotique, et les détails ne seront pas développés, mais l'étude des fonctions littéral permet d'avoir une meilleure intuition du comportement des arbres typiques calculant une fonction booléenne quelconque. Le cas général est donc traité en Section 2.5, grâce aux résultats sur les tautologies et grâce aux généralisations de la théorie des motifs de Kozik.

## 2.2 Définitions et préliminaires

### 2.2.1 Arbres associatifs, plans

**Définition 2.2.1**

*Un* **arbre associatif** *est un arbre dans lequel chaque nœud interne a pour arité un entier de* $\mathbb{N}\backslash\{1\}$*. Un* **arbre booléen associatif** *est un arbre associatif dont les nœuds internes sont étiquetés par* $\wedge$ *ou par* $\vee$ *de telle façon qu'un nœud interne ne peut avoir la même étiquette que son père, et dont les feuilles sont étiquetées par des littéraux choisis dans* $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$*. On notera* $\mathcal{A}_k$ *la famille des arbres booléens associatifs à* $k$ *variables,* $\mathcal{A}_{n,k}$ *la famille des arbres booléens associatifs de taille* $n$*,* $A_{n,k}$ *son cardinal, pour tout* $n \geqslant 1$ *et* $A(z) = \sum_{n\geqslant 1} A_{n,k} z^n$ *sa série génératrice.*

On dira que ces arbres booléens sont *stratifiés*, au sens où l'étiquette de la racine détermine les étiquettes de tous les nœuds de l'arbre : tous les nœuds internes d'une même génération ont la même étiquette.

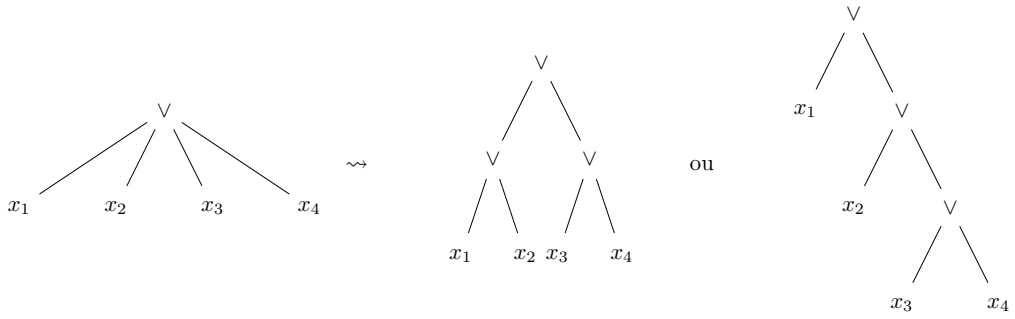



FIGURE 2.1 – *Deux équivalents binaires d'un même arbre associatif.*

**Définition 2.2.2**

*Pour toute fonction booléenne* $f \in \mathcal{F}_k$*, pour tout* $n \geqslant 1$*, on note* $A_{n,k}(f)$ *le nombre d'arbres booléens associatifs de taille* $n$ *calculant* $f$ *et on définit*

$$\mu^a_{n,k}(f) = \frac{A_{n,k}(f)}{A_{n,k}}$$

*comme étant la proportion d'arbres de taille* $n$ *calculant* $f$*.*

Dans cette partie préliminaire, nous allons montrer que cette distribution sur $\mathcal{F}_k$ converge vers une distribution asymptotique quand la taille $n$ des arbres considérés tend vers $+\infty$. Cette première étape nous permettra d'introduire les différentes fonctions génératrices qui nous seront utiles dans tout le reste du chapitre.

**Lemme 2.2.3**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$,*

$$\mu_k^a(f) = \lim_{n\to+\infty} \mu_{n,k}^a(f)$$

*existe et est strictement positive.*

Notons $\hat{\mathcal{A}}$ ($\check{\mathcal{A}}$) la famille des arbres associatifs réduits à une feuille, ou dont la racine est étiquetée par le connecteur $\wedge$ (resp. $\vee$), et notons $\hat{A}(z)$ (resp. $\check{A}(z)$) la fonction génératrice de cette famille. Notons que, par définition, ces deux familles contiennent les $2k$ arbres réduits à une feuille. Par la méthode symbolique, nous avons immédiatement $A(z) = \hat{A}(z) + \check{A}(z) - 2kz$, où le $-2kz$ permet de ne pas double-compter les $2k$ arbres de taille 1. Comme $\hat{A}(z) = \check{A}(z)$, on a

$$\hat{A}(z) = 2kz + \sum_{\ell\geqslant 2} \check{A}(z)^\ell = 2kz + \frac{\hat{A}^2(z)}{1-\hat{A}(z)}.$$

Cela implique que

$$A(z) = \frac{1}{2}\left(1 - 2kz - \sqrt{1-12kz+4k^2z^2}\right) \tag{2.1}$$

a pour singularité dominante

$$\alpha_k = \frac{3-2\sqrt{2}}{2k}. \tag{2.2}$$

De plus, $A(\alpha_k) = \sqrt{2}-1$.

**Démonstration du Lemme 2.2.3 :** Nous allons utiliser le théorème de Drmota-Lalley-Woods (cf. [FS09, Chapitre 8]) appliqué aux fonctions génératrices $\hat{A}_f(z)$ et $\check{A}_f(z)$ des arbres associatifs enracinés par $\wedge$ (resp. $\vee$) calculant $f$. Ces fonctions génératrices sont au nombre de $2^{2^k+1}$ et vérifient, via la méthode symbolique (cf [FS09] pour une introduction à la méthode symbolique), le système suivant :

$$\begin{cases} \hat{A}_f(z) = z\mathbb{1}_{\{f \text{ lit.}\}} + \displaystyle\sum_{i=2}^{\infty} \sum_{\substack{g_1,\dots,g_i,\\ g_1\wedge\cdots\wedge g_i = f}} \check{A}_{g_1}(z)\cdots\check{A}_{g_i}(z) \\ \check{A}_f(z) = z\mathbb{1}_{\{f \text{ lit.}\}} + \displaystyle\sum_{i=2}^{\infty} \sum_{\substack{g_1,\dots,g_i,\\ g_1\vee\cdots\vee g_i = f}} \hat{A}_{g_1}(z)\cdots\hat{A}_{g_i}(z). \end{cases}$$

Les fonctions $(\hat{A}_f, \check{A}_f)_{f\in\mathcal{F}_k}$ sont donc solutions d'un système algébrique non linéaire, qui vérifie les hypothèses du théorème de Drmota-Lalley-Woods (cf. [FS09, page 489]). Nous en déduisons que les fonctions $(\hat{A}_f, \check{A}_f)_{f\in\mathcal{F}_k}$, ainsi que leur somme $A(z)$, ont la même singularité $\alpha_k$, que cette singularité est de type racine-carrée, et que toutes ces fonctions sont $\Delta$-analytiques (cf. [FS09, page 389]). Nous pouvons donc appliquer le lemme de transfert de Flajolet et Odlyszko [FO90] : pour toute fonction booléenne $f$ de $\mathcal{F}_k$, il existe deux constantes $c_k(f), d_k(f) > 0$ telles que, asymptotiquement quand $n$ tend vers $+\infty$,

$$[z^n]\hat{A}_f \sim c_k(f) n^{-3/2}\alpha_k^{-n} \quad \text{et} \quad [z^n]\check{A}_f \sim d_k(f) n^{-3/2}\alpha_k^{-n}.$$

Dès lors, il existe une constante $cst = \sum_{f\in\mathcal{F}_k}(c_k(f)+d_k(f))$ telle que, asymptotiquement quand $n$ tend vers $+\infty$,

$$A_{n,k} = [z^n]A(z) \sim cst \cdot n^{-3/2}\alpha_k^{-n}.$$

Donc, asymptotiquement quand $n$ tend vers $+\infty$, pour toute fonction $f \in \mathcal{F}_k$,

$$\frac{A_{n,k}(f)}{A_{n,k}} \sim \frac{c_k(f)+d_k(f)}{cst},$$

ce qui implique immédiatement le Lemme 2.2.3. ∎

### 2.2.2 Arbres commutatifs, binaires

**Définition 2.2.4**

*Un* **arbre booléen commutatif** *est un arbre binaire non planaire dont les nœuds internes sont étiquetés par un connecteur $\wedge$ ou $\vee$ et dont les feuilles sont étiquetées par des littéraux de $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$. On notera $\mathcal{C}$ la famille de ces arbres, et $C_{n,k}$ le nombre d'arbres commutatifs de taille $n$ à $k$ variables.*

Les arbres binaires commutatifs vérifient la spécification suivante :

$$\mathcal{C} = 2k\mathcal{Z} + 2\{\mathcal{C}, \mathcal{C}\},$$

où la notation $\{\mathcal{C}, \mathcal{C}\}$ représente un multi-ensemble de deux éléments de $\mathcal{C}$. Cette spécification se traduit par l'équation suivante sur la fonction génératrice de ces arbres (pour une introduction au comptage de Pólya, le lecteur pourra se référer à [Kra11] ou [FS09]) :

$$C(z) = 2kz + C(z)^2 + C(z^2). \tag{2.3}$$

Cette équation ne peut permettre de trouver une expression explicite de la fonction génératrice $C(z)$. Toute la difficulté sera donc d'extraire un maximum d'informations de cette expression implicite.

**Lemme 2.2.5**

*Le rayon de convergence $\gamma_k$ de $C(z)$ vérifie :*

$$\gamma_k = \frac{1}{8k}\left(1 - \frac{1}{8k}\right) + \mathcal{O}\left(\frac{1}{k^3}\right).$$

**Démonstration :** Pour étudier la singularité de la solution de l'Équation (2.3), réécrivons-la sous la forme $F(z, C(z)) = C(z)$ où $F(X, Y) = 2kX + Y^2 + Q(X)$ et $Q(z) = C(z^2)$. Un résultat de Bender [Ben74] (cf. [Drm09, Théorème 2.19]) que l'on peut appliquer ici, en particulier car $Q(z) = C(z^2)$ est analytique sur le domaine d'analyticité de $C$, nous assure que le couple $(\gamma_k, C(\gamma_k))$ est solution du système

$$\begin{cases} F(X, Y) = Y \\ \partial_Y F(X, Y) = 1, \end{cases}$$

et donc de

$$\begin{cases} Y = 2kX + Y^2 + Q(X) \\ 1 = 2Y. \end{cases}$$

Dès lors, $C(\gamma_k) = \frac{1}{2}$ et $2k\gamma_k = \frac{1}{4} - Q(\gamma_k)$, ce qui, comme $Q$ est à coefficients positifs, implique $\gamma_k \leqslant \frac{1}{8k}$. De plus,

$$C(\gamma_k^2) = \sum_{n \geqslant 1} C_{n,k}\gamma_k^{2n} = \gamma_k \sum_{n \geqslant 1} C_{n,k}\gamma_k^{2n-1} \leqslant \frac{1}{8k}C(\gamma_k) = \frac{1}{16k},$$

ce qui implique donc bien que $\gamma_k = \frac{1}{8k} + \mathcal{O}\left(\frac{1}{k^2}\right)$. Le second ordre est obtenu en réinjectant ce premier développement asymptotique dans le système vérifié par $(\gamma_k, C(\gamma_k))$. ■

**Définition 2.2.6**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$, on notera $C_{n,k}(f)$ le nombre d'arbres commutatifs de*

*taille $n$ calculant $f$ et $C_f(z)$ la fonction génératrice de cette suite d'entiers. On notera*

$$\mu^c_{n,k}(f) = \frac{C_{n,k}(f)}{C_{n,k}}.$$

**Lemme 2.2.7**

*Pour toute fonction $f$,*

$$\mu^c_k(f) = \lim_{n\to\infty} \mu^c_{n,k}(f)$$

*existe et est strictement positive.*

**Démonstration :** Via la méthode symbolique, les fonctions génératrices $\{C_f(z)\}_{f\in\mathcal{F}_k}$ vérifient le système suivant :

$$C_f(z) = z\mathbb{1}_{\{f \text{ lit.}\}} + \frac{1}{2}\sum_{\substack{g,h\neq f\\ g\wedge h=f}} C_g(z)C_h(z) + \frac{1}{2}\sum_{\substack{g,h\neq f\\ g\vee h=f}} C_g(z)C_h(z) + C_f(z)^2 + C_f(z^2).$$

Ce système vérifie les hypothèses du théorème de Drmota-Lalley-Woods qui permet donc de conclure la preuve du Lemme 2.2.7 : les fonctions génératrices $(C_f(z))_{f\in\mathcal{F}_k}$ et $C(z)$ ont toutes le même rayon de convergence $\gamma_k$, et $\mu^c_{n,k}(f)$ converge vers un entier strictement positif quand $m$ tend vers $+\infty$. ■

### 2.2.3 Arbres associatifs et commutatifs

**Définition 2.2.8**

*Un* **arbre booléen général** *est un arbre non plan dont chaque nœud interne a pour arité un entier supérieur ou égal à 2, est étiqueté par un connecteur $\wedge$ ou $\vee$ différent de l'étiquette de son père (l'arbre est stratifié), et dont chaque feuille est étiquetée par un littéral de $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$. On note $\mathcal{P}$ cette famille d'arbres, $P(z)$ sa fonction génératrice, et $P_{n,k}$ le nombre de tels arbres de taille $n$.*

Tout comme dans le cas des arbres associatifs, la fonction génératrice $P(z)$ s'exprime en fonction des fonctions génératrices $\hat{P}(z)$ et $\check{P}(z)$ des arbres généraux dont la racine est étiquetée par un connecteur $\wedge$ (resp. $\vee$) :

$$P(z) = \hat{P}(z) + \check{P}(z) - 2kz = 2\hat{P}(z) - 2kz.$$

De plus, via la méthode symbolique, les fonctions $\hat{P}(z)$ et $\check{P}(z)$ vérifient la relation suivante :

$$\begin{cases} \hat{P}(z) = \exp\left(\displaystyle\sum_{i\geqslant 1} \frac{\check{P}(z^i)}{i}\right) - 1 - \check{P}(z) + 2kz \\ \check{P}(z) = \exp\left(\displaystyle\sum_{i\geqslant 1} \frac{\hat{P}(z^i)}{i}\right) - 1 - \hat{P}(z) + 2kz, \end{cases}$$

que l'on peut simplifier en remarquant que $\hat{P}(z) = \check{P}(z)$ :

$$\hat{P}(z) = \exp\left(\sum_{i\geqslant 1} \frac{\hat{P}(z^i)}{i}\right) - 1 - \hat{P}(z) + 2kz. \tag{2.4}$$

Encore une fois, nous ne pouvons déduire d'expression explicite de la fonction génératrice $P(z)$, mais cette relation implicite nous permettra de rassembler suffisamment d'informations, notamment au sujet de sa singularité.

**Lemme 2.2.9**

*La singularité de $P(z)$ vérifie :*

$$\delta_k = \frac{2\ln 2 - 1}{2k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

**Démonstration :** Pour montrer ce lemme, nous procédons comme dans la preuve du Lemme 2.2.5. L'Équation (2.4) peut s'écrire $\hat{P}(z) = F(z, \hat{P}(z))$ où $F(X,Y) = Q(X)\mathsf{e}^Y - 1 - Y + 2kX$ et $Q(z) = \exp\left(\sum_{i\geqslant 2} \frac{\hat{P}(z^i)}{i}\right)$. Comme $\delta_k < 1$, la fonction $Q$ est analytique sur le domaine d'analyticité de $\hat{P}$, et nous pouvons donc conclure que $(\delta_k, \hat{P}(\delta_k))$ est solution du système suivant :

$$\begin{cases} Y = Q(X)\mathsf{e}^Y - 1 - Y + 2kX \\ 1 = Q(X)\mathsf{e}^Y - 1. \end{cases}$$

Autrement dit,

$$\begin{cases} 2Y = 1 + 2kX \\ Q(X)\mathsf{e}^Y = 2. \end{cases}$$

Supposons que $(X,Y)$ soit solution de ce système, et que $X, Y \geqslant 0$. Par définition, $Q(X) \geqslant 1$. Donc,

$$2 = Q(X)\mathsf{e}^Y \geqslant \mathsf{e}^Y = \mathsf{e}^{kX+1/2},$$

ce qui implique que $X = \mathcal{O}\left(\frac{1}{k}\right)$ quand $k$ tend vers $+\infty$, et donc $Y = \mathcal{O}(1)$ quand $k$ tend vers $+\infty$. De plus, notons que pour tout entier $i \geqslant 2$,

$$P(X^i) = \sum_{n\geqslant 1} P_{n,k} X^{in} = X^{i-1} \sum_{n\geqslant 1} P_{n,k} X^{i(n-1)+1} \leqslant X^{i-1} \sum_{n\geqslant 1} P_{n,k} X^n = X^{i-1} P(X).$$

Nous avons donc, asymptotiquement quand $k$ tend vers $+\infty$,

$$\ln Q(X) = \sum_{i\geqslant 2} \frac{P(X^i)}{i} \leqslant P(X) \sum_{i\geqslant 2} \frac{X^{i-1}}{i} = \frac{Y}{X} \sum_{i\geqslant 2} \frac{X^i}{i} = \frac{Y}{X}(-\ln(1-X) - X) \sim \frac{Y}{X}\frac{X^2}{2} = \frac{YX}{2} \to 0.$$

Autrement dit, $Q(X) \sim 1$ et $Y \sim \ln 2$ quand $k$ tend vers $+\infty$, et $X \sim \frac{2\ln 2 - 1}{2k}$. Dès lors,

$$\ln Q(X) = \frac{YX}{2} = \mathcal{O}\left(\frac{1}{k}\right),$$

which implies

$$X = \frac{2\ln 2 - 1}{2k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

■

**Définition 2.2.10**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$, on note $P_{n,k}(f)$ le nombre d'arbres généraux qui calculent $f$, et $P_f(z)$ la fonction génératrice de cette suite. On définit*

$$\mu^g_{n,k}(f) = \frac{P_{n,k}(f)}{P_{n,k}}.$$

**Lemme 2.2.11**

*Pour toute fonction booléenne $f$,*

$$\mu^g_k(f) = \lim_{n\to+\infty} \mu^g_{n,k}(f)$$

*existe et est strictement positive.*

**Démonstration :** Si l'on note $\hat{P}_f(z)$ et $\check{P}_f(z)$ les fonctions génératrices des arbres généraux enracinés par $\wedge$ (resp. $\vee$) calculant $f$, via la méthode symbolique,

$$\begin{cases} \hat{P}_f(z) = z\mathbb{1}_{\{f \text{ lit.}\}} + \displaystyle\sum_{\ell=2}^{\infty} \sum_{\substack{g_1,\dots g_i,\\ g_1\wedge\dots\wedge g_\ell = f}} \prod_{j=1}^{\ell} \left( \exp\left( \sum_{i\geqslant 1} \frac{\check{P}_{g_j}(z^i)}{i} \right) - 1 \right) \\ \check{P}_f(z) = z\mathbb{1}_{\{f \text{ lit.}\}} + \displaystyle\sum_{\ell=2}^{\infty} \sum_{\substack{g_1,\dots g_i,\\ g_1\wedge\dots\wedge g_\ell = f}} \prod_{j=1}^{\ell} \left( \exp\left( \sum_{i\geqslant 1} \frac{\hat{P}_{g_j}(z^i)}{i} \right) - 1 \right). \end{cases}$$

Ce système vérifie les hypothèses du théorème de Drmota-Lalley-Woods qui permet donc de conclure la preuve : les fonctions génératrices $(P_f(z))_{f\in\mathcal{F}_k}$ et $P(z)$ ont la même singularité $\delta_k$, et toutes ces singularités sont de type racine-carrée. ■

### 2.2.4 Comportement des différents modèles

Nous avons montré dans cette partie l'existence des distributions limites qui seront l'objet de ce chapitre. Notre objectif est désormais de montrer que le Théorème 1.3.1 est vérifié par ces trois nouvelles distributions : autrement dit, associativité et commutativité des connecteurs n'influent pas sur le comportement global de la distribution induite sur l'ensemble des fonctions booléennes $\mathcal{F}_k$.

La première partie de ce chapitre (Section 2.3) sera consacrée à l'étude de la fonction `Vrai`, donc des tautologies : nous montrerons que la probabilité de `Vrai` est, dans tous les cas, d'ordre $1/k$ quand $k$ tend vers $+\infty$. Nos calculs seront suffisamment précis pour affirmer que $\mu_k(\texttt{Vrai})$ n'a pas la même valeur selon les propriétés logiques prises en compte. Dans la Section 2.4, nous montrerons rapidement que la probabilité des fonctions littéral est d'ordre $1/k^2$ quand $k$ tend vers $+\infty$, avant de généraliser à une fonction booléenne quelconque en Section 2.5. Nous montrerons que dans chaque modèle, tous comme dans le modèle des arbres de Catalan, pour toute fonction booléenne $f \in \mathcal{F}_k$ dont le nombre de variables essentielles ne dépend pas de $f$, il existe une constante $\lambda_f$ telle que

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

## 2.3 Tautologies

L'objectif de cette section est de calculer la probabilité de la fonction constante `Vrai`, autrement dit, la proportion limite des tautologies. Dans chaque modèle d'arbre, la stratégie est la même que dans le cas binaire planaire. Nous montrons tout d'abord que, asymptotiquement quand $n$ tend vers $+\infty$, presque toute tautologie est simple (cf. Définition 1.3.6), puis nous calculons la fraction limite des tautologies simples. Pour cette étude, nous aurons besoin de généraliser le lemme de Kozik (cf. Lemme 1.3.5).

### 2.3.1 Arbres associatifs

Dans cette partie consacrée à l'étude des tautologies associatives, nous allons montrer le résultat suivant :

**Proposition 2.3.1**

*Dans le modèle associatif, la probabilité de la fonction constante* `Vrai` *vérifie, asymptotiquement*

*quand $k$ tend vers $+\infty$,*

$$\mu_k^a(\texttt{Vrai}) = \frac{51 - 36\sqrt{2}}{k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

Pour montrer cette proposition, nous allons généraliser le Lemme 1.3.5 à nos arbres associatifs. Cette généralisation est sans difficultés, mais nous détaillons sa preuve pour information, car nous aurons besoin de reprendre cette preuve plus attentivement dans le cas des arbres non plans. Le lemme de Kozik nous permettra ensuite, tout comme dans le cas binaire plan (cf. Chapitre 1, Section 1.3.1), de montrer que, asymptotiquement quand $k$ tend vers $+\infty$, *presque toute tautologie est simple*, puis de calculer la fraction limite des tautologies simples dans le modèle de l'implication.

### Généralisation du lemme de Kozik

Dans cette partie, nous appellerons **squelette associatif** un arbre enraciné, plan, dont les nœuds internes ont au moins deux descendants, sont étiquetés par $\wedge$ ou $\vee$ de façon à être stratifiés, et dont les feuilles sont non étiquetées. On notera $\mathcal{S}$ la famille de ces arbres, $S_n$ le nombre de squelettes de taille $n$, et $S(z)$ leur fonction génératrice. La propriété de stratification des arbres associatifs nous oblige à revoir quelques définitions liées aux langages de motifs : en effet, lorsque l'on greffe un arbre dans un emplacement d'un motif, il faut que cet arbre soit enraciné par le connecteur qui assure la stratification. Ainsi, on notera abusivement $L[\mathcal{A}]$ l'ensemble des arbres obtenus en greffant dans un motif de $L$ des arbres de $\mathcal{A}$ de telle sorte que tout arbre de taille supérieure ou égale à 3 greffé dans un emplacement est enraciné par $\wedge$ (resp. $\vee$) si son parent est $\vee$ (resp. $\wedge$). Dès lors, nous dirons que le langage de motif $L$ est sous-critique pour la famille des squelettes $\mathcal{S}$ si, et seulement si, $L$ est sous-critique pour la famille $\hat{\mathcal{S}}$ des squelettes associatifs enracinés par $\wedge$ ou de taille 1.

Nous reprenons les définitions de la Section 1.3.1 concernant la théorie des motifs.

**Lemme 2.3.2**

*Soit $L$ un langage de motifs, non-ambigu, sous-critique pour la famille $\mathcal{S}$, et tel que $L[S] = S$. Soit $\Gamma$ un sous-ensemble de $\{x_1, \dots, x_k\}$ dont le cardinal ne dépend pas de $k$.*
*On note $A_{n,k}^{[p]}$ (resp. $A_{n,k}^{[\geqslant p]}$) le nombre d'arbres booléens associatifs de taille $n$ ayant exactement (resp. au moins) $p$ $(L,\Gamma)$-restrictions. Alors, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to\infty} \frac{A_{n,k}^{[p]}}{A_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right) \quad \textit{et} \quad \lim_{n\to\infty} \frac{A_{n,k}^{[\geqslant p]}}{A_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right).$$

La suite de cette sous-section est consacrée à la preuve de ce lemme. Cette preuve est quasi identique à celle développée dans [Koz08] pour le modèle plan binaire.

On notera $\gamma$ le cardinal de $\Gamma$. Soit $\ell, n \geqslant 1$ deux entiers, et soit $t$ un squelette associatif de taille $n$ ayant $\ell$ $L$-feuilles de motif. Pour tout entier $r \leqslant p$, le nombre d'étiquetages différents des $n$ feuilles de $t$ qui réalisent exactement $r$ $L$-répétitions et $p$ $(L,\Gamma)$-restrictions est donné par

$$\left\{ {\ell \atop \ell - r} \right\} \binom{\gamma}{p-r} (\ell - r)^{\underline{p-r}} (k-\gamma)^{\underline{\ell - r - (p-r)}} k^{n-\ell} 2^n,$$

où $x^{\underline{y}} = x(x-1)\dots(x-y+1)$, $\left\{ {y \atop x} \right\}$ sont les nombres de Stirling de seconde espèce [1]. Détaillons

1. Pour tout couple d'entiers $(x, y)$ le nombre de Stirling de seconde espèce $\left\{ {x \atop y} \right\}$ est égal au nombre de partitions d'un ensemble à $x$ éléments en $y$ parties non vides.

un peu cette formule : de gauche à droite, les différents facteurs représentent :
- le nombre de partitions des $L$-feuilles de motifs en $\ell - r$ parties non vides : les feuilles d'une même classe seront étiquetées par la même variable,
- le nombre de choix différents pour les $p - r$ variables de $\Gamma$ qui apparaîtront dans les $L$-feuilles de motif de $t$,
- le nombre de façons d'associer ces $p - r$ variables à $p - r$ parties parmi les $\ell - r$ du premier terme,
- le nombre de façons d'associer des variables de $\{x_1, \dots, x_k\}\backslash\Gamma$ aux parties restantes,
- le nombre de façons d'étiqueter les feuilles de $t$ qui ne sont pas des $L$-feuilles de motif,
- le nombre de choix pour le signe de chaque littéral, sur chaque feuille de $t$.

Dès lors, le lemme suivant est immédiat :

**Lemme 2.3.3**

*Soit $t$ un squelette associatif ayant $\ell$ feuilles de motif. Le nombre d'étiquetages différents des feuilles de $t$ tels que $t$ admet $p$ $(L, \Gamma)$-restrictions est donné par*

$$(k - \gamma)^{\underline{\ell - p}}\, k^{n-\ell}\, 2^n w_{\gamma,p}(\ell),$$

*où $w_{\gamma,p}(\ell) = \sum_{r=0}^{k} \left\{ {\ell \atop \ell - r} \right\} \binom{\gamma}{p - r} (\ell - r)^{\underline{p-r}}$ est un polynôme en $\ell$ de degré $p$.*

De plus, énonçons le résultat suivant, dont la preuve, développée dans [Koz08, Lemma 2.7], est omise. Nous l'énonçons dans un cadre plus général que celui des arbres associatifs, de façon à pouvoir l'utiliser ultérieurement.

**Proposition 2.3.4**

*Soit $\mathcal{S}$ une famille de squelettes dont la fonction génératrice $S(z) = \sum_{n\geqslant 1} S_n z^n$ admet une unique singularité dominante $\sigma > 0$. On suppose que cette singularité est de type racine carrée. Soit $L$ un langage de motifs non ambigu et sous-critique pour $\mathcal{S}$, et tel que $\mathcal{S} = L[\mathcal{S}]$. On notera $L[\mathcal{S}]_n(\ell)$ le nombre de squelettes de $L[\mathcal{S}]$ de taille $n$ ayant $\ell$ $L$-feuilles de motif. Enfin, soit $w$ un polynôme non nul de degré $\delta$.*

*Alors, il existe une constante $c_w \geqslant 0$ telle que*

$$\lim_{n\to\infty} \frac{\sum_{\ell\geqslant 0} L[\mathcal{S}]_n(\ell) w(\ell)}{S_n} = c_w.$$

*De plus, si $w(\ell)$ est à valeurs positives, s'il existe $\ell_0 \geqslant \delta$ tel que $w(\ell_0) > 0$ et tel qu'il existe un motif dans $L$ ayant $\ell_0$ feuilles de motif et au moins un emplacement, alors $c_w \neq 0$.*

Le Lemme 2.3.3 et la Proposition 2.3.4 vont nous permettre de conclure la preuve du Lemme 2.3.2 :

**Démonstration du Lemme 2.3.2 :** Soit $L$ un langage de motifs. Au vu du Lemme 2.3.3, et comme $\mathcal{S} = L[\mathcal{S}]$,

$$\frac{A^{[p]}_{n,k}}{A_{n,k}} = \frac{2^n \sum_{\ell\geqslant 0} L[\mathcal{S}]_n(\ell) w_{p,\gamma}(\ell)(k - \gamma)^{\underline{\ell-p}} k^{n-\ell}}{A_{n,k}},$$

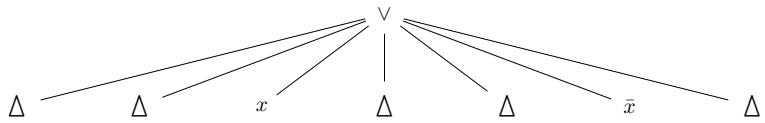
et donc,

$$\frac{A^{[p]}_{n,k}}{A_{n,k}} \leqslant \frac{2^n \sum_{\ell\geqslant 0} L[\mathcal{S}]_n(\ell) w_{p,\gamma}(\ell) k^{\ell-p} k^{n-\ell}}{(2k)^n S_n}.$$

Grâce à la Proposition 2.3.4, nous obtenons donc, asymptotiquement quand $n$ tend vers $+\infty$,

$$\lim_{n\to+\infty} \frac{A^{[p]}_{n,k}}{A_{n,k}} \leqslant \lim_{n\to\infty} \frac{2^n \sum_{\ell\geqslant 0} L[\mathcal{S}]_n(\ell) w_{p,\gamma}(\ell) k^{n-p}}{(2k)^n S_n} \sim \frac{c_{p,\gamma}}{k^p},$$

FIGURE 2.2 – *Une tautologie simple associative.*




où $c_{p,\gamma} = c_{w_{p,\gamma}}$ avec la notation de la Proposition 2.3.4. De plus, nous pouvons aisément vérifier que toutes les conditions sont réunies pour que $c_{p,\gamma}$ soit strictement positive (cf. Proposition 2.3.4).

Enfin, remarquons que

$$\frac{A_{n,k}^{[\geqslant p]}}{A_{n,k}} \leqslant \frac{2^n \sum_{\ell \geqslant 0} L[\mathcal{S}]_n(\ell) w_{p,\gamma}(\ell) k^{n-p}}{A_{n,k}},$$

ce qui implique, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{n\to+\infty} \frac{A_{n,k}^{[\geqslant p]}}{A_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right),$$

ce qui termine donc la preuve du Lemme 2.3.2. ■

**Tautologies associatives**

Rappelons qu'une tautologie simple dans le modèle et/ou (cf. Définition 1.3.6) est un arbre dans lequel deux feuilles distinctes sont reliées à la racine par deux chemins de $\vee$, et étiquetées par une variable et sa négation. Dans le modèle associatif, comme les arbres sont stratifiés, toute tautologie simple devient un arbre booléen dont la racine est étiquetée par un $\vee$, et dont deux feuilles du premier niveau sont étiquetées par une variable et sa négation (cf. Figure 2.2).

**Proposition 2.3.5**

*On notera $ST_{n,k}$ le nombre de tautologies simples de taille $n$. Alors, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty} \frac{ST_{n,k}}{A_{n,k}} = \frac{51 - 36\sqrt{2}}{k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

On dira qu'une tautologie est réalisée par la variable $x$ s'il existe deux feuilles de la première génération étiquetées respectivement par $x$ et sa négation. Bien entendu, une tautologie simple peut être réalisée par plusieurs variables simultanément. Soit $ST^x(z) = \sum_{n\geqslant 1} ST_n^x z^n$ la fonction génératrice des tautologies simples réalisées par $x$.

**Lemme 2.3.6**

*Pour toute variable $x \in \{x_1, \ldots, x_k\}$, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\mathcal{ST}^x) = \lim_{n\to+\infty} \frac{ST_n^x}{A_{n,k}} = \frac{51 - 36\sqrt{2}}{k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

**Démonstration :** Une tautologie simple réalisée par la variable $x$ est une racine étiquetée par $\vee$, dont les premiers sous-arbres, avant la première occurrence de $x$ (contribution $z$) sont des éléments de $\hat{\mathcal{A}}$

qui ne sont ni une feuille étiquetée par $x$, ni une feuille étiquetée par $\bar{x}$, dont les sous-arbres entre la première occurrence de $x$ et la première occurrence de $\bar{x}$ (contribution $z$) sont des éléments de $\hat{\mathcal{A}}$ qui ne sont pas une feuille étiquetée par $\bar{x}$, et dont les derniers sous-arbres sont des éléments de $\hat{\mathcal{A}}$ quelconques. Le tout multiplié par deux car $\bar{x}$ peut apparaître avant $x$. Par la méthode symbolique,

$$ST^x(z) = \frac{2z^2}{(1-\hat{A}(z)+2z)(1-\hat{A}(z)+z)(1-\hat{A}(z))},$$

et, via le Lemme 1.5.1 et l'Équation (2.2), la fraction limite des tautologies simples réalisée par $x$ est donnée par (cf. Lemme 1.5.1)

$$\begin{aligned}\lim_{n\to+\infty}\frac{ST^x_{n,k}}{A_{n,k}} = \lim_{z\to\alpha_k}\frac{1}{A'(z)}\Bigg(&\frac{4z}{(1-\hat{A}(z)+2z)(1-\hat{A}(z)+z)(1-\hat{A}(z))}\\ &+\frac{2z^2(\hat{A}'(z)-2)}{(1-\hat{A}(z)+2z)^2(1-\hat{A}(z)+z)(1-\hat{A}(z))}\\ &+\frac{2z^2(\hat{A}'(z)-1)}{(1-\hat{A}(z)+2z)(1-\hat{A}(z)+z)^2(1-\hat{A}(z))}\\ &+\frac{2z^2\hat{A}'(z)}{(1-\hat{A}(z)+2z)(1-\hat{A}(z)+z)(1-\hat{A}(z))^2}\Bigg)\end{aligned}$$

Nous savons que $\lim_{z\to\alpha_k}\hat{A}'(z) = +\infty$ car $\alpha_k$ est une singularité de type racine carrée. Donc, $\lim_{z\to\alpha_k}\frac{\hat{A}'(z)-2}{A'(z)} = \frac{1}{2}$ et $\lim_{z\to\alpha_k}\frac{1}{A'(z)} = 0$. De plus, $\alpha_k = \frac{3-2\sqrt{2}}{2k}$, et $A(\alpha_k) = \sqrt{2}-1$, donc $\hat{A}(\alpha_k) = 1-\frac{1}{\sqrt{2}}$. Nous avons donc

$$\lim_{n\to+\infty}\frac{ST^x_{n,k}}{A_{n,k}} = \frac{3\alpha_k^2}{(1-\hat{A}(\alpha_k)+o(1))^4} = \frac{51-36\sqrt{2}}{k^2}+\mathcal{O}\left(\frac{1}{k^3}\right).$$

asymptotiquement quand $k$ tend vers $+\infty$. ■

Posons $G(z) = \sum_{x\in\{x_1,\dots,x_k\}} ST^x(z) = kST^{x_1}(z)$. Et notons $\mathcal{DC}$ la famille comptée par cette série génératrice. Cette fonction génératrice compte les tautologies simples mais compte plusieurs fois chaque tautologie simple réalisée plusieurs variables distinctes : une tautologie simple peut être réalisée à la fois par $x_1$ et par $x_2$, par exemple. On notera $K_n^i$ le nombre de tautologies simples de taille $n$ réalisées simultanément par exactement $i$ variables distinctes.

Introduisons les langages de motifs suivants (généralisations aux arbres associatifs du langage de motifs évoqué en introduction, cf. Sous-Section 1.3.1) :

$$\begin{aligned}M &= \hat{M}\vee\hat{M}\,|\,\hat{M}\vee\hat{M}\vee\hat{M}\,|\dots\\ \hat{M} &= \bullet\,|\,\square\wedge\square\,|\,\square\wedge\square\wedge\square\,|\dots\end{aligned}. \tag{2.5}$$

Les $M$-feuilles de motif d'un arbre sont les feuilles de première génération dans un arbre enraciné par un $\vee$ : ce sont donc celles qui permettent de réaliser une tautologie simple.

**Lemme 2.3.7**

*Le langage de motifs $M$ est sous-critique pour la famille de squelettes associatifs $\check{\mathcal{S}}$.*

**Démonstration :** La série génératrice de $\hat{M}$ est donnée par

$$\hat{m}(x,y) = x+\frac{y^2}{1-y},$$

où $x$ marque les feuilles de motif et $y$ les emplacements, et celle de $M$ par

$$m(x,y) = \frac{\hat{m}(x,y)^2}{1-\hat{m}(x,y)}.$$

Cette fonction est donc analytique sur l'ensemble $\Omega = \{(x,y) \in \mathbb{C}^2 \,|\, y \neq 1, x + \frac{y^2}{1-y} \neq 1\}$.

La série génératrice des squelettes associatifs enracinés par $\wedge$ se déduit aisément de celle des arbres booléens associatifs (cf. Équation (2.1)), et l'on a :

$$\hat{S}(z) = \frac{1}{4}\left(1 + z - \sqrt{1 - 6z + z^2}\right),$$

et sa singularité dominante $\rho$ vérifie

$$\hat{S}(\rho) = 1 - \frac{1}{\sqrt{2}} \text{ et } \rho = 3 - 2\sqrt{2}.$$

Supposons $(x,y) \in D = \{(x,y) \in \mathbb{C} \,|\, |x| \leqslant \frac{1}{4}, |y| \leqslant \frac{1}{2}\}$. Alors, de manière évidente $y \neq 1$, et

$$\left|x + \frac{y^2}{1-y}\right| \leqslant \frac{1}{4} + \frac{1}{4|1-y|} \leqslant \frac{1}{4} + \frac{1}{4(1-|y|)} \leqslant \frac{3}{4},$$

donc $D \subset \Omega$, et comme $|\rho| < \frac{1}{4}$ et $|\hat{S}(\rho)| < \frac{1}{2}$, $M$ est bien sous-critique pour la famille $\hat{\mathcal{S}}$ des squelettes associatifs enracinés par $\wedge$. ■

**Lemme 2.3.8**

*Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty} \frac{[z^n]G(z)}{A_{n,k}} = \mu_k(\mathcal{ST}) + \mathcal{O}\left(\frac{1}{k^2}\right).$$

**Démonstration :** Réécrivons $[z^n]G(z)$ comme suit :

$$[z^n]G(z) = ST_{n,k} + \sum_{i=2}^{k}(i-1)|K_n^i|.$$

En effet, nous avons $ST_{n,k} = \sum_{i=1}^{k}|K_n^i|$. Pour tout entier $i \geqslant 3$, toute tautologie de $K_n^i$ admet au moins 3 $M$-répétitions, donc au moins 3 $(M,\varnothing)$-restrictions. Au vu du lemme de Kozik 2.3.2 :

$$\begin{aligned}\lim_{n\to+\infty} \frac{\sum_{i=2}^{k}(i-1)|K_n^i|}{A_{n,k}} &= \lim_{n\to+\infty} \frac{\sum_{i=2}^{3}(i-1)|K_n^i|}{A_{n,k}} + \lim_{n\to+\infty} \frac{\sum_{i=4}^{k}(i-1)|K_n^i|}{A_{n,k}} \\ &\leqslant 2 \lim_{n\to+\infty} \frac{A_{n,k}^{[\geqslant 3]}}{A_{n,k}} + (k-3)(k-1) \lim_{n\to+\infty} \frac{A_{n,k}^{[\geqslant 4]}}{A_{n,k}} \\ &= \mathcal{O}\left(\frac{1}{k^3}\right) + k^2\mathcal{O}\left(\frac{1}{k^4}\right) = \mathcal{O}\left(\frac{1}{k^2}\right).\end{aligned}$$

Nous avons donc bien

$$\lim_{n\to+\infty} \frac{[z^n]G(z)}{A_{n,k}} = \lim_{n\to+\infty} \frac{ST_{n,k}}{A_{n,k}} + \mathcal{O}\left(\frac{1}{k^2}\right).$$ ■

**Démonstration de la Proposition 2.3.5 :** Résumons les résultats obtenus dans les lemmes précédents : asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k(\mathcal{ST}) = \frac{51 - 36\sqrt{2}}{k} + \mathcal{O}\left(\frac{1}{k^2}\right),$$

ce qui conclut la preuve. ■

Nous avons calculé la fraction limite des tautologies simples : nous devons maintenant montrer la proposition suivante, afin de déterminer la probabilité de la fonction constante `Vrai`.

**Proposition 2.3.9**

*Dans le modèle associatif, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple.*

La preuve de ce résultat est très similaire à celle évoquée en introduction (cf. Section 1.3.1) et développée dans [Koz08]. Nous allons utiliser la théorie des motifs, et utiliser plus précisément les langages de motifs suivants, généralisations du langage de motifs $N$ défini en introduction.

$$\begin{cases} \hat{N} = \bullet | \check{N} \wedge \square | \check{N} \wedge \square \wedge \square | \dots \\ \check{N} = \bullet | \hat{N} \vee \hat{N} | \hat{N} \vee \hat{N} \vee \hat{N} | \dots \\ R = \hat{N} \cup \check{N}, \end{cases} \tag{2.6}$$

où $R = \hat{N} \cup \check{N}$ contient tous les arbres de $\check{N}$ et de $\hat{N}$. Il est facile de voir que $\hat{N}$, $\check{N}$ et $R$ sont non ambigus. De plus,

**Lemme 2.3.10**

*Le langage de motifs $R$ est sous-critique pour la famille $\mathcal{S}$ des squelettes associatifs, au sens où $\hat{N}$ est sous-critique par rapport à $\check{\mathcal{S}}$ et $\check{N}$ est sous-critique par rapport à $\hat{\mathcal{S}}$.*

**Démonstration :** La fonction génératrice de $R$ est donnée par

$$p(x, y) = \hat{p}(x, y) + \check{p}(x, y) - x,$$

où $\hat{p}(x, y)$ (resp. $\check{p}(x, y)$) est la fonction génératrice du langage de motifs $\hat{N}$ (resp. $\check{N}$). Nous allons montrer, successivement, que les langages $\check{N}$ et $\hat{N}$ sont sous-critiques pour la famille $\hat{\mathcal{S}}$ des squelettes enracinés par $\wedge$, ce qui impliquera le Lemme 2.3.10. Via la méthode symbolique, ces deux fonctions génératrices vérifient le système suivant :

$$\begin{cases} \hat{p}(x, y) = x + \dfrac{y}{1-y}\check{p}(x, y) \\ \check{p}(x, y) = x + \dfrac{\hat{p}(x, y)^2}{1 - \hat{p}(x, y)}, \end{cases}$$

ce qui implique

$$\begin{cases} \hat{p}(x, y) = \dfrac{1}{2}\left(1 - y + x - \sqrt{(y - x - 1)^2 - 4x}\right) \\ \check{p}(x, y) = x + \dfrac{\hat{p}(x, y)^2}{1 - \hat{p}(x, y)}. \end{cases}$$

La fonction $\hat{p}(x, y)$ est analytique sur l'ensemble $\hat{\Omega} = \{(x, y) \in \mathbb{C}^2 \mid (1 + x - y)^2 - 4x \neq 0\}$, et la fonction $\check{p}(x, y)$ est analytique sur $\check{\Omega} = \{(x, y) \in \mathbb{C}^2 \mid (1 + x - y)^2 - 4x \neq 0 \text{ et } \hat{p}(x, y) \neq 1\}$. Notons que $(x, y) \in \mathbb{C}$ est un élément de $\hat{\Omega}$ si, et seulement si,

$$x^2 - 2(1 + y)x + (1 - y)^2 = 0,$$

c'est à dire si, et seulement si,

$$x = (1 - \sqrt{y})^2 \text{ ou } x = (1 + \sqrt{y})^2.$$

Rappelons que la famille $\check{\mathcal{S}}$ des squelettes associatifs dont la racine est étiquetée par $\vee$ a pour singularité $\rho = 3 - 2\sqrt{2}$ et que sa série génératrice vérifie $\check{S}(\rho) = 1 - \frac{1}{\sqrt{2}}$. Comme $\rho < (1 - \sqrt{\check{S}(\rho)})^2$, le langage de motif $\hat{N}$ est sous-critique pour $\check{\mathcal{S}}$. Pour montrer que le langage de motif $\check{N}$ est sous-critique

pour $\hat{\mathcal{S}}$, il nous reste à prouver que $\hat{p}(x,y) \neq 1$ pour tout $(x,y) \in D$. Comme $\hat{p}(x,y)$ est une série entière bi-variée à coefficients positifs,

$$|\hat{p}(x,y)| \leqslant \hat{p}(|x|,|y|) \leqslant \hat{p}\left(\frac{1}{8},\frac{1}{4}\right),$$

pour tout $(x,y) \in D$. Comme

$$\hat{p}\left(\frac{1}{8},\frac{1}{4}\right) < 1,$$

nous en déduisons que $D \subseteq \check{\Omega}$, et donc que le langage $R$ est sous-critique pour $\mathcal{S}$. ■

**Remarque :** Tout comme le langage de motifs $N$ défini en introduction (cf. Section 1.3.1), le langage de motifs $R$ vérifie la propriété suivante : si toute les $R$-feuilles de motif d'un arbre booléen sont assignées à la valeur `Faux`, alors l'arbre booléen entier calcule `Faux`, quelles que soient les valeurs des autres variables booléennes. Cette propriété se vérifie par induction à partir de la définition du langage de motifs $R$ (cf. Équation (2.6)).

La Figure 2.3 donne un exemple d'arbre de $\mathcal{A}$, exhibe les motifs de $R$ et $R[R] := \hat{N}[\check{N}] \cup \check{N}[\hat{N}]$ à partir desquels il est construit.

**Démonstration de la Proposition 2.3.9 :** Soit $t$ une tautologie associative, i.e. un arbre booléen associatif calculant la fonction constante `Vrai`. On pose $\Gamma = \varnothing$. Supposons que $t$ a exactement une $(R[R], \varnothing)$-restriction. Comme $\Gamma = \varnothing$, cette restriction est une répétition.

Si cette répétition est due à deux apparitions d'un même littéral $\alpha$ dans deux feuilles de motifs, alors, nous pouvons affecter tous les littéraux étiquetant les $R$-feuilles de motifs à `Faux` (en particulier, on affecte $\alpha$ à `Faux`). Pour cette affectation des variables, l'arbre $t$ calcule `Faux`, par définition du langage $R$. C'est impossible puisque $t$ est une tautologie.

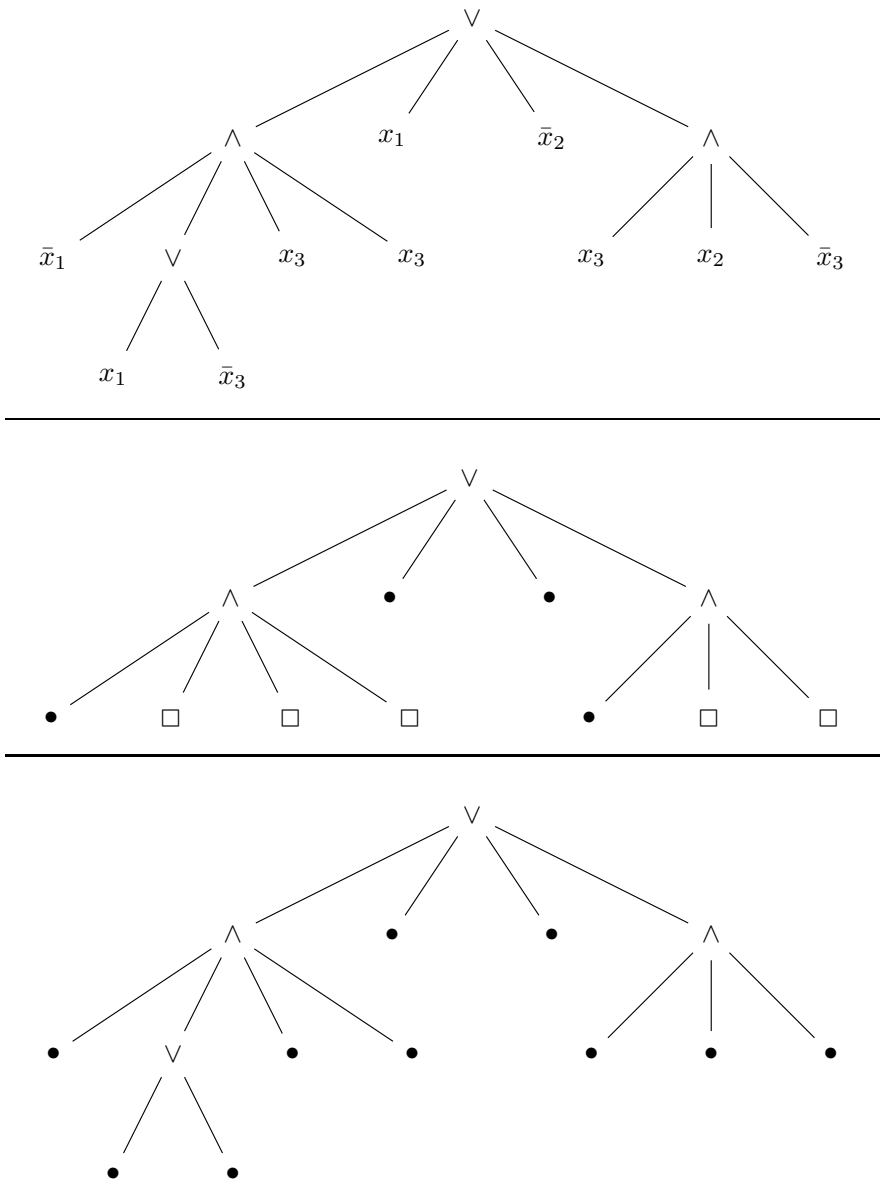
Par conséquent, la répétition doit être due à l'apparition d'un littéral $\alpha$ et de sa négation $\bar{\alpha}$ parmi les $R[R]$-feuilles de motif : supposons que seul $\alpha$ (resp. seul $\bar{\alpha}$) apparaisse parmi les $R$-feuilles de motifs. Dans ce cas, toutes les $R$-feuilles de motif peuvent être assignées à `Faux`, ce qui est impossible. Dès lors, $\alpha$ et $\bar{\alpha}$ apparaissent parmi les $R$-feuilles de motif. Supposons qu'il existe un nœud interne $\nu$, étiqueté par $\wedge$ entre $\alpha$ (resp. $\bar{\alpha}$) et la racine. On notera $t_1, \dots t_r$ les sous-arbres de $\nu$. Supposons par exemple que la $R$-feuille de motif étiquetée par $\alpha$ soit dans $t_1$ (cf. Figure 2.4). Alors, il n'y a pas de répétition parmi les $R[R]$-feuilles de motif de $t_2, \dots, t_r$, et on peut donc toutes les affecter à `Faux`, et ce sans fixer la valeur de $\alpha$. Ainsi, le sous-arbre enraciné en $\nu$ calcule `Faux`, et ce sans avoir fixé aucune des variables étiquetant les $R[R]$-feuilles de motifs de $t \backslash \{t_1 \cup \dots \cup t_r\}$ : nous pouvons donc affecter toutes ces feuilles à `Faux`, et il est facile de voir que $t$ calcule `Faux` pour cette affectation. C'est impossible, et $t$ est donc une tautologie simple : toute tautologie ayant exactement une $R[R]$-restriction est une tautologie simple.

De plus, une tautologie admet au moins une $R$-répétition, donc au moins une $R[R]$-répétition (car sinon, on peut affecter toutes les $R$-feuilles de motifs de cette tautologie à `Faux`, forçant ainsi l'arbre total à calculer `Faux`). Enfin, l'ensemble des tautologies de taille $n$ ayant au moins 2 $R[R]$-restrictions est inclus dans $A_{n,k}^{[\geqslant 2]}$ et d'après le Lemme 2.3.2, cet ensemble admet une fraction limite d'ordre $\mathcal{O}\left(\frac{1}{k^2}\right)$ quand $k$ tend vers $+\infty$. Ces tautologies sont donc négligeables devant l'ensemble des tautologies simples. ■

### 2.3.2 Arbres commutatifs

L'objectif de cette partie est de calculer la probabilité de `Vrai` dans le modèle des arbres commutatifs :

FIGURE 2.3 – *Un arbre associatif suivi du motif de $R$ à partir duquel il est construit, puis du motif de $R[R] := \hat{N}[\check{N}] \cup \check{N}[\hat{N}]$ à partir duquel il est construit.*



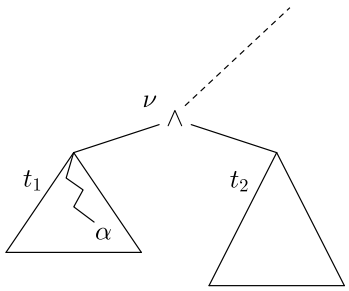

**Théorème 2.3.11**

*La probabilité de la fonction* `Vrai` *dans le modèle des arbres commutatifs vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^c(\texttt{Vrai}) = \frac{641}{1024k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

Tout comme dans les cas classique et associatif, la preuve se fait en deux étapes : nous généralisons tout d'abord le lemme de Kozik 1.3.5 pour pouvoir l'appliquer dans le contexte des arbres

FIGURE 2.4 – *Le nœud $\nu$ étiqueté par $\wedge$ et séparant la feuille $\alpha$ de la racine dans la preuve de la Proposition 2.3.9 (dans le cas particulier $r = 2$).*




commutatifs, puis, nous montrons que presque toute tautologie est simple, asymptotiquement quand $k$ tend vers $+\infty$. La théorie des motifs exige que les motifs considérés soient plans, alors que les arbres commutatifs ne le sont pas. C'est pourquoi nous allons plonger partiellement les arbres commutatifs dans le plan : un *mobile* sera un arbre dont certaines branches proches de la racine seront plongées dans le plan, et d'autres branches ne le seront pas. Ce plongement partiel est la clef de la généralisation du lemme de Kozik 1.3.5 (la Figure 2.5 donne l'exemple de deux mobiles).

De plus, dans le cadre commutatif, il n'est plus vrai que le nombre de squelettes de taille $n$ est relié au nombre d'arbres commutatifs de taille $n$ via un facteur $(2k)^n$. Nous devrons donc greffer dans les motifs des arbres déjà étiquetés, et la notion de sous-criticalité devra donc concerner les langages de motifs aux feuilles de motif étiquetées par rapport aux arbres commutatifs.

### Généralisation du lemme de Kozik aux *mobiles*

Étant donné un langage de motifs $L$ (les arbres d'un langage de motifs sont toujours plans), on appellera **mobile** un arbre de $L[\mathcal{C}]$. Notons que toute la partie *motif* d'un mobile est plane alors que les arbres greffés dans les emplacements sont non plans : c'est pourquoi nous parlerons de plongement partiel. De plus, les feuilles de motif des arbres de $L[C]$ sont non étiquetées alors que les feuilles qui ne sont pas des feuilles de motif sont étiquetées.

**Lemme 2.3.12**

*Soit $\Gamma$ un sous-ensemble de $\{x_1, \dots, x_k\}$ dont le cardinal ne dépend pas de $k$. Soit $L$ un langage de motifs (plans) de fonction génératrice $\ell(x, y)$. À partir de la fonction génératrice des motifs dont les feuilles de motif sont étiquetées, on définit les fonctions $(A_j(y))_{j\geqslant 0}$ comme suit :*

$$A_j(y) = [x^j]\ell(2kx, y), \forall j \geqslant 0.$$

*Supposons que les coefficients $A_j(y)$ soient sous-critiques pour la famille des arbres booléens commutatifs $\mathcal{C}$. On note $L[\mathcal{C}]_{n,k}^{[p]}$ (resp. $L[\mathcal{C}]_{n,k}^{[\geqslant p]}$) l'ensemble des éléments de $L[\mathcal{C}]$ de taille $n$, dont les feuilles de motif ont été étiquetées sur $k$ variables, et ayant exactement (resp. au moins) $p$ $(L, \Gamma)$-restrictions. Enfin, on note $L[\mathcal{C}]_{n,k}$ le nombre de mobiles de taille $n$ dont on a étiqueté les feuilles de motif sur $k$ variables.*

*Alors, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty} \frac{L[\mathcal{C}]_{n,k}^{[\geqslant p]}}{C_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right) \quad \text{and} \quad \lim_{n\to\infty} \frac{L[\mathcal{C}]_{n,k}^{[p]}}{C_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right).$$

Nous allons montrer ce lemme en suivant la stratégie utilisée dans le cas planaire (cf. Section 2.3.1).

Soit $t$ un mobile de $L[\mathcal{C}]$ de taille $n$ et ayant $\ell$ feuilles de motif : remarquons que les feuilles des arbres greffés dans les emplacements sont étiquetées alors que les feuilles de motif ne le sont pas. On notera $\gamma = |\Gamma|$. Pour tout $r \leqslant p$, le nombre d'étiquetages différents des feuilles de motif de $t$ tels que $t$ admet $r$ $L$-répétitions et $p$ $(L,\Gamma)$-restrictions est donné par

$$\left\{ \begin{matrix} \ell \\ \ell - r \end{matrix} \right\} \binom{\gamma}{p-r} (\ell - r)^{\underline{p-r}} (k-\gamma)^{\underline{\ell - r - (p-r)}} 2^{\ell},$$

où, tout comme dans le cas plan, $x^{\underline{y}} = x(x-1)\dots(x-y+1)$ et $\left\{ \begin{matrix} y \\ x \end{matrix} \right\}$ sont les nombres de Stirling de seconde espèce.

Dès lors, le lemme suivant (modification du Lemme 2.3.3) est immédiat :

**Lemme 2.3.13**

*Soit $t$ un mobile de $L[\mathcal{C}]$ ayant $\ell$ $L$-feuilles de motif. Le nombre d'étiquetages différents des feuilles de motifs de $t$ tels que $t$ a exactement $p$ $(L,\Gamma)$-restrictions est donné par*

$$(k-\gamma)^{\underline{\ell - p}} 2^{\ell} w_{\gamma,p}(\ell),$$

*où $w_{\gamma,p}(\ell) = \sum_{r=0}^{p} \left\{ \begin{matrix} \ell \\ \ell - r \end{matrix} \right\} \binom{\gamma}{p-r} (\ell - r)^{\underline{p-r}}$ est un polynôme en $\ell$.*

La Proposition 2.3.4 doit elle aussi être adaptée au modèle commutatif : les langages $N = \bullet | N \vee N | N \wedge \square$ et le langage $S = \bullet | S \vee S | \square \wedge \square$ dont les feuilles de motifs sont étiquetées ne sont pas sous-critiques pour la famille des arbres commutatifs $\mathcal{C}$. Nous montrons dans la proposition suivante qu'une notion plus faible de sous-criticalité suffit :

**Proposition 2.3.14**

*Soit $L$ un langage de motifs non ambigu de fonction génératrice $\Lambda(x,y)$. On définit les fonctions $(A_j(y))_{j \geqslant 0}$ à partir de la fonction génératrice des motifs de $L$ dont les feuilles de motifs sont étiquetées :*

$$A_j(y) = [x^j]\Lambda(2kx, y), \forall j \geqslant 0.$$

*La série génératrice $A_j(z)$ compte les motifs ayant $j$ feuilles de motifs. Supposons que, pour tout $j \geqslant 0$, $A_j(y)$ soit sous-critique pour $\mathcal{C}$. On notera $L[\mathcal{C}]_n(\ell)$ le nombre d'arbres de $L[\mathcal{C}]$ de taille $n$ ayant exactement $\ell$ feuilles de motif. Soit $w$ un polynôme de degré $\lambda$.*

*Alors, il existe une constante $c_w \geqslant 0$ et un entier $N > 0$ tels que*

$$\lim_{n \to +\infty} \frac{\sum_{\ell=0}^{N} (2k)^{\ell} L[\mathcal{C}]_n(\ell) w(\ell)}{C_{n,k}} = c_w.$$

**Démonstration :** Soit $N$ un entier. La fonction génératrice de la suite $((2k)^{\ell} L[\mathcal{C}]_n(\ell) w(\ell) \mathbb{1}_{\ell \leqslant N})_{\ell \geqslant 0}$ sera notée $\Lambda_w(x,y)$, où $x$ marque les feuilles de motif et $y$ marque les emplacements. Écrivons le polynôme $w(\ell)$ sous la forme

$$w(\ell) = \sum_{j=0}^{\lambda} w_j \ell^{\underline{j}}.$$

Enfin, soit $\Lambda_N(x,y) = \sum_{\ell=0}^{N} A_\ell(y) x^\ell$ la troncature de la série génératrice $\Lambda(2kx, y)$.

Notons que

$$x^j \frac{\partial^j \Lambda_N(x,y)}{\partial x^j} = \sum_{\ell=0}^{N} \underline{\ell^j} A_\ell(y) x^\ell,$$

ce qui implique

$$\sum_{j=0}^{\lambda} w_j x^j \frac{\partial^j \Lambda_N(x,y)}{\partial x^j} = \sum_{\ell=0}^{N} w(\ell) A_\ell(y) x^\ell.$$

Dès lors, la fonction génératrice $\Lambda_w(x,y)$ est une combinaison linéaire des dérivées de $\Lambda_N(x,y)$ par rapport à $x$, lesquelles sont des sommes finies de termes sous-critiques par rapport à $\mathcal{C}$. Ainsi, $\Lambda_w(z, C(z))$ et $C(z)$ ont même rayon de convergence et sont toutes deux $\Delta$-analytiques (cf. [FS09, page 389]). De plus, au vu de [Koz08, Observation 2.3] chaque terme sous-critique a une expansion en racine carrée : ainsi, la singularité de $\Lambda_w(z, C(z))$ est de type racine carrée ou bien d'ordre plus grand s'il y a des simplifications.

Nous pouvons donc appliquer le lemme de transfert de Flajolet et Odlyzsko [FO90] à $C(z)$ et $\Lambda_w(z, C(z))$, et, asymptotiquement quand $n$ tend vers $+\infty$, il existe une constante $cst \geqslant 0$ telle que

$$\frac{[z^n]\Lambda_w(z, C(z))}{C_{n,k}} \sim cst.$$

Dès lors, il existe une constante $c_w \geqslant 0$ telle que,

$$\lim_{n\to+\infty} \frac{\sum_{\ell\geqslant 0}(2k)^\ell L[\mathcal{C}]_n(\ell) w(\ell)}{C_{n,k}} = c_w.$$

La constante $c_w$ est non nulle s'il n'y a pas eu de simplification entre les différents termes sous-critiques. ■

Nous pouvons donc désormais démontrer le Lemme 2.3.12 :

**Démonstration du Lemme 2.3.12 :** D'après le Lemme 2.3.13,

$$\frac{L[\mathcal{C}]^{[p]}_{n,k}}{C_{n,k}} = \frac{\sum_{\ell=0}^{N} L[\mathcal{C}]_n(\ell) w_{p,\gamma}(\ell)(k-\gamma)^{\underline{\ell-p}} 2^\ell}{C_{n,k}},$$

où $N = k - \gamma + p$, car pour $\ell > N$, le facteur $(k-\gamma)^{\underline{\ell-p}}$ est nul. Nous avons donc

$$\frac{L[\mathcal{C}]^{[p]}_{n,k}}{C_{n,k}} \leqslant \frac{\sum_{\ell=0}^{N} L[\mathcal{C}]_n(\ell) w_{p,\gamma}(\ell) k^{\ell-p}\, 2^\ell}{C_{n,k}} = \frac{\sum_{\ell=0}^{N} (2k)^\ell L[\mathcal{C}]_n(\ell) w_{p,\gamma}(\ell)}{C_{n,k}} \cdot \frac{1}{k^p}.$$

Finalement, au vu de la Proposition 2.3.14, nous obtenons

$$\lim_{n\to+\infty} \frac{L[\mathcal{C}]^{[p]}_{n,k}}{C_{n,k}} \leqslant \frac{c_w}{k^p}.$$

Un raisonnement très similaire permet l'étude de $\lim_{n\to+\infty} \frac{L[\mathcal{C}]^{[\geqslant p]}_{n,k}}{C_{n,k}}$. ■

### Sous-criticalité

Tout comme dans le cas binaire planaire, nous allons avoir besoin d'utiliser deux langages de motifs : le langage $N = \bullet | N \vee N | N \wedge \square$ et le langage $S = \bullet | S \vee S | \square \wedge \square$. En préliminaire, montrons que ces deux langages de motifs vérifient l'hypothèse de sous-criticalité du Lemme 2.3.12.

**Lemme 2.3.15**

*Soit $n(x,y)$ la fonction génératrice du langage de motifs $N$. Définissons les $(A_j(y))_{j\geqslant 0}$ comme suit :*

$$A_j(y) = [x^j]n(2kx, y),$$

*alors, les fonctions $(A_j(y))_{j\geqslant 0}$ sont sous-critiques pour la famille $\mathcal{C}$.*

**Démonstration :** Par méthode symbolique, nous obtenons

$$n(x,y) = x + n(x,y)^2 + yn(x,y),$$

d'où nous déduisons

$$n(x,y) = \frac{1}{2}\left(1 - y - \sqrt{(y-1)^2 - 4x}\right).$$

Calculons les $(A_j(y))_{j\geqslant 0}$ : comme $n(0,0) = 0$,

$$\begin{aligned} n(2kx,y) &= \frac{1-y}{2} - \frac{1}{2}\sqrt{(y-1)^2}\sqrt{1 - \frac{8kx}{(y-1)^2}} \\ &= \frac{1-y}{2} - \frac{1}{2}(1-y)\sum_{j\geqslant 0}\binom{1/2}{j}(-8k)^j(y-1)^{-2j}x^j, \end{aligned}$$

où $\binom{1/2}{j} = 1/2(1/2-1)\dots(1/2-(j-1))1/j!$. Dès lors, pour tout $j \geqslant 1$, $A_j(y) = -\frac{1}{2}(1-y)\binom{1/2}{j}(-8k)^j(y-1)^{-2j}$ est une fonction rationnelle de $y$ et son rayon de convergence est 1, et $A_0(y)$ est une fonction entière de $y$. Les $(A_j(y))_{j\geqslant 0}$ sont donc sous-critiques pour la famille $\mathcal{C}$. ■

**Lemme 2.3.16**

*Soit $s(x,y)$ la fonction génératrice du langage de motifs $S$. Définissons les $(A_j(y))_{j\geqslant 0}$ comme suit :*

$$A_j(y) = [x^j]s(2kx, y),$$

*alors, les fonctions $(A_j(y))_{j\geqslant 0}$ sont sous-critiques pour la famille $\mathcal{C}$.*

**Démonstration :** La fonction génératrice du langage de motifs $S$ vérifie $s(x,y) = x + s(x,y)^2 + y^2$ ce qui implique, comme $s(0,0) = 0$,

$$s(x,y) = \frac{1 - \sqrt{1-x-y^2}}{2}.$$

Dès lors,

$$\begin{aligned} s(2kx,y) &= \frac{1}{2} - \frac{\sqrt{1-y^2}}{2}\sqrt{1 - \frac{2kx}{1-y^2}} \\ &= \frac{1}{2} - \frac{\sqrt{1-y^2}}{2}\sum_{j\geqslant 0}\binom{1/2}{j}\frac{(2k)^j}{(1-y^2)^j}x^j. \end{aligned}$$

Dès lors, $A_j(y) = -\frac{1}{2}\binom{1/2}{j}(2k)^j(1-y^2)^{1/2-j}$ est une fonction rationnelle de $y$ et son rayon de convergence est 1. Les $(A_j(y))_{j\geqslant 0}$ sont donc sous-critiques pour la famille $\mathcal{C}$. ■

Enfin, le lemme suivant nous assure que, si un langage de motifs $L$ vérifie l'hypothèse de sous-criticalité du Lemme 2.3.12, alors $L[L]$ aussi, ainsi que toute puissance de $L$ au sens suivant :

**Définition 2.3.17**

*Étant donné un langage de motifs $L$, pour tout entier $r > 1$, on définit $L^r$ comme suit :*

$$L^1 = L \;\; et \;\; L^{r+1} = L^r[L].$$

*Sa série génératrice est donnée par*

$$\ell^{*r}(x,y) = \underbrace{\ell(x,(\ell(x,\cdots\ell(x}_{r \text{ fois}},y)\cdots))),$$

*et, pour tout $j \geqslant 0$, on définit*

$$A_j^{*r}(y) = [x^j]\ell^{*r}(2kx,y).$$

**Lemme 2.3.18**

*Soit $L$ un langage de motifs de fonction génératrice $\ell(x,y)$. On définit les $(A_j(y))_{j\geqslant 0}$ comme suit :*

$$A_j(y) = [x^j]\ell(2kx,y), \forall j \geqslant 0.$$

*Supposons $A_0(y) = 0$. Si, pour tout $j \geqslant 1$, $A_j(y)$ est sous-critique pour la famille des arbres commutatifs $\mathcal{C}$, alors, pour tout $r \geqslant 1$, pour tout $j \geqslant 1$, $A_j^{*r}(y)$ est sous-critique pour la famille $\mathcal{C}$.*

**Démonstration :** Il est important de noter que l'hypothèse $A_0(y) = 0$ nous assure que tout motif de $L$ a au moins une feuille de motif. Il est facile de voir que cette propriété est aussi vérifiée par $A_0^{*r}(y)$, pour tout $r \geqslant 1$.

Nous allons raisonner par récurrence sur $r$. Le cas $r = 1$ correspond à $A_j^{*1}(y) = A_j(y) = \sum_{\ell\geqslant 0} a_{\ell,j}x^\ell y^j$ qui est sous-critique pour $\mathcal{C}$ par hypothèse. Soit $r \geqslant 1$, supposons que pour tout $j \geqslant 1$, $A_j^{*r}(y)$ est sous-critique pour $\mathcal{C}$. Nous voulons montrer que $[x^j]\ell^{*(r+1)}(2kx,y)$ est sous-critique pour $\mathcal{C}$ :

$$\ell^{*(r+1)}(2kx,y) = \ell(2kx,\ell^{*r}(x,y)) = \sum_{j\geqslant 0} A_j(\ell^{*r}(2kx,y))x^j.$$

Dès lors, pour tout $\lambda \geqslant 0$,

$$\begin{aligned}
[x^\lambda]\ell^{*(r+1)}(x,y) &= [x^\lambda]\sum_{j\geqslant 0} A_j(\ell^{*r}(2kx,y))x^j \\
&= [x^\lambda]\sum_{j\geqslant 0}\sum_{\ell\geqslant 0} a_{\ell,j}\left(\ell^{*r}(2kx,y)\right)^\ell x^j \\
&= [x^\lambda]\sum_{j\geqslant 0}\sum_{\ell\geqslant 0} a_{\ell,j}\left(\sum_{\mu\geqslant 0} x^\mu A_\mu^{*r}(y)\right)^\ell x^j \\
&= [x^\lambda]\sum_{j\geqslant 0}\sum_{\ell\geqslant 0} a_{\ell,j}\sum_{\mu_1,\ldots,\mu_\ell} x^{\sum\mu_i+j}A_{\mu_1}^{*r}\cdots A_{\mu_\ell}^{*r} \\
&= \sum_{j\geqslant 0}\sum_{\ell\geqslant 0} a_{\ell,j}\sum_{\mu_1+\ldots+\mu_\ell=\lambda-j} A_{\mu_1}^{*r}\cdots A_{\mu_\ell}^{*r}.
\end{aligned}$$

Comme $A_0^{*r}(y) = 0$, pour tout $i \in \{1,\ldots,j\}$, $\mu_i > 0$. Dès lors, dans chaque terme de la somme ci-dessus, nous avons $\lambda$ facteurs, et,

$$[x^\lambda]\ell^{*(r+1)}(x,y) = \sum_{j=0}^{\lambda}\sum_{\ell=0}^{\lambda} a_{\ell,j}\sum_{\substack{\mu_1,\ldots,\mu_\ell,\\ \mu_1+\ldots+\mu_\ell=\lambda-j}} A_{\mu_1}^{*r}(y)\cdots A_{\mu_\ell}^{*r}(y)$$

est une somme finie de produits de termes sous-critiques pour $\mathcal{C}$, ce qui conclut la preuve. ■

### Tautologies commutatives

**Proposition 2.3.19**

*Dans le modèle commutatif, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple (cf. Définition 1.3.6).*

Tout comme dans le modèle associatif, nous allons utiliser le langage de motifs $N = \bullet | N \vee N | N \wedge \square$, et à tout arbre de $\mathcal{C}$, nous allons associer un arbre de $N[\mathcal{C}]$ de la façon suivante : partons de la racine, et choisissons un ordre de gauche à droite pour ses descendants ; si la racine est étiquetée par un $\wedge$, procédons récursivement pour la racine du sous-arbre gauche, le sous-arbre droit reste non-plan, il est greffé dans un emplacement ; si la racine est étiquetée par un $\vee$, procédons récursivement pour tous ses sous-arbres : si la racine est en fait une feuille, alors, c'est une $N$-feuille de motif. Nous obtenons bien un élément de $N[\mathcal{C}]$, que nous appellerons un **semi-plongement** de l'arbre de départ. Bien entendu, un même arbre de $\mathcal{C}$ admet plusieurs semi-plongements. Par contre, un mobile de $N[\mathcal{C}]$ est le semi-plongement d'un unique arbre de $\mathcal{C}$.

Le procédé de semi-plongement peut-être appliqué pour n'importe quel langage de motifs $L$ non ambigu. Soit $\Gamma \subseteq \{x_1, \ldots, x_k\}$ un ensemble de variables dont le cardinal ne dépend pas de $k$. Pour chaque arbre $t$ de $\mathcal{C}$, on appellera **semi-plongement minimal** de $t$ un semi-plongement qui minimise le nombre de $(L, \Gamma)$-restrictions.

**Lemme 2.3.20**

*Soit $t$ un arbre commutatif calculant* `Vrai`*. Tout semi-plongement de $t$ a au moins une $N[N]$-répétition.*

**Démonstration :** Soit $\hat{t}$ un semi-plongement de $t$. Raisonnons par l'absurde et supposons que $\hat{t}$ n'ait pas de $N[N]$-répétition. Alors, nous pouvons affecter simultanément toutes ses $N$-feuilles de motif à `Faux`, et $\hat{t}$ calcule `Faux` (cette propriété est due à la définition du langage de motifs $N$). Seulement, le semi-plongement d'un arbre ne change pas la fonction qu'il calcule : la fonction calculée par $\hat{t}$ est la même que celle calculée par $t$, à savoir `Vrai`. C'est donc impossible. ■

**Lemme 2.3.21**

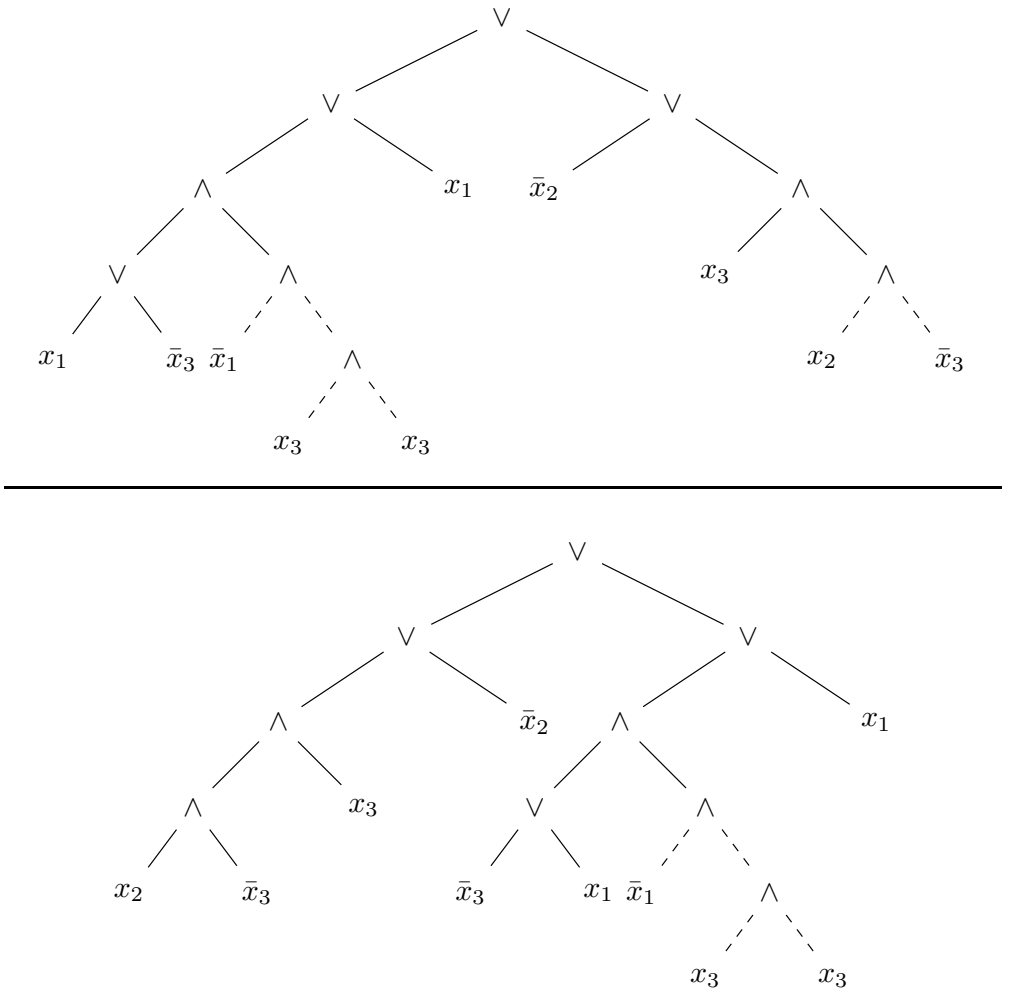
*Soit $t$ un arbre commutatif dont les semi-plongements minimaux ont exactement une $N[N]$-répétition. Alors, $t$ est une tautologie simple.*

**Démonstration :** Soit $\hat{t}$ un semi-plongement minimal de $t$. Il existe une variable $x$ qui apparaît comme étiquette de deux $N[N]$-feuilles de motifs de $\hat{t}$. Si les deux occurrences de cette variable sont deux occurrences du même littéral, alors, nous pouvons affecter les $N$-feuilles de motifs de $\hat{t}$ à `Faux`, et, pour cette affectation, $\hat{t}$, et donc $t$, calculent `Faux`, ce qui est impossible. Avec un raisonnement similaire, nous pouvons affirmer que $\hat{t}$ a exactement une $N$-répétition.

Supposons qu'il y ait un nœud $\nu$ étiqueté par $\wedge$ entre la $N$-feuille de motif étiquetée par $x$ et la racine de $\hat{t}$. Dès lors, On notera $t_\ell$ le sous-arbre gauche de $\nu$ et $t_r$ son sous-arbre droit. Supposons par exemple que la $N$-feuille de motif étiquetée par $x$ soit dans $t_\ell$. Alors, il n'y a pas de répétition parmi les $N[N]$-feuilles de motif de $t_r$, et on peut donc toutes les affecter à `Faux`, et ce sans fixer la valeur de $x$. Ainsi, le sous-arbre enraciné en $\nu$ calcule `Faux`, et ce sans avoir fixé aucune des variables étiquetant les $N[N]$-feuilles de motifs de $t \backslash t_r$ : nous pouvons donc affecter toutes ces feuilles à Faux, et il est facile de voir que $\hat{t}$, et donc $t$, calcule `Faux` pour cette affectation, ce qui est absurde.

Le même raisonnement nous assure que la $N$-feuille de motif de $\hat{t}$ étiquetée par $\bar{x}$ est reliée à la racine par un chemin de connecteurs $\vee$ : $t$ est donc bien une tautologie simple.

FIGURE 2.5 – *Les deux arbres ci-dessous sont deux semi-plongements selon le langage de motifs $N$ d'un même arbre commutatifs. Les arêtes dessinées en trait plein sont plongées dans le plan, celles en pointillés ne le sont pas. Le premier semi-plongement admet* 5 *$(N, \{x_1, x_2\})$-restrictions alors que le second en admet* 6.




■

**Démonstration de la Proposition 2.3.19 :** Soit $t$ un arbre commutatif calculant `Vrai`. Alors, il existe au moins une variable $x$ qui apparaît au moins deux fois comme étiquette d'une feuille de $t$ (car sinon, $t$ ne peut être une tautologie). Notons $C_{n,k}^{[p]}$ le nombre d'arbres de $\mathcal{C}$ de taille $n$ ayant $p$ $N$-restrictions. Comme un mobile de $N[\mathcal{C}]$ est le semi-plongement d'un unique arbre de $\mathcal{C}$, nous savons que

$$C_{n,k}^{[p]} \leqslant N[\mathcal{C}]_{n,k}^{[p]}.$$

Les Lemmes 2.3.18 et 2.3.15 nous autorisent à appliquer le Lemme 2.3.26, ce qui implique,

$$\lim_{n\to+\infty} \frac{C_{n,k}^{[p]}}{C_{n,k}} \leqslant \lim_{n\to+\infty} \frac{N[\mathcal{C}]_{n,k}^{[p]}}{C_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right).$$

Dès lors, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie a exactement une $N[N]$-répétition et est donc une tautologie simple. ■

Il nous reste donc à énumérer les tautologies simples. Tout comme dans le cas associatif, nous procèderons par étapes : comptons d'abord les tautologies réalisées par une variable booléenne fixée $x$, puis déduisons-en la fraction limite des tautologies simples en étant attentifs aux double-comptages que nous pourrions faire.

Fixons $x \in \{x_1, \dots, x_k\}$. On notera $ST^x_{n,k}$ le nombre de tautologies simples de taille $n$ réalisées par $x$.

**Lemme 2.3.22**

*Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty} \frac{ST^x_{n,k}}{C_{n,k}} = \frac{384}{512\,k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

**Démonstration :** Soit $g_x(z)$ la fonction génératrice des arbres commutatifs qui ont une feuille étiquetée par $x$ et reliée à la racine par un chemin de $\vee$. Dès lors, $g_x(z) = C(z) - \bar{g}_x(z)$ où

$$\bar{g}_x(z) = (2k-1)z + \frac{1}{2}\left(C^2(z) + C(z^2)\right) + \frac{1}{2}\left(\bar{g}^2_x(z) + \bar{g}_x(z^2)\right), \tag{2.7}$$

car un arbre compté par $g_x(z)$ doit être enraciné par un $\vee$, et un de ses deux sous-arbres doit avoir une feuille étiquetée par $x$ et reliée à la racine du sous-arbre par un chemin de $\vee$. La fonction génératrice $ST^x(z)$ qui compte les tautologies simples réalisés par la variable $x$ est donnée par $ST^x(z) = C(z) - \overline{ST}^x(z)$, où $\overline{ST}^x(z)$ compte les arbres qui ne sont pas des tautologies simples réalisées par $x$. Un tel arbre est soit enraciné par un $\wedge$, soit enraciné par un $\vee$. S'il est enraciné par un $\vee$, alors aucun de ses deux sous-arbres n'est une tautologie simple réalisée par $x$, et si l'un a une feuille étiquetée par $x$ reliée à la racine par un chemin de $\vee$, alors, le second ne contient pas de feuille $\bar{x}$ reliée à la racine par un chemin de $\vee$. Dès lors, la fonction génératrice $\overline{ST}^x(z)$ vérifie l'équation implicite suivante.

$$\overline{ST}^x(z) = 2kz + \frac{1}{2}\left(C^2(z) + C(z^2)\right) + \frac{1}{2}\left((\overline{ST}^x)^2(z) + \overline{ST}^x(z^2)\right) - (g_x(z) - ST^x(z))(g_{\bar{x}}(z) - ST^x(z)). \tag{2.8}$$

Dans un premier temps, calculons la fraction limite des tautologies simples réalisées par la variable $x$ : cette fraction limite est donnée par $1 - \lim_{z\to\gamma_k} \frac{(\overline{ST}^x)'(z)}{C'(z)}$. On notera $u_k := \bar{g}_x(\gamma_k)$, $v_k := \bar{g}_x(\gamma_k^2), U_k := \overline{ST}^x(\gamma_k)$ et $V_k := \overline{ST}^x(\gamma_k^2)$, et on calculera le développement asymptotique de $U_k$ quand $k$ tend vers $+\infty$ à l'ordre $\frac{1}{k^2}$. Au vu de (2.7), et du Lemme 2.2.5,

$$u_k = (2k-1)\frac{1}{8k}\left(1 - \frac{1}{8k}\right) + \frac{1}{2}\left(\frac{1}{4} + C(\gamma_k^2)\right) + \frac{1}{2}(u_k^2 + v_k) + \mathcal{O}\left(\frac{1}{k^2}\right) \tag{2.9}$$

$$v_k = (2k-1)\frac{1}{64k^2}\left(1 - \frac{1}{8k}\right)^2 + \frac{1}{2}(C^2(\gamma_k^2) + C(\gamma_k^4)) + \frac{1}{2}(v_k^2 + \bar{g}_x(\gamma_k^4)) + \mathcal{O}\left(\frac{1}{k^2}\right) \tag{2.10}$$

Nous savons que $C(z) = 2kz + C(z)^2 + C(z^2)$, donc,

$$C(z^2) = 2kz^2 + C(z^2)^2 + C(z^4).$$

De plus, pour tout $z \leqslant \gamma_k$,

$$C(z^2) = \sum_{n\geqslant 0} C_{n,k} z^{2n} \leqslant z \sum_{n\geqslant 0} C_{n,k} z^{2n-1} \leqslant zC(z).$$

Donc, $C(\gamma_k^2) \leqslant \frac{1}{16k}$, et, asymptotiquement quand $k$ tend vers $+\infty$,

$$C(\gamma_k^4) \leqslant \frac{1}{64k^2}\frac{1}{16k} = \mathcal{O}\left(\frac{1}{k^3}\right).$$

Donc,

$$C(\gamma_k^2)(1-C(\gamma_k^2)) = 2k\frac{1}{64k^2}(1+\mathcal{O}\left(\frac{1}{k}\right)) + \mathcal{O}\left(\frac{1}{k^3}\right),$$

ce qui implique

$$C(\gamma_k^2) = \frac{1}{32k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

Si l'on utilise ces développements dans l'équation (2.10), on obtient $v_k = \frac{1}{32k} + \mathcal{O}\left(\frac{1}{k^2}\right)$. De même, l'Équation (2.9) permet d'obtenir $u_k = \frac{1}{2} - \frac{1}{4k} + \mathcal{O}\left(\frac{1}{k^2}\right)$. Finalement, les équations (2.7) et (2.8) permettent d'obtenir

$$V_k = \frac{1}{32k} - \frac{7}{1024k^2} + \mathcal{O}\left(\frac{1}{k^3}\right) \quad \text{et} \quad U_k = \frac{1}{2} - \frac{129}{1024k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

Dérivons $\overline{ST}^x(z)$ et $\bar{g}_x(z)$ :

$$\bar{g}_x'(z) = 2k - 1 + C(z)C'(z) + zC'(z^2) + \bar{g}_x(z)\bar{g}_x'(z) + z\bar{g}_x'(z^2),$$
$$(\overline{ST}^x)'(z) = 2k + C(z)C'(z) + zC'(z^2) + \overline{ST}^x(z)\bar{g}_x'(z) + z(\overline{ST}^x)'(z^2) - 2(\bar{g}_x(z) - \overline{ST}^x(z))(\bar{g}_x'(z) - (\overline{ST}^x)'(z)).$$

Cela implique, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{z\to\gamma_k}\frac{\bar{g}_x'(z)}{C'(z)} = \lim_{z\to\gamma_k}\frac{1}{1-\bar{g}_x(z)}\left(\frac{2k-1}{C'(z)} + C(z) + \frac{zC'(z^2)}{C'(z)} + \frac{z\bar{g}_x'(z^2)}{C'(z)}\right)$$
$$\sim \frac{1}{2(1-u_k)} = 1 - \frac{1}{2k} + \frac{1}{4k^2} + \mathcal{O}\left(\frac{1}{k^3}\right);$$

$$\begin{aligned}
X_n &:= \lim_{z\to\gamma_k}\frac{(\overline{ST}^x)'(z)}{C'(z)} \\
&= \lim_{z\to\gamma_k}\frac{1}{1-\overline{ST}^x(z)}\left(\frac{2k}{C'(z)} + C(z) + \frac{zC'(z^2)}{C'(z)} + \frac{z(\overline{ST}^x)'(z^2)}{C'(z)} - \frac{2(\bar{g}_x(z)-\overline{ST}^x(z))(\bar{g}_x'(z)-(\overline{ST}^x)'(z))}{C'(z)}\right) \\
&\sim \frac{1}{1-U_k}\left(\frac{1}{2} - 2(U_k-u_k)\lim_{z\to\gamma_k}\frac{(\overline{ST}^x)'(z)-\bar{g}_x'(z)}{C'(z)}\right) \\
&\sim \left(2 - \frac{129}{512k^2}\right)\left(\frac{1}{2} - \frac{1-X_n}{2k}\right),
\end{aligned}$$

ce qui implique

$$X_n = 1 - \frac{385}{512k^2} + \mathcal{O}\left(\frac{1}{n^3}\right).$$

■

**Lemme 2.3.23**

*Soit $G(z) = kST^x(z)$ où $ST^x(z)$ est la fonction génératrice des $ST^x_{n,k}$. Alors, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty}\frac{[z^n]G(z)}{C_{n,k}} = \mu_k(\mathcal{ST}) + \mathcal{O}\left(\frac{1}{k^2}\right).$$

**Démonstration :** Posons $G(z) = \sum_{i=1}^k ST^{x_i}(z) = kST^x(z)$ : cette fonction génératrice compte les tautologies simples mais compte plusieurs fois certaines d'entre elles. Une tautologie réalisée simultanément par $i$ variables booléennes distinctes est comptée $i$ fois dans $G(z)$. Autrement dit, si l'on note $K_n^i$ l'ensemble des tautologies simples réalisées par exactement $i$ variables simultanément :

$$[z^n]G(z) = \sum_{i=1}^k i|K_n^i| = ST_{n,k} + \sum_{i=2}^k (i-1)|K_n^i|.$$

Le langage de motifs $S = \bullet | S \vee S | \square \wedge \square$ permet de sélectionner les feuilles d'un arbre qui sont en position de réaliser des tautologies simples. En effet, un arbre de $K_n^i$ admet au moins $i$ $(S, \varnothing)$-restrictions. Donc, d'après le Lemme 2.3.12 appliqué au langage de motifs $S$,

$$\begin{aligned}\frac{\sum_{i=2}^{k}(i-1)|K_n^i|}{C_{n,k}} &\leqslant \sum_{i=2}^{k}(i-1)\frac{S[\mathcal{C}]_{n,k}^{[i]}}{C_{n,k}} \\ &\leqslant \frac{S[\mathcal{C}]_{n,k}^{[2]}}{C_{n,k}} + k\frac{S[\mathcal{C}]_{n,k}^{[\geqslant 3]}}{C_{n,k}} \\ &= \mathcal{O}\left(\frac{1}{k^2}\right).\end{aligned}$$

■

**Démonstration du Théorème 2.3.11 :** Les Lemmes 2.3.22 et 2.3.23 nous indiquent que

$$\mathbb{P}_k(\texttt{Vrai}) = \lim_{n\to+\infty}\frac{ST_{n,k}}{C_{n,k}} = \lim_{n\to+\infty}\frac{[z^n]G(z)}{C_{n,k}} + \mathcal{O}\left(\frac{1}{k^2}\right),$$

et

$$\lim_{n\to+\infty}\frac{[z^n]G(z)}{C_{n,k}} = k\lim_{n\to+\infty}\frac{ST_{n,k}^x}{C_{n,k}} \sim \frac{129}{1024k},$$

ce qui conclut la preuve. ■

### 2.3.3 Arbres commutatifs et associatifs

**Théorème 2.3.24**

*Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^g(\texttt{Vrai}) = \frac{(2\ln 2 - 1)^2}{4k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

Tout comme dans les parties précédentes, nous allons montrer que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple. Nous allons de nouveau utiliser des *mobiles*, ainsi que les langages de motifs $M$ et $R$ (cf. Équations (2.5) et (2.6)), utilisés pour les arbres associatifs plans. En préliminaire, nous allons montrer que ces deux langages de motifs sont sous-critiques pour la famille des arbres associatifs et commutatifs, au sens suivant :

**Lemme 2.3.25**

*Pour tout $j \geqslant 0$, on définit*

$$A_j(y) = [x^j]\hat{p}(2kx, y), \forall j \geqslant 0,$$

*où $\hat{p}(x, y)$ est la fonction génératrice du langage de motifs $\hat{N}$. Alors, les fonctions $(A_j(y))_{j\geqslant 0}$ sont sous-critiques pour la famille d'arbres booléens généraux $\mathcal{P}$. De plus, pour tout $j \geqslant 0$, on définit*

$$B_j(y) = [x^j]\check{p}(2kx, y), \forall j \geqslant 0,$$

*où $\check{p}(x, y)$ est la fonction génératrice du langage de motifs $\check{N}$. Alors, les fonctions $(B_j(y))_{j\geqslant 0}$ sont sous-critiques pour la famille d'arbres booléens généraux $\mathcal{P}$.*

**Démonstration :** Rappelons (cf. preuve du Lemme 2.3.10) que la fonction génératrice du langage de motifs $R$ est donnée par $p(x, y) = \hat{p}(x, y) + \check{p}(x, y) - x$ où

$$\hat{p}(x, y) = \frac{1}{2}\left(1 - y + x - \sqrt{(y - x - 1)^2 - 4x}\right).$$

Dès lors,

$$\begin{aligned}\hat{p}(2kx,y) &= \frac{1-y}{2} + \frac{2kx}{2} - \frac{1-y}{2}\sqrt{1+\frac{2kx(2kx-2(1-y)-2)}{(1-y)^2}} \\ &= \frac{1-y}{2} + kx - \frac{1-y}{2}\sum_{j\geqslant 0}\binom{1/2}{j}\frac{(2kx)^j(2kx-2(1-y)-2)^j}{(1-y)^{2j}}.\end{aligned}$$

Nous pouvons en déduire que les $(A_j(y))_{j\geqslant 0}$ sont des fonctions rationnelles de rayon de convergence 1. Comme $1 > \delta_k$, les $(A_j(y))_{j\geqslant 0}$ sont sous-critiques pour $\mathcal{P}$.

A partir de l'identité

$$\check{p}(2nx,y) = 2nx + \frac{\hat{p}^2(2nx,y)}{1-\hat{p}(2nx,y)},$$

en remarquant que $\hat{p}(0,y) = 0$ car aucun motif de $\hat{N}$ n'a aucune feuille de motif, nous en déduisons que les $(B_j(y))_{j\geqslant 0}$ s'écrivent comme sommes finies d'éléments de $(A_j(y))_{j\geqslant 0}$ et sont donc sous-critiques pour $\mathcal{P}$. ■

**Généralisation du lemme de Kozik**

**Lemme 2.3.26**

*Soit $L$ un langage de motifs de fonction génératrice $\ell(x,y)$. On définit*

$$A_j(y) = [x^j]\ell(2kx,y),$$

*et on suppose que ces fonctions sont sous-critiques pour $\mathcal{P}$.*

*Soit $\Gamma \subseteq \{x_1,\dots,x_k\}$ un ensemble dont le cardinal ne dépend pas de $k$. On note $L[\mathcal{P}]_{n,k}^{[p]}$ (resp. by $L[\mathcal{P}]_{n,k}^{[\geqslant p]}$) le nombre d'arbres de $L[\mathcal{P}]$ de taille $n$, dont on a étiqueté les feuilles de motifs, et qui ont exactement (resp. au moins) $p$ $(L,\Gamma)$-restrictions. De plus, on note $L[\mathcal{P}]_n$ le nombre d'arbres de $L[\mathcal{P}]$ de taille $n$.*

*Alors, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty}\frac{L[\mathcal{P}]_{n,k}^{[\geqslant p]}}{P_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right) \text{ et } \lim_{n\to+\infty}\frac{L[\mathcal{P}]_{n,k}^{[p]}}{P_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right).$$

La preuve de ce lemme est inspirée de celles des Lemmes 2.3.2 et 2.3.12. Il est facile de voir que le Lemme 2.3.13 reste vrai si l'on remplace toute occurrence de $\mathcal{C}$ par $\mathcal{P}$.

**Démonstration du Lemme 2.3.26 :** Le Lemme 2.3.13 nous donne

$$\frac{L[\mathcal{P}]_{n,k}^{[p]}}{P_{n,k}} = \frac{\sum_{\ell\in\mathbb{N}} L[\mathcal{P}]_n(\ell)w_{p,\gamma}(\ell)(k-\gamma)^{\underline{\ell-p}}2^\ell}{P_{n,k}};$$

ce qui implique

$$\frac{L[\mathcal{P}]_{n,k}^{[p]}}{P_{n,k}} \leqslant \frac{\sum_{\ell\in\mathbb{N}} L[\mathcal{P}]_n(\ell)w_{p,\gamma}(\ell)k^{\ell-p}2^\ell}{P_{n,k}} = \frac{\sum_{\ell\in\mathbb{N}}(2k)^\ell L[\mathcal{P}]_n(\ell)w_{p,\gamma}(\ell)}{P_{n,k}}\frac{1}{k^p}.$$

La Proposition 2.3.14 permet donc de conclure la preuve. ■

### Tautologies associatives et commutatives

Dans ce modèle, les tautologies simples sont des arbres enracinés par un $\vee$ et tels que deux feuilles de la première génération sont étiquetées par une variable et sa négation.

**Proposition 2.3.27**

*Dans le modèle associatif et commutatif, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple.*

De même que dans le modèle commutatif, les motifs d'un langage de motifs seront toujours planaires, ce qui leur permet de rester non ambigus, et nous devrons donc considérer des mobiles. Ainsi, si l'on considère le langage de motifs $R$, nous pourrons associer à tout arbre de $\mathcal{P}$ des semi-plongements de $R[\mathcal{P}]$. Bien entendu, il existe plusieurs semi-plongements d'un même arbre de $\mathcal{P}$, mais un mobile est le semi-plongement d'un unique arbre de $\mathcal{P}$.

Étant donné un ensemble $\Gamma$ de variables, nous appellerons semi-plongement minimal d'un arbre de $\mathcal{P}$, un semi-plongement qui minimise le nombre de ses $(R,\Gamma)$-restrictions. Tout comme dans le cas binaire, nous pouvons montrer que toute tautologie ayant exactement une $(R[R],\varnothing)$-restriction est une tautologie simple : nous ne détaillons pas cet argument. La preuve de la Proposition 2.3.27 est très similaire à celle de la Proposition 2.3.19 : elle est donc omise.

**Démonstration du Théorème 2.3.24 :** On note $ST^x(z)$ la fonction génératrice des tautologies simples réalisées par $x$ et telles que $x$ et $\bar{x}$ apparaissent chacun exactement une fois parmi les feuilles de la première génération.

$$\begin{aligned} ST^x(z) &= z^2 \sum_{\ell\geqslant 0} Z_\ell((\hat{P}(z)-2z, \hat{P}(z^2)-2z^2, \dots) \\ &= z^2 \exp\left(\sum_{\ell\geqslant 1} \frac{\hat{P}(z^\ell)-2z^\ell}{\ell}\right), \end{aligned} \tag{2.11}$$

où $Z_\ell(s_1, s_2, \dots)$ représente l'indice de cycle du groupe symétrique sur $\ell$ éléments (une introduction au comptage de Pólya, et aux indices de cycles peut être lue dans l'article [PR87] ou dans la thèse de V. Kraus [Kra11]). Dès lors,

$$\begin{aligned} (ST^x)'(z) &= 2z \exp\left(\sum_{\ell\geqslant 1} \frac{\hat{P}(z^\ell)-2z^\ell}{\ell}\right) + z^2 \exp\left(\sum_{\ell\geqslant 1} \frac{\hat{P}(z^\ell)-2z^\ell}{\ell}\right)\left(\sum_{\ell\geqslant 1} z^{\ell-1}(\hat{P}'(z^i)-2)\right) \\ &= \frac{2}{z} ST^x(z) + ST^x(z)\left(\hat{P}'(z) - 2 + \sum_{\ell\geqslant 2} z^{\ell-1}(\hat{P}'(z^i)-2)\right) \end{aligned}$$

En $z=\delta_k \sim \frac{2\ln 2-1}{2k}$ (la singularité de $P(z)$), $ST^x(z)$ est égal à

$$ST^x(\delta_k) = \delta_k^2 \underbrace{\exp\left(\sum_{i\geqslant 1} \frac{\hat{P}(\delta_k^i)}{i}\right)}_{=2} \underbrace{\exp\left(\sum_{i\geqslant 1} \frac{-2\delta_k^i}{i}\right)}_{=(1-\delta_k)^2\sim 1} \sim 2\delta_k^2,$$

Comme $P(z) = 2\hat{P}(z) - 2kz$,

$$\begin{aligned} \lim_{z\to\delta_k} \frac{(ST^x)'(z)}{P'(z)} &= \lim_{z\to\delta_k} \frac{ST^x(z)\hat{P}'(z)}{P'(z)} \\ &= \lim_{z\to\delta_k} \frac{ST^x(z)\hat{P}'(z)}{2\hat{P}'(z)-2k} = \frac{(2\ln 2-1)^2}{4k^2} \end{aligned}$$

et

$$\lim_{z\to\delta_k}\frac{(ST^x)'(z)}{P'(z)} = k\lim_{z\to\delta_k}\frac{(ST^x)'(z)}{P'(z)} \sim \frac{(2\ln 2-1)^2}{4k}.$$

La Proposition 2.3.27 permet de conclure la preuve. ■

## 2.4 Littéraux

L'objet de cette section est le calcul de la probabilité des fonctions littéral, c'est à dire des fonctions de complexité 1. Tout comme nous l'avons fait pour les tautologies, nous allons définir une sous-famille des arbres qui calculent une fonction littéral $((x_1,\dots,x_k)\mapsto x_\ell)$ ou $((x_1,\dots,x_k)\mapsto \bar{x}_\ell)$ $(\ell\in\{1,\dots,k\})$ : les simple-$x$.

La définition suivante est le *dual* des tautologies et tautologies simples.

**Définition 2.4.1**

*Une* **contradiction** *est un arbre booléen qui calcule la fonction constante* `Faux`*. Une* **contradiction simple** *est un arbre et/ou dont deux feuilles reliées à la racines par un chemin de $\wedge$ sont étiquetées respectivement par un littéral $\alpha$ et sa négation.*

**Définition 2.4.2 (cf. Figure 2.6)**

*Un* **simple-$x$ de type** $T$ *est un arbre dont la racine a deux descendants, dont l'un des deux sous-arbres est réduit à une feuille étiquetée par un littéral $x$, dont la racine est étiquetée par $\wedge$ (resp. $\vee$), et dont l'autre sous-arbre est une tautologie simple (resp. une contradiction simple).*

*Un* **simple-$x$ de type** $X$ *est un arbre dont la racine a deux descendants, dont l'un des deux sous-arbres est réduit à une feuille étiquetée par un littéral $x$, dont la racine est étiquetée par $\wedge$ (resp. $\vee$), et dont l'autre sous-arbre a une feuille étiquetée par $x$ reliée à sa racine par un chemin de $\vee$.*

*Un* **simple-$x$** *est soit un simple-$x$ de type $T$, soit un simple-$x$ de type $X$.*

**Remarque :** Remarquons qu'un simple-$x$ de type $X$ dans le cas associatif est un arbre dont la racine a deux descendants, dont l'un des deux sous-arbres est réduit à une feuille étiquetée par un littéral $x$, dont la racine est étiquetée par $\wedge$ (resp. $\vee$), et dont l'autre sous-arbre, enraciné par $\vee$ (car l'arbre est stratifié) a une feuille de première génération étiquetée par $x$.

Dans les différents modèles, nous montrerons que

**Proposition 2.4.3**

*Asymptotiquement quand $k$ tend vers $+\infty$, presque tout arbre calculant $x$ est un simple-$x$.*

Nous ne détaillerons pas la preuve de cette proposition dans les différents modèles car elle se fait par des arguments similaires à ceux utilisés pour montrer qu'asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple. Nous nous contenterons donc de calculer dans chaque modèle la fraction limite des simple-$x$. Nous verrons en fin de chapitre comment le cas des fonctions littéral peut être vu comme un cas particulier du cas d'une fonction générale (cf. Section 2.6).

Dans le cas binaire plan, nous avons le résultat suivant :

**Théorème 2.4.4 ([GGKM13])**

*Dans le modèle binaire plan, la probabilité de la fonction littéral $((x_1,\dots,x_k)\mapsto x_\ell)$ (ou celle de la fonction littéral $((x_1,\dots,x_k)\mapsto \bar{x}_\ell)$), pour tout $\ell\in\{1,\dots,k\}$ vérifie, asymptotiquement*

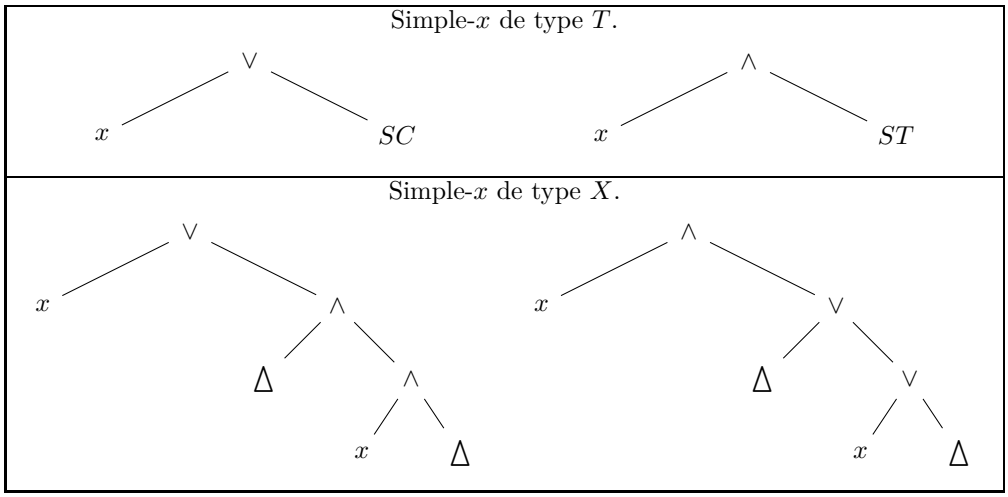



FIGURE 2.6 – *Les simple-$x$ dans le cas binaire. La notation $ST$ représente une tautologie simple et la notation $SC$ représente une contradiction simple.*

*quand $k$ tend vers $+\infty$ :*

$$\mu_k(x_\ell) = \frac{5}{16k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

### 2.4.1 Arbres associatifs

**Théorème 2.4.5**

*Dans le modèle associatif plan, la probabilité de la fonction littéral $((x_1, \dots, x_k) \mapsto x)$, pour tout $x \in \{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$, vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^a(x) = \frac{546 - 386\sqrt{2}}{k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

**Démonstration :** Calculons tout d'abord la fraction limite des simple-$x$ de type $T$. Les tautologies simples, tout comme les contradictions simples, sont énumérées par la série génératrice $ST(z)$. Dès lors, la série génératrice des simple-$x$ de type $T$ est donnée par $SX_T(z) = 4zST(z)$ où le terme $4z$ compte les différents étiquetages de la racine et le choix du sous-arbre réduit à une feuille $x$ (notons qu'un simple-$x$ a par définition une racine binaire). Dès lors, au vu de l'Équation (2.2),

$$\lim_{n\to+\infty} \frac{[z^n]SX_T(z)}{A_{n,k}} = \lim_{z\to\alpha_k} 4z \lim_{z\to\alpha_k} \frac{ST'(z)}{A'(z)} = 4\alpha_k\mu_k^a(\texttt{Vrai}) = \frac{594 - 420\sqrt{2}}{k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

La contribution des simple-$x$ de type $X$ est donnée par la série génératrice $SX_X(z) = 4zg(z)$, où $g(z)$ compte les arbres enracinés par $\vee$ et dont exactement une feuille de première génération est étiquetée par $x$, et tels qu'aucune autre feuille de première génération est étiquetée par $\bar{x}$ (sinon, nous compterions de nouveau les simple-$x$ de type $T$). La fonction $g(z)$ est donc donnée par

$$g(z) = z\sum_{\ell\geqslant 2} \ell(\hat{A}(z) - 2z)^{\ell-1} = \frac{1}{(1 - \hat{A}(z) + 2z)^2}.$$

Après calcul, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{n\to+\infty}\frac{[z^n]SX_X(z)}{A_{n,k}} = 4\alpha_k \lim_{z\to\alpha_k}\frac{g'(z)}{A'(z)} = \frac{2(3-2\sqrt{2})}{k}\frac{3\sqrt{2}-2}{k} + \mathcal{O}\left(\frac{1}{k^3}\right) = \frac{34\sqrt{2}-48}{k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

Le Théorème 2.4.5 est donc prouvé par addition des deux fractions limites. ■

### 2.4.2 Arbres commutatifs

**Théorème 2.4.6**

*La probabilité d'une fonction littéral $((x_1,\ldots,x_k)\mapsto x)$, pour tout $x\in\{x_1,\bar{x}_1,\ldots,x_k,\bar{x}_k\}$ vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^c(x) = \frac{1153}{4096k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

**Démonstration :** La fonction génératrice qui compte les simple-$x$ de type $T$ est donnée par $SX_T(z) = 2zST(z)$ où le facteur $2z$ compte les deux choix pour l'étiquetage de la racine (il n'y a plus de différence entre sous-arbre gauche et sous-arbre droit). Dès lors,

$$\lim_{n\to+\infty}\frac{[z^n]SX_X(z)}{C_{n,k}} = \lim_{z\to\gamma_k} 2z\frac{ST'(z)}{C'(z)} = 2\gamma_k\mu_k(\texttt{Vrai}) = \frac{641}{4096k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

La fonction génératrice qui compte les simple-$x$ de type $X$ est donnée par $SX_X(z) = 2zg(z)$ où $g(z)$ est la fonction génératrice des arbres dont une feuille étiquetée par $x$ est relié à la racine par un chemin de $\vee$. Cette fonction génératrice, ou plutôt son complément $\bar{g}(z)$ a déjà été calculée (cf. Équation (2.8)). Le calcul de $\lim_{z\to\gamma_k}\frac{\bar{g}'(z)}{C'(z)}$ est détaillé dans la preuve du Théorème 2.3.11. En réutilisant ces résultats, nous obtenons que

$$\lim_{n\to+\infty}\frac{[z^n]SX_X(z)}{C_{n,k}} = 2\gamma_k \lim_{n\to+\infty}\frac{[z^n]g_x(z)}{[z^n]C(z)} = \frac{1}{4k}\left(1+\frac{1}{8k}\right)\frac{1}{2k} + \mathcal{O}\left(\frac{1}{k^3}\right) = \frac{1}{8k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

En sommant les contributions des simple-$x$ de type $T$ et de type $X$, nous concluons la preuve du Théorème 2.4.6. ■

### 2.4.3 Arbres associatifs et commutatifs

**Théorème 2.4.7**

*La probabilité de la fonction littéral $((x_1,\ldots,x_k)\mapsto x)$, pour tout $x\in\{x_1,\bar{x}_1,\ldots,x_k,\bar{x}_k\}$ vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^g(x) = \frac{(2\ln 2-1)^2(2\ln 2+1)}{4k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

**Démonstration :** Les simple-$x$ de type $T$ sont comptés par la fonction génératrice $SX_T(z) = 2zST(z)$. Dès lors,

$$\frac{[z^n]SX_T(z)}{P_{n,k}} = \lim_{z\to\delta_k} 2z\frac{ST'(z)}{P'(z)} = 2\delta_k\mu_k^g(\texttt{Vrai}) = \frac{(2\ln 2-1)^3}{8k^2} + \mathcal{O}\left(\frac{1}{k^3}\right).$$

De plus, la série génératrice des simple-$x$ de type $X$ est donnée par $SX_X(z) = 2zg_x(z)$ où $g_x(z)$ est la série génératrice des arbres enracinés par $\vee$, dont une feuille de première génération est étiquetée

par $x$ et tels qu'aucune autre feuille de première génération n'est étiquetée par $x$ ou par $\bar{x}$. Par la méthode symbolique, nous avons

$$g_x(z) = z + z\left(\exp\left(\sum_{\ell\geqslant 1}\frac{\hat{P}(z^\ell)-2z^\ell}{\ell}\right)-1\right),$$

et

$$g_x'(z) = 1+\frac{1}{z}(g_x(z)-z)+g_x(z)\left(\sum_{\ell\geqslant 1} z^{\ell-1}(\hat{P}(z^\ell)-2)\right).$$

Comme $g_x(\delta_k)\sim 2\delta_k$, nous obtenons :

$$\lim_{z\to\delta_k}\frac{g_x'(z)}{P'(z)} = \lim_{z\to\delta_k}\frac{g_x(z)\hat{P}'(z)}{P'(z)} = \frac{2\delta_k}{2} = \frac{2(\ln 2-1)}{2k}+\mathcal{O}\left(\frac{1}{k^2}\right),$$

ce qui implique

$$\frac{[z^n]SX_X(z)}{P_{n,k}} \sim 2\delta_k \lim_{z\to\delta_k}\frac{g_x'(z)}{P'(z)} = \frac{(2\ln 2-1)^2}{4n^2}+\mathcal{O}\left(\frac{1}{n^3}\right). \qquad \blacksquare$$

## 2.5 Probabilité d'une fonction quelconque

Dans les parties précédentes, nous avons étudié les fonctions de complexité 0 et 1. Dans cette partie, nous nous intéressons à une fonction de complexité arbitraire. Dans sa preuve du Théorème 1.3.1, Kozik a montré que, asymptotiquement quand $k$ tend vers $+\infty$, presque tout arbre calculant une fonction booléenne fixée $f$ a une forme *simple*. Plus précisément, c'est un arbre minimal de $f$ dans lequel a été greffée une unique *expansion*. Dans cette partie, nous généraliserons cette approche aux modèles associatifs et commutatifs.

### 2.5.1 Arbres associatifs

Nous montrons le théorème suivant :

**Théorème 2.5.1**

*Dans le modèle associatif plan, fixons $f$ une fonction de complexité $L(f)\geqslant 2$. Alors, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^a(f) \sim \frac{\lambda_f^a}{k^{L(f)+1}},$$

*où $\lambda_f^a > 0$ est une constante. De plus,*

$$\left(\frac{3-2\sqrt{2}}{2}\right)^k\left[(145-102\sqrt{2})L(f)+153-108\sqrt{2}\right]m_f \leqslant \lambda_f$$

$$\lambda_f \leqslant \left(\frac{3-2\sqrt{2}}{2}\right)^k\left[(9\sqrt{2}-12)L(f)^2+(247-174\sqrt{2})L(f)+(36\sqrt{2}-51)\right]m_f,$$

*où $m_f$ est le nombre d'arbres minimaux de $f$.*

Pour montrer le théorème 2.5.1, nous procédons comme dans le papier de Kozik [Koz08].

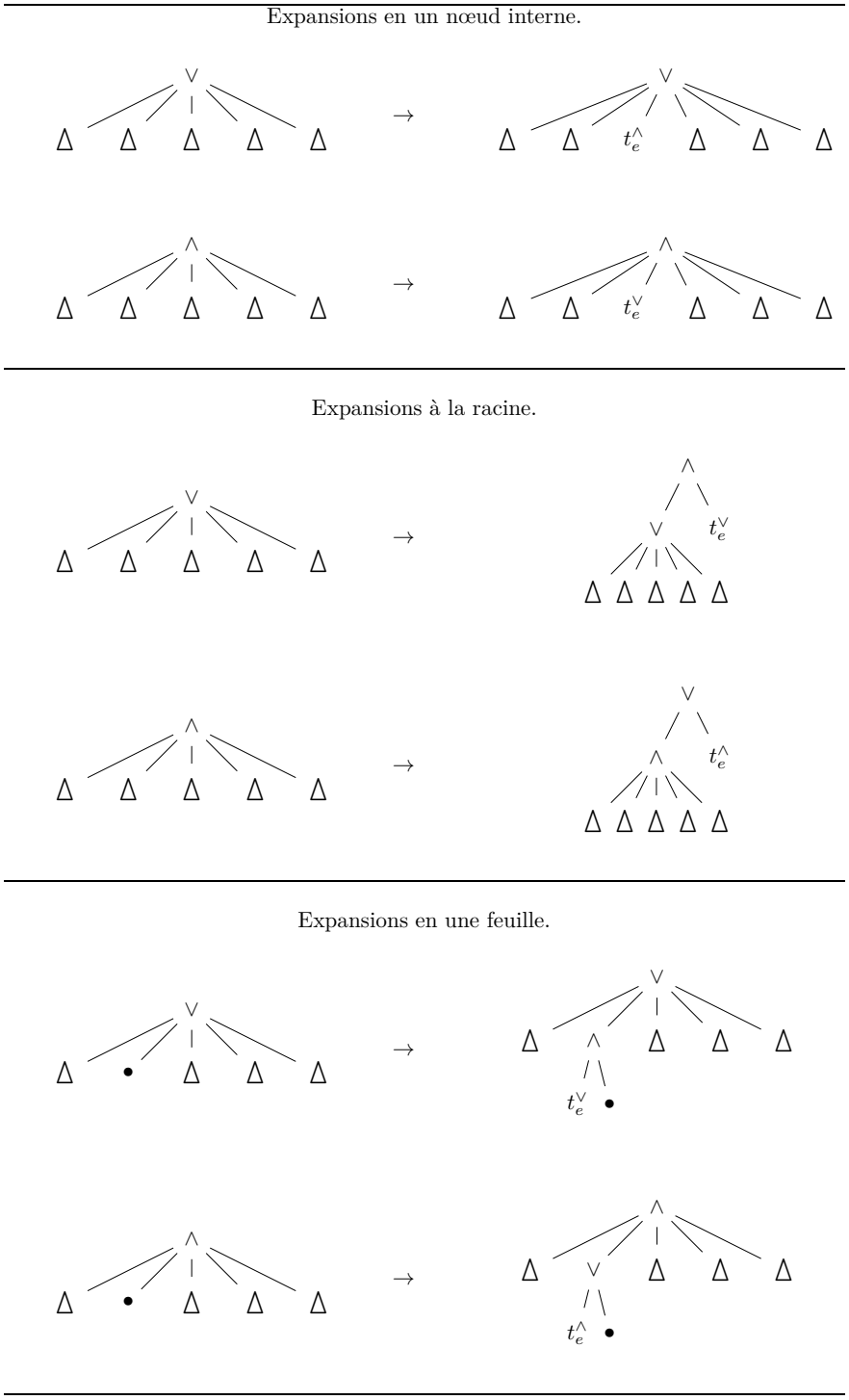



FIGURE 2.7 – *Les différentes expansions valides dans le modèle associatif. Ici, $t_e^{\vee}$ (resp. $t_e^{\wedge}$) est un arbre associatif enraciné par $\vee$ (resp. $\wedge$). L'arbre représenté dans les dessins est uniquement l'arbre enraciné en $\nu$ avant et après expansion.*

**Expansions associatives**

Au vu de la structure stratifiée des arbres associatifs, la définition des expansions est différente de celle utilisée par Kozik : l'arbre greffé doit respecter la stratification.

**Définition 2.5.2 (cf. Figure 2.7)**

*Soit $t$ un arbre et/ou associatif calculant $f$. On définit deux types d'expansions de $t$ :*
- *Soit $\nu$ un nœud interne de $t$ (éventuellement sa racine). On note $t_1, \ldots, t_j$ ($j \geqslant 2$) les sous-arbres de $\nu$. Une expansion de premier type de $t$ en $\nu$ est un arbre obtenu en ajoutant un sous-arbre $t_e$ à $\nu$.*
- *Soit $\nu$ la racine de $t$. L'arbre obtenu en remplaçant le sous-arbre $t$ par $t_e \diamond t$, où $\diamond$ est un connecteur choisi entre $\wedge$ et $\vee$ de telle sorte que l'arbre obtenu soit stratifié, est une expansion de second type de $t$.*

*On dira qu'une telle expansion est* **valide** *quand l'arbre expansion calcule toujours $f$.*

**Remarque :** A la racine, les deux types d'expansions sont possibles, et les expansions de second type ne sont possibles qu'en la racine.

**Proposition 2.5.3**

*Asymptotiquement quand $k$ tend vers $+\infty$, presque tout arbre calculant $f$ est une expansion valide d'un arbre minimal de $f$ de l'un des deux types suivants :*
- *les $T$-expansions : une expansion valide est une $T$-expansion si l'arbre greffé $t_e$ est une tautologie simple (resp. une contradiction simple), et si la nouvelle étiquette de $\nu$ est $\wedge$ (resp. $\vee$) ;*
- *les $X$-expansions : une expansion valide est une $X$-expansion s'il existe une variable essentielle $x$ de $f$ telle que l'arbre greffé $t_e$ a une feuille étiquetée par $x$ ou $\bar{x}$ reliée à la racine par un chemin de $\wedge$ (resp. $\vee$), et si la nouvelle étiquette de $\nu$ est $\vee$ (resp. $\wedge$).*

Avant de développer la preuve de ce résultat, nous devons introduire les langages de motifs nécessaires :

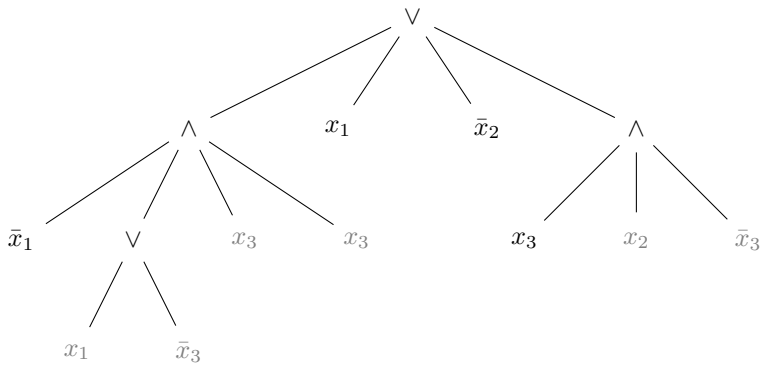
$$\begin{cases} \hat{P} = \bullet | \check{P} \wedge \check{P} | \check{P} \wedge \check{P} \wedge \check{P} | \ldots \\ \check{P} = \bullet | \hat{P} \vee \square | \hat{P} \vee \square \vee \square | \ldots \\ Q = \{\hat{P}, \check{P}\}; \\ \\ \hat{N} = \bullet | \check{N} \wedge \square | \check{N} \wedge \square \wedge \square | \ldots \\ \check{N} = \bullet | \hat{N} \vee \hat{N} | \hat{N} \vee \hat{N} \vee \hat{N} | \ldots \\ R = \{\hat{N}, \check{N}\}, \end{cases} \tag{2.12}$$

**Remarque :**
- Le langage de motifs $Q$ vérifie la propriété suivante : si toutes les $Q$-feuilles de motif d'un arbre $t$ sont affectées à `Vrai`, alors $t$ calcule `Vrai` pour cette affectation partielle des variables.
- Le langage de motifs $R$ vérifie la propriété suivante : si toutes les $R$-feuilles de motif d'un arbre $t$ sont affectées à `Faux`, alors $t$ calcule `Faux` pour cette affectation partielle des variables.

**Définition 2.5.4 (cf. Figure 2.8)**

*Soit $L$ un langage de motif. Rappelons que l'on note $L^r$ la composée $r$ fois du langage de motifs $L$ (cf. Définition 2.3.17). Si une feuille d'un arbre $t$ est une $L^r$-feuille de motif mais non une $L^{r-1}$ feuille de motif, on dira que cette feuille est une feuille de motif de niveau $r$.*

FIGURE 2.8 – *Dans l'arbre ci-dessous, nous avons colorié en noir les R-feuilles de motif de premier niveau et en gris les R-feuilles de motif de niveau 2. Cet arbre n'a pas de R-feuilles de motif de niveau trois. La Figure 2.3 peut aider à comprendre cet exemple.*




**Démonstration de la Proposition 2.5.3 :** Cette preuve est inspirée de celle développée par Kozik pour les arbres binaire plans [Koz08, Théorème 6.1]. Soit $f$ une fonction booléenne fixée de complexité $L(f)$, nous noterons $\Gamma_f$ l'ensemble de ses variables essentielles. Nous allons devoir utiliser les langages de motifs $L = R^{(L(f)+1)}[R \oplus Q]$ et $\bar{L} = R^{(L(f)+1)}[(R \oplus Q)^2]$, où la somme de deux langages de motifs $R \oplus Q$ est définie comme suit : les $(R \oplus Q)$-feuilles de motif d'un arbre $t$ sont ses $Q$-feuilles de motif et ses $R$-feuilles de motif. On peut démontrer que si $R$ et $Q$ sont sous-critiques pour la famille $\mathcal{A}$, alors $R \oplus Q$ l'est aussi. Le langage de motifs $L$ et toutes ses puissances sont donc sous-critiques pour la famille $\mathcal{A}$.

Montrons tout d'abord qu'un arbre typique calculant $f$ a exactement $L(f)+1$ $(L, \Gamma_f)$-restrictions. Soit $t$ un arbre minimal de $f$. Considérons la famille des arbres dont la racine est étiquetée par $\wedge$ (resp. $\vee$), ayant deux sous-arbres, l'un d'entre eux étant $t$ et le second étant une tautologie simple (resp. une contradiction simple). La série génératrice de cette famille est donnée par $4z\, z^{L(f)} ST(z)$ et tous les arbres de cette famille calculent $f$. Dès lors, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k^a(f) \geqslant 4\alpha_k^{L(f)+1} \mu_k(\texttt{Vrai}) \sim cst \cdot \frac{1}{k^{L(f)+1}}.$$

Cette borne inférieure nous indique que la famille des arbres calculant $f$ et ayant au moins $L(f)+2$ $\bar{L}$-restrictions est négligeable. Il nous suffit donc de considérer les arbres calculant $f$ et ayant au plus $L(f)+1$ $\bar{L}$-restrictions.

La seconde étape consiste à montrer qu'un arbre calculant $f$ ne peut avoir moins de $L(f)+1$ $L$-restrictions. Cette preuve est identique à celle développée dans [Koz08], nous ne la développons pas ici.

Enfin, considérons un arbre calculant $f$ ayant exactement $L(f)+1$ $\bar{L}$-restrictions, et montrons que cet arbre est une $T$-expansion ou une $X$-expansion d'un arbre minimal de $f$. Nous savons que les variables qui apparaissent comme étiquettes des feuilles de motif du niveau $L(f)+3$ ne sont pas des variables essentielles de $f$, et ne sont pas répétées dans les $\bar{L}$-feuilles de motif. Dès lors, chaque sous-arbre de $t$ enraciné au niveau $L(f)+3$ et dont le parent est au niveau $L(f)+2$ peut être remplacé par une $\star$, signifiant ainsi qu'il peut être affecté à `Vrai` ou `Faux` indépendamment des autres $\bar{L}$-feuilles de motif. En effet, il suffit d'affecter soit toutes les $R$-feuilles de motifs de ces arbres à `Faux`, soit toutes leurs $Q$-feuilles de motif à `Vrai`. Après cette opération, toutes les feuilles de l'arbre, qui ne sont pas étiquetées par $\star$, sont des $\bar{L}$-feuilles de motif.

Remplaçons enfin par une $\star$ toute $\bar{L}$-feuille de motif de $t$ étiquetée par une variable ni essentielle ni répétée. L'arbre obtenu après toutes ces opérations sera noté $t^\star$. Cet arbre n'est bien sûr pas un arbre et/ou puisque certaines de ses feuilles sont étiquetées par $\star$. Remarquons que cet arbre calcule

toujours $f$, quelles que soient les valeurs des différentes $\star$.

Simplifions l'arbre de façon à éliminer les $\star$, sans toutefois changer la fonction calculée par l'arbre : pour cela, nous suivons les règles suivantes :

$$\star \vee \ldots \vee \star \equiv \star \qquad \star \wedge \ldots \wedge \star \equiv \star \tag{2.13}$$

$$\star \vee \ldots \star \vee t_1 \vee \ldots \vee t_j \equiv \texttt{Vrai} \qquad \star \wedge \ldots \star \wedge t_1 \wedge \ldots \wedge t_j \equiv \texttt{Faux} \tag{2.14}$$

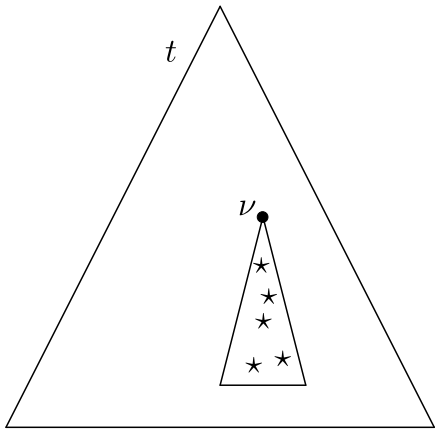
$$\texttt{Vrai} \vee t_1 \vee \ldots \vee t_j \equiv \texttt{Vrai} \qquad \texttt{Faux} \wedge t_1 \wedge \ldots \wedge t_j \equiv \texttt{Faux} \tag{2.15}$$

$$\texttt{Faux} \vee t_1 \vee \ldots \vee t_j \equiv t_1 \vee \ldots \vee t_j \qquad \texttt{Vrai} \wedge t_1 \wedge \ldots \wedge t_j \equiv t_1 \wedge \ldots \wedge t_j \tag{2.16}$$

où $t_1, \ldots, t_j$ sont des sous-arbres ne contenant pas de $\star$.

L'arbre noté $\hat{t}$ que nous obtenons après ces simplifications calcule toujours $f$ : montrons que c'est un arbre minimal de $f$ et que le dernier ancêtre commun [2] des $\star$ dans $t^\star$ a été simplifié (cf. Figure 2.9).

FIGURE 2.9 – *Le nœud $\nu$ est le dernier ancêtre commun des $\star$.*




L'arbre $t^\star$ contient au moins une $\star$ car $t$ est suffisamment grand pour avoir au moins une feuille de motif au niveau $L(f)+3$. Remarquons de plus que l'arbre $t^\star$ a exactement $L(f)+1$ $\bar{L}$-restrictions. Notons qu'une règle de type (2.14) a dû être appliquée au moins une fois car seules ces règles permettent de supprimer une $\star$. Mais appliquer une de ces règles simplifie de surcroit au moins une autre feuille de motif, non étiquetée par une $\star$, i.e. une feuille de motif étiquetée par une variable soit essentielle, soit répétée. Dès lors, les simplifications ont au moins supprimé une restriction et l'arbre $\hat{t}$ a au plus $L(f)$ feuilles : c'est donc un arbre minimal de $f$.

Soit $\nu$ le dernier ancêtre commun des $\star$ dans $t^\star$. Supposons que $\nu$ est toujours dans $\hat{t}$ : cela implique qu'au moins deux $\star$ ont été simplifiées indépendamment. Autrement dit, au moins deux règles de type (2.14) ont été appliquées, et deux restrictions ont donc été simplifiées durant le processus de simplification. L'arbre simplifié $\hat{t}$ calcule $f$ et contient au plus $L(f)-1$ restrictions, donc au plus $L(f)-1$ feuilles, ce qui est impossible car la complexité de $f$ est $L(f)$. Nous avons donc montré que $\nu$ doit être simplifié lors de la simplification des $\star$.

Remarquons enfin que la dernière règle que l'on applique lors de la simplification doit être de type (2.16). Nous avons donc montré qu'un arbre typique calculant $f$ est en effet une expansion d'un arbre minimal de $f$. Il nous reste à comprendre quelles sont les expansions typiques valides d'un arbre minimal de $f$.

Le lemme de Kozik 2.3.2 nous assure que la famille des expansions d'un arbre minimal de $f$ telles que l'arbre greffé $t_e$ a au moins 2 $((R \oplus Q)^2, \Gamma_f)$-restrictions est de fraction limite négligeable

2. on dira que le dernier ancêtre commun d'un ensemble de nœuds est le nœud $\nu$ le plus éloigné de la racine tel que le sous-arbre enraciné en $\nu$ contient tous les nœuds considérés.

asymptotiquement quand $k$ tend vers $+\infty$. Supposons par ailleurs que $t_e$ n'a pas de $((R \oplus Q)^2, \Gamma_f)$-restriction. Alors, $t_e$ peut être affecté à la valeur `Vrai` ou à la valeur `Faux` indépendamment du reste de l'arbre. Nous pouvons donc choisir la valeur de $t_e$ de façon à simplifier son ancêtre, et donc au moins une autre feuille de l'arbre. Cette simplification nous donne un arbre de taille strictement inférieure à $L(f)$, et qui calcule $f$. C'est impossible. Il nous suffit donc de considérer les expansions d'arbres minimaux de $f$ telles que l'arbre greffé $t_e$ a exactement une $((R \oplus Q)^2, \Gamma_f)$-restriction.

Supposons dans un premier temps que $t_e$ contient une $(R \oplus Q)^2$-répétition. Alors, $t_e$ doit calculer une fonction constante `Vrai` ou `Faux`. En effet, si l'on suppose que cet arbre ne calcule pas une fonction constante, alors il peut être affecté à `Vrai` ou `Faux`, et ce indépendamment du reste de l'arbre. Nous pouvons donc effectuer une simplification de l'ancêtre de $t_e$, et donc d'au moins une feuille extérieure à $t_e$. L'arbre obtenu après simplification calcule $f$ et est de taille strictement inférieure à $L(f)$, ce qui est impossible. Dès lors, $t_e$ est une tautologie ou une contradiction, typiquement une tautologie simple ou une contradiction simple, et comme cette expansion doit être valide, $t$ est une $T$-expansion d'un arbre minimal de $f$.

Supposons par ailleurs que $t_e$ ne contienne aucune répétition mais une occurrence d'une variable essentielle $x$ de $f$ parmi ses $(R \oplus Q)^2$-feuilles de motifs. Cette variable essentielle doit apparaître à la première génération de $t_e$. Sinon, $t_e$ est de la forme $s_1 \wedge (s_2 \vee x)$ ou $s_1 \vee (s_2 \wedge x)$ où $s_1$ et $s_2$ n'ont aucune $(R \oplus Q, \Gamma_f)$-restriction. Nous pouvons donc affecter $s_1$ et $s_2$ à `Vrai` ou `Faux` indépendamment l'un de l'autre, et indépendamment du reste de l'arbre. Nous pouvons donc affecter $t_e$ à `Vrai` ou `Faux` indépendamment du reste de l'arbre et simplifier ainsi son ancêtre ainsi qu'une feuille extérieure à $t_e$. Nous obtenons donc un arbre calculant $f$ et de taille strictement inférieure à $L(f)$, ce qui est impossible. ■

### 2.5.2 Arbres commutatifs

**Théorème 2.5.5**

*Dans le modèle commutatif binaire, soit $f$ une fonction booléenne non constante, de complexité $L(f)$. Alors il existe une constante $\lambda_f^c > 0$ telle que, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^c(f) \sim \frac{\lambda_f^c}{k^{L(f)+1}}.$$

*La constante $\lambda_f^c$ dépend du nombre d'expansions valides d'arbres minimaux de $f$, et*

$$\frac{2306L(f) - 641}{1024 \cdot 8^{L(f)}} m_f \leqslant \lambda_f^c \leqslant \frac{(2L(f) - 1)(1024L(f) + 641)}{1024 \cdot 8^{L(f)}} m_f,$$

*où $m_f$ est le nombre d'arbres minimaux de $f$.*

**Remarque :** Il est intéressant de voir que lorsque $L(f) = 1$, les bornes de $\lambda_f^c$ sont égales.

**Démonstration :** La preuve est très similaire à celle effectuée dans Kozik [Koz08], qui est elle-même presque identique à celle développée dans le cadre du modèle associatif ci-dessus. Les langages de motifs utilisés sont ici $N$ et $P$ définis comme suit :

$$N = \bullet | N \vee N | N \wedge \square$$
$$P = \bullet | P \vee \square | P \wedge P.$$

Lorsque l'on affecte toutes les $N$-feuilles de motif (resp. $P$-feuilles de motif) d'un arbre à `Faux` (resp. `Vrai`), cet arbre calcule `Faux` (resp. `Vrai`) indépendamment des autres variables non affectées. Les lemmes 2.3.18 et 2.3.15 montrent que la Proposition 2.3.14 peut être appliquée au langage de motifs

$N \oplus P$ car la fonction génératrice du langage $P$ est la même que celle du langage $N$. On notera $L = N^{(L(f))}[N \oplus P]$ et $\bar{L} = N^{(L(f))}[(N \oplus P)^2]$.

Soit $f$ une fonction booléenne fixée. Nous noterons $L(f)$ sa complexité et $\Gamma_f$ l'ensemble des ses variables essentielles. Nous pouvons montrer (les détails sont omis) qu'un arbre typique calculant $f$ est un arbre qui a exactement $L(f)+1$ $(\bar{L}, \Gamma_f)$-restrictions. De plus, un arbre ayant exactement $L(f)+1$ $(\bar{L}, \Gamma_f)$-restrictions est une expansion d'un arbre minimal de $f$. Enfin, on montre qu'un arbre typique calculant $f$ est une expansion valide de type $T$ ou de type $X$ d'un arbre minimal de $f$. Tous ces résultats se montrent comme dans le cas associatif, en travaillant sur un semi-plongement minimal de $t$ dans $L[\mathcal{C}]$ ou $\bar{L}[\mathcal{C}]$, et en utilisant la Proposition 2.3.14.

On note $w_1^c$ la fraction limite des tautologies simples. Fixons un littéral $x$, $w_2^c$ sera la fraction limite de la famille des arbres tels que $x$ apparaît comme étiquette d'une feuille de première génération. Au vu de la Partie 2.3.2, $w_1^c \sim \frac{641}{1024k}$, et au vu de la Partie 2.4.2, $w_2^c \sim \frac{1}{2k}$. Donc, comme un arbre typique calculant $f$ est une $X$-expansion ou une $T$-expansion d'un arbre minimal calculant $f$,

$$\mu_k^c(f) \sim \gamma_k^{L(f)}(\lambda_T w_1^c + \lambda_X w_2^c),$$

où $\lambda_T$ (resp. $\lambda_X$) est la somme sur tous les arbres minimaux de $f$ du nombre d'emplacements où une expansion de type $T$ peut être réalisée dans l'arbre minimal considéré (resp. où une expansion de type $X$ peut être réalisée pour chaque littéral possible). Il est facile de voir qu'une $T$-expansion est possible en chaque nœud d'un arbre minimal (aussi bien en greffant une tautologie simple qu'une contradiction simple). Comme chaque arbre minimal a $2L(f)-1$ nœuds,

$$\lambda_T = (2L(f)-1)m_f.$$

En chaque feuille d'un arbre minimal, au moins deux $X$-expansions sont possibles (l'arbre greffé peut être enraciné par $\wedge$ ou par $\vee$), donc

$$2L(f)m_f \leqslant \lambda_X$$

On peut au maximum faire $2L(f)$ expansions en chaque nœud, nous avons

$$\lambda_X \leqslant 2L(f)^2 m_f.$$

Il ne reste plus qu'à rappeler que $\gamma_k \sim \frac{1}{8k}$ pour conclure la preuve. ■

### 2.5.3 Arbres associatifs et commutatifs

**Théorème 2.5.6**

*Dans le modèle associatif et commutatif, pour toute fonction booléenne fixée $f$ de complexité notée $L(f)$, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k^g(f) \sim \frac{\lambda_f^g}{k^{L(f)+1}},$$

*où $\lambda_f^g$ est une constante. On a*

$$\left(\frac{2\ln 2 - 1}{2}\right)^k \left(\left(\ln^2 2 - \frac{1}{4}\right) L(f) + \ln^2 2 - 2\ln 2 + \frac{1}{2}\right) m_f \leqslant \lambda_f^g$$

$$\lambda_f^g \leqslant \left(\frac{2\ln 2 - 1}{2}\right)^k \frac{(2\ln 2 - 1)(L(f) + 1 + 4\ln 2)L(f)}{4},$$

*où $m_f$ est le nombre d'arbres minimaux de $f$.*

**Démonstration :** Ce résultat se montre en appliquant les arguments des Sections 2.5.2 et 2.5.1, et en utilisant les langages de motifs $L = R^{(L(f))}[R \oplus Q]$ et $\bar{L} = R^{(L(f))}[(R \oplus Q)^2]$. Nous obtenons que, pour toute fonction booléenne $f$ fixée,

$$\mu_k^g(f) = \delta_k^{L(f)}(\lambda_T^g w_1^g + \lambda_X^g w_2^g).$$

Nous savons que $w_1^g \sim \frac{(2\ln 2-1)^2}{4k}$ (cf. Section 2.3), et $w_2^{a,c} \sim \frac{2\ln 2-1}{4k}$ (cf. Section 2.4). De plus, $\delta_k \sim \frac{2\ln 2-1}{2k}$ asymptotiquement quand $k$ tend vers $+\infty$, et nous pouvons montrer que :

$$(L(f)+2)m_f \leqslant \lambda_T^g \leqslant 2L(f)m_f$$
$$2L(f)m_f \leqslant \lambda_X^g \leqslant (L(f)^2 + 3L(f))m_f$$

ce qui conclut la preuve. ■

## 2.6 Conclusion

Nous avons compris dans ce chapitre l'influence de l'associativité et de la commutativité des connecteurs logiques sur le comportement de la loi induite sur l'ensemble des fonctions booléennes. Nous avons montré que prendre en compte les propriétés logiques des connecteurs $\wedge$ et $\vee$ ne change pas le comportement global de la distribution induite sur l'espace des fonctions booléennes : la probabilité d'une fonction $f$ fixée se comporte en $\Theta\left(\frac{1}{k^{L(f)+1}}\right)$ quand $k$ tend vers $+\infty$. Cependant, grâce à l'étude des tautologies et des fonctions littéral (cf. Figure 2.10), nous savons que les quatre distributions étudiées dans ce chapitre ne sont pas toutes égales. Pour montrer le résultat principal de ce chapitre, nous avons montré que, quel que soit le modèle étudié, un arbre typique calculant une fonction booléenne $f$ fixée est une expansion d'un arbre minimal de $f$.

| | Arbres de Catalan | Arbres associatifs (non binaires) | Arbres commutatifs (non plans) | Arbres commutatifs et associatifs |
|---|---|---|---|---|
| `Vrai` | $\frac{3}{4} = 0.75$ | $51 - 36\sqrt{2} \approx 0.0883$ | $\frac{385}{512} \approx 0.7520$ | $\frac{(2\ln 2-1)^2}{4} \approx 0.0373$ |
| $x$ | $\frac{5}{16} = 0.3125$ | $546 - 386\sqrt{2} \approx 0.1136$ | $\frac{641}{2048} \approx 0.3130$ | $\frac{(2\ln 2-1)^2(2\ln 2+1)}{4} \approx 0.0890$ |

FIGURE 2.10 – *Les différentes constantes $\lambda$ telles que $\mu_k(\texttt{Vrai}) \sim \frac{\lambda}{k}$ et $\mu_k(x) \sim \frac{\lambda}{k^2}$ asymptotiquement quand $k$ tend vers $+\infty$, selon le modèle d'arbre considéré.*

Les preuves des résultats de ce chapitre sont fondées sur la théorie des motifs de Kozik [Koz08] établie pour les arbres de Catalan. Nous avons montré comment cette théorie pouvait être généralisée aux cas associatif et commutatif et comment elle nous permettait de prouver les résultats principaux de ce chapitre.

Maintenant que nous avons traité le cas d'une fonction booléenne $f$ quelconque, il est intéressant de noter que le cas des fonctions littéral n'est qu'un cas particulier de la Section 2.5 : en effet, les simple-$x$ définis en Section 2.4 sont exactement les expansions de l'arbre minimal de la fonction littéral $x$, i.e. l'arbre de taille 1 dont l'unique feuille est étiquetée par $x$.

Enfin, notons que si notre résultat est vrai pour une fonction $f$ dont le nombre de variables essentielles $E(f)$ ne dépend pas de $k$. Rappelons que l'effet Shannon (cf. Théorème 1.2.9) est exhibé lorsque, asymptotiquement quand $k$ tend vers $+\infty$, presque toute fonction booléenne selon la distribution de probabilité étudiée est de complexité exponentielle en $k$. Si l'on veut infirmer l'effet Shannon, il faut exhiber une fonction $g(k)$ sous-exponentielle en $k$ telle que

$$\mu_k(\{f \in \mathcal{F}_k, L(f) \leqslant g(k)\}).$$

Nos résultats ne permettent pas de conclure quant à l'effet Shannon, car il faudrait alors les appliquer à des fonctions dont la complexité, et donc éventuellement le nombre de variables essentielles dépend de $k$. Il serait cependant d'étudier l'effet Shannon dans ces modèles. Il a été démontré par Genitrini et Gittenberger [GG10] que, dans le modèle de l'implication (binaire, plan), la distribution induite sur l'ensemble des fonctions booléennes n'exhibe pas d'effet Shannon, et il serait donc naturel de conjecturer un comportement similaire dans nos nouveaux modèles, ainsi que dans le modèle des arbres et/ou de Catalan.

Par ailleurs, remarquons que la notion de taille en termes de nombre de feuilles, c'est à dire en termes de nombre de nœuds étiquetés par des littéraux, est pertinente pour pouvoir comparer ces résultats à ceux obtenus dans le cadre binaire plan. Cependant, dans le cas d'arbres non binaires, dans lesquels le nombre de nœuds internes n'est plus égal au nombre de feuilles moins 1, cette notion de taille est peut-être moins naturelle que celle qui prendrait en compte tous les nœuds de l'arbre. C'est cette modification, qui paraît naturelle du point de vue de l'espace mémoire, que nous étudierons dans le chapitre suivant.

# Chapitre 3

# Une nouvelle notion de taille pour les arbres et/ou non binaires

## 3.1 Une notion de taille naturelle

Ce chapitre est une extension naturelle du chapitre précédent : étant donné un arbre et/ou associatif, la taille en termes de nombre total de nœuds et non en nombre de feuilles semble plus naturelle. Si l'on parle par exemple en terme d'espace mémoire nécessaire pour stocker un tel arbre, la notion de *taille* pertinente semble bien être le nombre total de nœuds. Définir une nouvelle notion de taille pour des arbres booléens définit bien une nouvelle notion de complexité d'une fonction booléenne, et donc une toute nouvelle distribution sur $\mathcal{F}_k$, a priori.

Dans ce chapitre, nous considérons la distribution uniforme sur les arbres associatifs plans de taille $n$ (où la taille est désormais le nombre total de nœuds d'un arbre) étiquetés à $k$ variables, et montrons que la suite de distributions induites sur $\mathcal{F}_k$ converge vers une distribution asymptotique $\mu_k$. Les méthodes utilisées dans le Chapitre 2 s'avèrent inapplicables à ce nouveau modèle : la théorie des motifs de Kozik, par exemple, ne s'applique pas. Intuitivement, la théorie des motifs repose sur le fait que *imposer une contrainte sur les étiquettes d'une famille d'arbre rajoute un facteur* $\frac{1}{k}$ *à sa fraction limite* : cette heuristique n'est plus vraie dans ce nouveau modèle. Nous introduirons donc de nouvelles méthodes, elles-aussi à base de combinatoire analytique, pour étudier ce nouveau modèle.

Il est surprenant de constater que, même si les méthodes usuelles sont inefficaces pour ce nouveau modèle, le comportement de la distribution étudiée $\mu_k$ est très similaire au comportement de la distribution usuelle des arbres de Catalan, ou à celles introduites et étudiées dans le Chapitre 2 : nous conjecturons que les fonctions de faible complexité sont favorisées par cette nouvelle distribution et asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple. Si la plupart des preuves s'avèrent plus délicates dans ce nouveau modèle, nous verrons que montrer que cette nouvelle distribution n'exhibe pas d'effet Shannon est relativement immédiat.

Le plan de ce chapitre est le suivant : la Section 3.2 contient la définition du modèle, la preuve de la convergence vers une distribution asymptotique via le théorème de Drmota-Lalley-Woods [Drm97, Lal93, Woo97], et l'énoncé des résultats principaux et de la conjecture concernant ce nouveau modèle. La Section 3.3 est consacrée à l'étude générale du modèle : nous montrons dans cette section des propriétés générales du modèle qui permettent de développer une bonne intuition du comportement des arbres associatifs *typiques* et qui seront utiles dans la suite du Chapitre. La Section 3.4 est consacrée à l'étude de la fonction constante `Vrai`, et donc des arbres tautologiques, la Section 3.5 à l'étude des fonctions littéral, et enfin, la Section 3.6 contient quelques pistes non-

abouties en vue de prouver la conjecture résumant le comportement global de la distribution $\mu_k$ induite sur $\mathcal{F}_k$.s

## 3.2 Modèle, résultats et conjecture

Nous reprenons dans ce chapitre la notion d'arbre booléen associatif du Chapitre 2 (cf. Définition 2.2.1), mais définissons une nouvelle notion de taille d'un arbre, et donc de complexité d'une fonction booléenne :

**Définition 3.2.1**

*La* **taille** $|t|$ *d'un arbre booléen associatif $t$ est le nombre de ses nœuds (nœuds internes et feuilles). On notera $\mathcal{A}_{n,k}$ l'ensemble des arbres booléens de taille $n$, et $A_{n,k}$ son cardinal.*

La notion de complexité est donc modifiée pour cette notion de taille. Par exemple, la fonction XOR est de complexité 7 dans ce modèle. Nous posons $L(\texttt{Vrai}) = L(\texttt{Faux}) = 0$. Par ailleurs, pour des raisons qui seront claires ultérieurement, la complexité des fonctions littéral sera par définition 2. En effet, les deux arbres minimaux de la fonction $x$ seront l'arbre ayant un nœud interne (étiqueté soit par $\wedge$, soit par $\vee$) et une feuille étiquetée par $x$, même si ces arbres ne sont pas, à proprement parler, des arbres associatifs booléens : un arbre associatif booléen ne peut être de taille 2.

De manière usuelle, nous noterons $\mu_{n,k}(f)$ la proportion d'arbres de taille $n$ étiquetés sur $k$ variables calculant $f$, et étudierons le comportement de cette distribution limite quand $n$ tend vers $+\infty$. Voici les résultats principaux de ce chapitre, suivis d'une conjecture plus générale dont la preuve est incomplète dans ce manuscrit.

**Lemme 3.2.2**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$,*

$$\mu_k(f) = \lim_{n \to +\infty} \mu_{n,k}(f)$$

*existe et est strictement positive.*

**Théorème 3.2.3**

*Il existe deux constantes $\alpha$ et $\beta$ telles que, pour tout $k \geqslant 1$,*

$$0 < \alpha \leqslant \mu_k(\texttt{Vrai}) = \mu_k(\texttt{Faux}) \leqslant \beta < \frac{1}{2}.$$

*De plus, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple.*

**Théorème 3.2.4**

*La probabilité d'une fonction littéral vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

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

**Conjecture 3.2.5**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$ telle que $L(f) \geqslant 3$, et telle que $E(f)$ ne dépend pas de $k$, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(f) = \Theta\left(\frac{1}{k^{L(f)}}\right).$$

Ce modèle s'avère donc assez similaire au modèle associatif étudié dans le chapitre précédent : la nouvelle notion de taille ne change pas le comportement global de la distribution induite sur l'ensemble des fonctions booléennes. Nous remarquons tout de même que l'exposant de $\frac{1}{k}$ dans la conjecture est $L(f)$ au lieu de $L(f)+1$ dans le modèle classique, et les deux fonction constantes, `Vrai` et `Faux` ont une probabilité en $\Theta(1)$ quand $k$ tend vers $+\infty$. Nous verrons par la suite que, bien que ces résultats soient très similaires au cas classique, les preuves développées dans le Chapitre 2 ne seront pas applicables dans ce nouveau modèle, pour des raisons que nous développeront notamment dans la Section 3.3.

Montrons tout d'abord que la distribution asymptotique des $\mu_{n,k}$ existe quand la taille des arbres considérés $n$ tend vers $+\infty$, autrement dit, montrons le Lemme 3.2.2. Comme dans les autres modèles d'arbres booléens, cette démonstration se fait via les fonctions génératrices et le théorème de Drmota-Lalley-Woods.

**Démonstration du Lemme 3.2.2 :** Les arbres associatifs dont la taille est mesurée en nombre total de nœuds sont décrits par la spécification suivante (rappelons que les arbres associatifs sont stratifiés) :

$$\hat{\mathcal{A}} = \mathcal{L} + \{\wedge\} \times \mathtt{Seq}_{\geqslant 2}(\check{\mathcal{A}}),$$

où $\hat{\mathcal{A}}$ (resp. $\check{\mathcal{A}}$) est la famille des arbres associatifs de taille 1 ou enracinés par un $\wedge$ (resp. $\vee$), où $\mathcal{L} = \{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ a pour série génératrice $2kz$, et où $\{\wedge\}$ a pour série génératrice $z$ (car chaque nœud interne contribue pour 1 dans la taille). Dès lors, la méthode symbolique nous permet d'induire

$$\hat{A}(z) = 2kz + z \cdot \frac{\check{A}(z)^2}{1-\check{A}(z)},$$

où $\hat{A}(z)$ (resp. $\check{A}(z)$) est la série génératrice de $\hat{\mathcal{A}}$ (resp. $\check{\mathcal{A}}$). Comme, par symétrie du modèle, $\check{A}(z) = \hat{A}(z)$, nous obtenons, après calcul,

$$\hat{A}(z) = \frac{2kz + 1 - \sqrt{(4k^2-8k)z^2 - 4kz + 1}}{2(z+1)}. \tag{3.1}$$

Enfin, si l'on note $A(z) = \sum_{n\geqslant 0} A_{n,k} z^n$ la fonction génératrice des arbres associatifs (dont la taille est mesurée en nombre total de nœuds), alors

$$A(z) = 2\hat{A}(z) - 2kz.$$

Pour toute fonction booléenne $f \in \mathcal{F}_k$, nous noterons $\hat{A}_f(z)$ (resp. $\check{A}_f(z)$) la série génératrice des arbres enracinés par $\wedge$ (resp. $\vee$) et calculant $f$. Via la méthode symbolique, nous avons

$$\hat{A}_f(z) = z\mathbb{1}_{f\text{lit}} + \sum_{\ell\geqslant 2} \sum_{g_1 \wedge \dots \wedge g_\ell = f} \check{A}_{g_1} \dots \check{A}_{g_\ell}$$
$$\check{A}_f(z) = z\mathbb{1}_{f\text{lit}} + \sum_{\ell\geqslant 2} \sum_{g_1 \vee \dots \vee g_\ell = f} \hat{A}_{g_1} \dots \hat{A}_{g_\ell},$$

où $\mathbb{1}_{f\text{lit}} = 1$ si $f$ est une fonction littéral et $\mathbb{1}_{f\text{lit}} = 0$ sinon. L'indice $\ell$ de la somme représente le nombre d'enfants de la racine de l'arbre considéré : ce nombre est donc bien supérieur ou égal à 2, par définition. Nous ne détaillons pas les arguments, mais le théorème de Drmota-Lalley-Woods peut être appliqué à ce système. Nous en concluons que les $(A_f(z))_{f\in\mathcal{F}_k}$ ont toutes la même singularité dominante que nous noterons $\rho_k$, que cette singularité est de type racine carrée pour toute fonction $f \in \mathcal{F}_k$. Nous pouvons donc en conclure que, pour toute fonction $f \in \mathcal{F}_k$, il existe une constante strictement positive $c_k(f)$ telle que, asymptotiquement quand $n$ tend vers $+\infty$,

$$[z^n]A_f(z) \sim c_k(f) n^{-3/2} \rho_k^{-n}.$$

Nous en déduisons

$$[z^n]A(z) \sim c n^{-3/2} \rho_k^{-n},$$

où $c = \sum_{f \in \mathcal{F}_k} c_k(f) > 0$. Dès lors, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mu_{n,k}(f) = \frac{[z^n]A_f(z)}{[z^n]A(z)} \sim \frac{c_k(f)}{c} > 0,$$

ce qui conclut la preuve. ■

Au détour de la preuve du Lemme 3.2.2 (cf. Équation (3.1)), nous avons montré la proposition suivante, qui sera fort utile par la suite. Nous avons en effet vu dans le Chapitre 2 que la connaissance du comportement de la singularité dominante quand $k$ tend vers $+\infty$, ainsi que celle de la série génératrice elle-même évaluée en sa singularité, sont utiles dès que l'on applique le Lemme 1.5.1.

**Proposition 3.2.6**

*La singularité $\rho_k$ de $A(z)$ vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\rho_k = \cdot\frac{1}{2(k+\sqrt{2k})} = \frac{1}{2k} - \frac{1}{k\sqrt{2k}} + \mathcal{O}\left(\frac{1}{k^2}\right),$$
$$A(\rho_k) = 1 - \frac{1}{k} + \mathcal{O}\left(\frac{1}{k\sqrt{k}}\right),$$
$$\hat{A}(\rho_k) = 1 - \frac{1}{\sqrt{2k}} + \mathcal{O}\left(\frac{1}{k}\right).$$

*De plus, si l'on note $\hat{B}(z)$ la série génératrice des arbres associatifs enracinés par $\wedge$, et de taille au moins 3, alors $\hat{B}(z) = \hat{A}(z) - 2kz$, et*

$$\hat{B}(\rho_k) = \frac{1}{\sqrt{2k}} + \mathcal{O}\left(\frac{1}{k}\right).$$

## 3.3 Propriétés générales du modèle

L'objectif de cette Section est de mieux comprendre, intuitivement, notre nouveau modèle d'arbres. Par exemple, les arbres n'ayant aucune feuille de première génération sont *peu probables*, ce qui n'était pas le cas dans le modèle usuel étudié dans le Chapitre 2, et la famille des arbres ayant au moins une feuille de première génération étiquetée par $x_1$ a une proportion limite d'ordre $\frac{1}{\sqrt{k}}$, et non d'ordre $\frac{1}{k}$ comme dans le modèle classique. Ces différences, a priori contre-intuitives, viennent du fait qu'il existe maintenant deux atomes de taille 1 : les nœuds internes qui choisissent entre deux étiquettes différentes et les feuilles qui choisissent entre $2k$ étiquettes différentes. Par ailleurs, dans le modèle classique, un arbre booléen aléatoire de taille $n$ étiqueté sur $k$ variables est un arbre non-étiqueté de taille $n$ que l'on a étiqueté uniformément au hasard : **cette décomposition forme/étiquetage n'est plus vraie dans notre nouveau modèle**. Une autre différence qui sera fondamentale est que, dans ce nouveau modèle, *un arbre typique contient beaucoup de répétitions*. Ce comportement augure des difficultés pour adapter telles quelles les méthodes utilisées dans les modèles usuels, notamment la théorie des motifs de Kozik, qui reposent entièrement sur la *faible probabilité des répétitions*. Ce sont ces raisons qui font de ce nouveau modèle de taille un problème particulièrement intéressant.

### 3.3.1 Quelques propriétés

Dans cette section, nous énumérons donc des propriétés surprenantes de ce modèle qui, en sus de développer notre intuition, seront toutes utiles par la suite.

**Proposition 3.3.1**

*La fraction limite de la famille $\mathcal{A}^{(0)}$ des arbres associatifs booléens n'ayant aucune feuille de première génération vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\mathcal{A}^{(0)}) = \frac{1}{k\sqrt{2k}} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

**Démonstration :** Via la méthode symbolique, nous pouvons établir que la série génératrice de la famille $\mathcal{A}^{(0)}$ est donnée par (cf. Proposition 3.2.6 pour la définition de $\hat{B}(z)$)

$$A^{(0)}(z) = 2z\frac{\hat{B}(z)^2}{1-\hat{B}(z)}.$$

Dès lors, via le Lemme 1.5.1,

$$\begin{aligned}\mu_k(\mathcal{A}^{(0)}) = \lim_{n\to+\infty} \frac{[z^n]A^{(0)}(z)}{[z^n]A(z)} &= \lim_{z\to\rho_k} \frac{A'^{(0)}(z)}{A'(z)} \\ =&2\rho_k\frac{2\hat{B}(\rho_k)}{1-\hat{B}(\rho_k)} \lim_{z\to\rho_k} \frac{\hat{B}'(z)}{A'(z)} + 2\rho_k\frac{\hat{B}(\rho_k)^2}{(1-\hat{B}(\rho_k))^2} \lim_{z\to\rho_k} \frac{\hat{B}'(z)}{A'(z)} \\ &+ 2\frac{\hat{B}(\rho_k)^2}{1-\hat{B}(\rho_k)} \lim_{z\to\rho_k} \frac{1}{A'(z)}.\end{aligned}$$

Nous avons que $\rho_k$ est une singularité de type racine carrée pour $A(z)$. Dès lors, $\lim_{z\to\rho_k} A'(z) = +\infty$ et le troisième terme de la somme ci-dessus vaut donc 0. Par ailleurs, $\lim_{z\to\rho_k} {}^{\hat{B}(z)}\!/_{A'(z)} = {}^1\!/_2$, et au vu de la Proposition 3.2.6, nous obtenons

$$\mu_k(\mathcal{A}^{(0)}) = \frac{1}{k\sqrt{2k}} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

■

**Proposition 3.3.2**

*Soit $\Gamma$ un sous-ensemble de $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$ dont le cardinal, noté $\gamma$, ne dépend pas de $k$. La fraction limite (définie en Section 1.5) de la famille $\mathcal{A}_\Gamma$ des arbres associatifs booléens ayant au moins une feuille de première génération étiquetée par un littéral de $\Gamma$ vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\mathcal{A}_\Gamma) = \frac{\gamma\sqrt{2}}{\sqrt{k}} + \mathcal{O}\left(\frac{1}{k}\right).$$

**Démonstration :** Soit $A_\Gamma(z)$ la série génératrice de la famille $\mathcal{A}_\Gamma$. Via la méthode symbolique, la racine apporte une contribution en $2z$ (étiquette $\wedge$ ou $\vee$) ; la suite des sous-arbres peut être ensuite décomposée comme suit : (1) une première sous-suite d'arbres qui sont, soit de taille au moins 3, soit non-étiquetés par un littéral de $\Gamma$, (2) un premier arbre de taille 1 étiqueté par un littéral de $\Gamma$, (3) puis une suite d'arbres quelconques. De plus, l'une au moins des deux sous-suites évoquées dans les points (1) et (3) précédents est nn vide, ce qui ajoute une correction $-2\gamma z^2$. Dès lors,

$$A_\Gamma(z) = 2z\frac{1}{1-(\hat{A}(z)-\gamma z)}\gamma z\frac{1}{1-\hat{A}(z)} - 2\gamma z^2.$$

Dès lors, via le Lemme 1.5.1,

$$\mu_k(\mathcal{A}_\Gamma) = \lim_{n\to\infty} \frac{[z^n]A_\Gamma(z)}{[z^n]A(z)} = \lim_{z\to\rho_k} \frac{A'_\Gamma(z)}{A'(z)} = \gamma\sqrt{\frac{2}{k}} + \mathcal{O}\left(\frac{1}{k}\right),$$

asymptotiquement quand $k$ tend vers $+\infty$. ■

### 3.3.2 Une famille d'arbres utile

Cette partie se concentre sur une sous-famille d'arbres associatifs. Il est assez difficile d'expliquer en quoi cette famille sera cruciale plus tard, et la définition de cette famille peut sembler assez peu naturelle. Cette définition provient de l'étude des tautologies, et notamment de la preuve que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple (cf. Section 3.4). Comme l'étude de cette sous-famille est trop longue pour être incluse dans l'étude des tautologies, elle est développée ici, à part, en tant que parenthèse technique.

**Définition 3.3.3 (cf. Figure 3.1)**

*Soient $q, r, \ell, p$ quatre entiers, soit $\{\gamma_1, \dots, \gamma_p\}$ un sous-ensemble de $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ ne pouvant contenir à la fois une variable et sa négation. Soit $\mathcal{M}^p_{q,\ell,r}$ les famille des arbres associatifs enracinés par un $\vee$, et tels que,*

- *exactement $q$ littéraux différents $\alpha_1, \dots, \alpha_q$ sont étiquettes des feuilles de première génération, et parmi ces étiquettes aucune variable ne peut apparaître à la fois positivement et négativement ;*
- *exactement $\ell$ enfants de la racine sont des nœuds internes ;*
- *au moins l'un des sous-arbres enracinés en un des ces enfants de la racine appartient à la famille $\mathcal{J}^p_{q,r}$ définie ci-dessous.*

*La famille $\mathcal{J}^p_{q,r}$ contient tous les arbres enracinés par un $\wedge$ et tels qu'il existe un ensemble $\beta_1, \dots, \beta_r$ de $r$ littéraux deux à deux distincts (et différents des $\alpha_1, \dots, \alpha_q, \gamma_1, \dots, \gamma_p$ et de leurs négations) vérifiant :*

- *les feuilles de première génération sont étiquetées par des littéraux de $\{\alpha_1, \dots, \alpha_q, \gamma_1, \dots, \gamma_p, \beta_1, \dots, \beta_r\}$ ou par leurs négations ;*
- *les étiquettes $\beta_1, \dots, \beta_r$ apparaissent toutes parmi les feuilles de première génération.*

**Lemme 3.3.4**

*Si, asymptotiquement quand $k$ tend vers $+\infty$, $q = \Omega(k^{1/4})$, $\ell = \mathcal{O}(k^{1/8})$ et $r \leqslant \ell$, alors, la fraction limite de la famille $\mathcal{M}_{q,\ell,r}$ vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\mathcal{M}_{q,\ell,r}) = \mathcal{O}\left(\frac{1}{k^{3/2}}\right).$$

**Démonstration :** Via des arguments symboliques, la fonction génératrice de $\mathcal{J}^p_{q,r}$ est donnée par

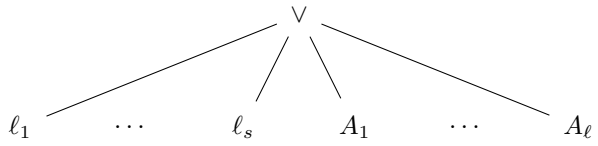
$$J(z) = \frac{z^{r+1}}{(1 - (\hat{B}(z) + (q+p)z)) \dots (1 - (\hat{B}(z) + (q+p+r)z))},$$

où $\hat{B}(z)$ est la série génératrice des arbres de taille au moins 3 (cf. Proposition 3.2.6). La fonction génératrice $M(z)$ de $\mathcal{M}^p_{q,\ell,r}$ est donc majorée par

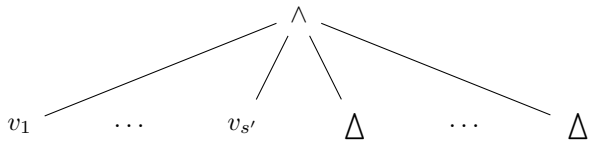
$$M(z) \leqslant \Delta(z) := z\binom{k}{q+r} 2^{q+r} q! \frac{1}{\ell!} \left( z^{q+\ell} \prod_{m=1}^{q} \frac{1}{1-mz} \right)^{(\ell)} J(z)\hat{B}(z)^{\ell-1}.$$

FIGURE 3.1

Un arbre de $\mathcal{M}^p_{q,\ell,r}$ est un arbre de la forme ci-dessous, à permutation des sous-arbres près, où les étiquettes $\ell_1, \dots, \ell_s$ des feuilles de première génération forment un ensemble $\{\alpha_1, \dots, \alpha_q\}$ de $q$ étiquettes deux à deux distinctes telles qu'une variable et sa négation ne peuvent faire partie simultanément de cet ensemble, où les arbres $A_1, \dots, A_\ell$ sont de taille au moins 3, tels qu'il existe $i \in \{1, \dots, \ell\}$ tel que $A_i \in \mathcal{J}^p_{q,r}$.




Un arbre de $\mathcal{J}^p_{q,r}$ est un arbre de la forme suivante, à permutation des sous-arbres près, où l'ensemble des étiquettes $v_1, \dots, v_{s'}$ privé des littéraux $\alpha_1, \dots, \alpha_q, \gamma_1, \dots, \gamma_p$ et de leurs négations est de cardinal $r$ et ne contient pas simultanément une variable et sa négation, et où les arbres $\bigtriangleup$ sont des arbres de taille au moins 3.




En effet, le facteur $z$ représente la racine ; le facteur $\binom{k}{q+r} 2^{q+r} q!$ correspond au choix des $\alpha_1, \dots, \alpha_k$, $\beta_1, \dots, \beta_r$ tels qu'une variable et sa négation ne peuvent être toutes les deux choisies ; le facteur $\frac{1}{\ell!}\left(z^{q+\ell} \prod_{m=1}^{q} \frac{1}{1-mz}\right)^{(\ell)}$ correspond au choix des emplacements des $\ell$ enfants de la racine qui sont des nœuds internes[1] ; $J(z)$ représente le sous-arbre de $\mathcal{J}^p_{q,\ell,r}$ et $\hat{B}(z)^{\ell-1}$ représente les $\ell - 1$ autres sous-arbres. Il y a du double-comptage effectué dans cette dernière remarque puisque plusieurs sous-arbres peuvent être des éléments de $\mathcal{J}^p_{q,\ell,r}$, et nous comptons ici comme s'il y en avait un seul. C'est pour cela que nous obtenons uniquement une majoration de $M(z)$.

$$\Delta(z) = z\binom{k}{q+r} 2^{q+r} q! \frac{1}{\ell!} \left( z^{q+\ell} \prod_{m=1}^{q} \frac{1}{1-mz} \right)^{(\ell)} \frac{z^{r+1}}{\prod_{m=0}^{r}[1 - (\hat{B}(z) + (q+p+m)z)]} \hat{B}(z)^{\ell-1}. \quad (3.2)$$

Soit

$$K(z) = \frac{\hat{B}(z)^{\ell-1}}{\prod_{m=0}^{r}(1 - (\hat{B}(z) + (q+p+m)z))}.$$

Remarquons au passage que la singularité de $\Delta(z)$ est $\rho_k$. En effet, les dénominateurs $(1 - (\hat{B}(z) + (q+p+m)z))_{m\in\{0,\dots,r\}}$ ne s'annulent pas avant $\rho_k$ (car $q = o(\sqrt{k})$) et $\rho_k$ est la singularité de $\hat{B}(z)$. De plus, les dénominateurs $(1 - mz)_{m\in\{1,\dots,q\}}$ ne s'annulent pas avant $\rho_k$ car $\rho_k \sim \frac{1}{2k}$, et $q = o(\sqrt{k})$.

1. Rappelons que $[z^n](zf(z))^{(\ell)} = (n+\ell)\dots(n+1)[z^n]f(z)$, où l'exposant $(\ell)$ représente la dérivée $\ell^{\text{ième}}$.

Au vu du Lemme 1.5.1, la fraction limite de la famille $\mathcal{M}^p_{q,\ell,r}$ est donnée par

$$\lim_{z\to\rho_k}\frac{M'(z)}{A'(z)}=\binom{k}{q+r}2^{q+r}k!\frac{1}{\ell!}\left(\rho_k^{q+\ell}\prod_{m=1}^{q}\frac{1}{1-m\rho_k}\right)^{(\ell)}\rho_k^2\rho_k^{r+2}\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}.$$

Afin d'évaluer cette fraction limite quand $k$ tend vers $+\infty$, nous devons remarquer que, si $f(z)=z^{q+\ell}\prod_{m=1}^{q}\frac{1}{1-mz}$, alors, d'après le théorème de Cauchy,

$$\frac{f^{(\ell)}(z)}{\ell!}=\frac{1}{2i\pi}\oint_{\gamma}\frac{f(z)}{(z-\rho_k)^{\ell+1}}dz,$$

pour tout contour $\gamma$ autour de $\rho_k$. Soit $\gamma$ le cercle défini par l'équation $|z-\rho_k|=\frac{1}{k^{3/2}}$. Dès lors,

$$\begin{aligned}|f^{(\ell)}(z)|&\leqslant\frac{1}{2\pi}\ell!\oint_{\gamma}\frac{|f(z)|}{|z-\rho_k|^{\ell+1}}dz\\&\leqslant\ell!f\left(\rho_k\left(1+\frac{2}{k}\right)\right)k^{3/2(\ell+1)}\frac{1}{k^{3/2}}\end{aligned}$$

car le maximum de $|f(z)|$ sur $\gamma$ est atteint en $z=\rho_k+\frac{1}{k^{3/2}}=\rho_k\left(1+\frac{2}{\sqrt{k}}\right)$ puisque $f(z)$ est une fonction génératrice à coefficients positifs. Nous en déduisons que, asymptotiquement quand $z$ tend vers $\rho_k$ ($\sim\frac{1}{2k}$ quand $k$ tend vers $+\infty$),

$$\begin{aligned}|f^{(\ell)}(z)|&\leqslant\ell!k^{3/2\ell}\left(\frac{1}{2k}\right)^{q+\ell}\left(1+\frac{2}{\sqrt{k}}\right)^{q+\ell}\prod_{m=1}^{q}\frac{1}{1-\frac{m}{2k}}\\&\leqslant\ell!k^{\ell/2-q}2^{-q-\ell}\left(1+\frac{2}{\sqrt{k}}\right)^{q+\ell}\left(\frac{1}{1-\frac{q}{2k}}\right)^{q}.\end{aligned}$$

Comme $q\leqslant k$, nous avons, asymptotiquement quand $k$ tend vers $+\infty$,

$$|f^{(\ell)}(z)|\leqslant cst\cdot\ell!2^{-q-\ell}k^{\ell/2-q}.$$

Au vu de l'Équation (3.2), asymptotiquement quand $k$ tend vers $+\infty$ [2],

$$\begin{aligned}\lim_{z\to\rho_k}\frac{M'(z)}{A'(z)}&\sim cst\cdot2^{-\ell-q}k^{-q+\ell/2}\binom{k}{q+r}2^{q+r}q!\left(\frac{1}{2k}\right)^{r+2}\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}\\&\leqslant cst\cdot k^{\ell/2}2^{r-\ell}\frac{(k-q)_r}{(q+r)_r}\frac{(k)_q}{k^q}\left(\frac{1}{2k}\right)^{r+2}\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}\\&\leqslant cst\cdot k^{\ell/2}2^{-\ell}\frac{k^r}{q^r}\frac{\prod_{m=q}^{q+r-1}\left(1-\frac{m}{k}\right)}{\prod_{m=1}^{r}\left(1+\frac{m}{q}\right)}\prod_{m=1}^{q-1}\left(1-\frac{m}{k}\right)\left(\frac{1}{k}\right)^{r}\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}\\&\leqslant cst\cdot2^{-\ell}k^{\ell/2-2}q^{-r}\left(1+\frac{r}{q}\right)^{r}\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}\\&\leqslant cst\cdot2^{-\ell}k^{\ell/2-2}q^{-r}\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}\end{aligned}$$

car $\left(1+\frac{r}{q}\right)^r=\left(1+\mathcal{O}\left(\frac{1}{k^{1/8}}\right)\right)^r=\mathcal{O}(1)$. De plus,

$$\lim_{z\to\rho_k}\frac{K'(z)}{A'(z)}=\frac{\hat{B}'(z)}{A'(z)}\frac{1}{\prod_{m=0}^{r}(1-\hat{B}(\rho_k)-(q+p+m)\rho_k)}\left[(\ell-1)\hat{B}(\rho_k)^{\ell-2}+\sum_{m=0}^{r}\frac{\hat{B}(\rho_k)^{\ell-1}}{1-\hat{B}(\rho_k)-(q+p+m)\rho_k}\right],$$

2. Rappelons que pour tout réel $x$ et pour tout entier $m$, $(x)_m=x(x-1)\dots(x-(m-1))$.

et comme $\lim_{z\to\rho_k} \frac{\hat{B}'(z)}{A'(z)} = 1/2$ et $\hat{B}(\rho_k) \sim \frac{1}{\sqrt{2k}}$ (cf. Proposition 3.2.6), nous obtenons

$$\lim_{z\to\rho_k} \frac{K'(z)}{A'(z)} = \mathcal{O}\left(r2^r\ell\left(\frac{1}{2k}\right)^{\frac{\ell-1}{2}}\right).$$

Nous avons donc finalement, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k(\mathcal{M}^p_{q,\ell,r}) = \mathcal{O}\left(r\ell 2^{r-\ell}2^{-\frac{\ell-1}{2}}k^{-3/2}\right) = o\left(\frac{1}{k^{3/2}}\right).$$

■

**Lemme 3.3.5**

*La fraction limite des arbres associatifs ayant moins de $k^{1/4}$ étiquettes différentes parmi les feuilles de première génération est d'ordre $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$, asymptotiquement quand $k$ tend vers $+\infty$.*

**Démonstration :** La fonction génératrice des arbres ayant exactement $q$ étiquettes différentes parmi les feuilles de première génération (et tels qu'une variable et sa négation n'apparaissent pas toutes deux dans la première génération) est donnée par

$$G_q(z) = \binom{k}{q} 2^q q! z \prod_{m=0}^{q} \frac{z}{1-mz-\hat{B}(z)},$$

et leur fraction limite est donc donnée par

$$\lim_{z\to\rho_k} \frac{G'_q(z)}{A'(z)} = \binom{k}{q} 2^q q! z \prod_{m=0}^{q} \frac{z}{1-m\rho-\hat{B}(\rho_k)} \sum_{m=0}^{q} \frac{1}{1-m\rho_k-\hat{B}(\rho_k)} \lim_{z\to\rho_k} \frac{\hat{B}'(z)}{A'(z)}.$$

Comme $\lim_{z\to\rho_k} \frac{\hat{B}'(z)}{A'(z)} = \frac{1}{2}$, nous avons, asymptotiquement quand $k$ tend vers $+\infty$,

$$\begin{aligned}\lim_{z\to\rho_k} \frac{G'_q(z)}{A'(z)} &\leqslant \frac{1}{4k} \prod_{m=0}^{q-1}\left(1-\frac{m}{k}\right) \prod_{m=0}^{q} \frac{1}{1-\frac{m}{2k}-\frac{1}{\sqrt{2k}}} \left(\sum_{m=0}^{q} \frac{1}{1-\frac{m}{2k}-\frac{1}{\sqrt{2k}}}\right) \\ &\leqslant \frac{1}{4k}\left(\frac{1}{1-\frac{q}{2k}-\frac{1}{\sqrt{2k}}}\right)^{q+1} (q+1)\frac{1}{1-\frac{q}{2k}-\frac{1}{\sqrt{2k}}},\end{aligned}$$

car $q \leqslant k^{1/4}$ Dès lors, la fraction limite des arbres ayant moins de $k^{1/4}$ étiquettes différentes parmi les feuilles de première génération vérifie

$$\begin{aligned}\sum_{q=0}^{k^{1/4}} \lim_{z\to\rho_k} \frac{G'_q(z)}{A'(z)} &\leqslant \sum_{q=0}^{k^{1/4}} \frac{1}{4k}\left(\frac{1}{1-\frac{q}{2k}-\frac{1}{\sqrt{2k}}}\right)^{q+1} (q+1)\frac{1}{1-\frac{q}{2k}-\frac{1}{\sqrt{2k}}} \\ &\leqslant k^{1/4}\frac{1}{4k}\left(\frac{1}{1-\frac{k^{1/4}}{2k}-\frac{1}{\sqrt{2k}}}\right)^{k^{1/4}+1} (k^{1/4}+1)\frac{1}{1-\frac{k^{1/4}}{2k}-\frac{1}{\sqrt{2k}}} \\ &= \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).\end{aligned}$$

■

**Lemme 3.3.6**

*La fraction limite des arbres ayant au moins $k^{1/8}$ sous-arbres de taille au moins 3 est d'ordre $\Theta\left(\frac{k^{1/8}}{2^{k^{1/8}}}\right)$, asymptotiquement quand $k$ tend vers $+\infty$.*

**Démonstration :** La fonction génératrice des arbres ayant exactement $\ell$ sous-arbres de taille au moins 3 est donnée par

$$H_\ell(z) = 2z\frac{\hat{B}^\ell(z)}{(1-2kz)^{\ell+1}}.$$

Dès lors, la fraction limite de cette famille vérifie, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{z\to\rho_k}\frac{H'_\ell(z)}{A'(z)} = \frac{1}{k}\frac{\ell\hat{B}^{\ell-1}(\rho_k)}{(1-2k\rho_k)^{\ell+1}}\lim_{z\to\rho_k}\frac{\hat{B}'(z)}{A'(z)} \sim \frac{1}{2k}\frac{\ell\left(\frac{1}{\sqrt{2k}}\right)^{\ell-1}}{\left(\sqrt{\frac{2}{k}}\right)^{\ell+1}}$$

$$\sim \frac{1}{2k}\frac{\ell\left(\frac{1}{\sqrt{2k}}\right)^{\ell-1}}{\left(\sqrt{\frac{2}{k}}\right)^{\ell+1}} \sim \frac{\ell}{2^{\ell+1}}.$$

Dès lors, la fraction limite des arbres ayant au moins $k^{1/8}$ enfants de la racine qui sont des nœuds internes vérifie, asymptotiquement quand $k$ tend vers $+\infty$,

$$\sum_{\ell\geqslant k^{1/8}}\frac{\ell}{2^{\ell+1}} = \Theta\left(\frac{k^{1/8}}{2^{k^{1/8}}}\right).$$

■

## 3.4 Tautologies.

L'objet de cette partie est de démontrer le Théorème 3.2.3. Comme dans tout modèle d'arbres booléens, étudier la fonction `Vrai` est certes l'étude d'un cas particulier simple, mais aussi la première étape du cas général, i.e. de l'étude de la probabilité de n'importe quelle fonction booléenne. Cette Section est divisée en trois sous-sections : la première permet de montrer l'existence de la borne inférieure $\alpha$ du Théorème 3.2.3, la seconde montrera l'existence de la borne supérieure $\beta$, et la troisième montrera que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est une tautologie simple (cf. Définition 1.3.6).

### 3.4.1 Une famille non-négligeable de tautologies.

Dans cette partie nous définissons une famille de tautologies dont la fraction limite est bornée inférieurement, et uniformément pour tout $k \geqslant 1$. Rappelons la définition d'une tautologie simple dans le cas d'arbres associatifs (cf. Définition 1.3.6) :

**Définition 3.4.1**

*Une* **tautologie simple** *réalisée par la variable $x$ est un arbre enraciné par $\vee$, et tel que $x$ et $\bar{x}$ apparaissent comme étiquettes de deux feuilles de première génération. On notera $\mathcal{ST}$ la famille des tautologies simples.*

Soit $\mathcal{E}^q$ la famille des arbres enracinés par $\vee$ ayant exactement $q$ feuilles de première génération et exactement un enfant de la racine qui ne soit pas une feuille (la racine est donc d'arité $q+1$). On prendra $q \in \{\lfloor\sqrt{k}\rfloor, \ldots, 15\lfloor\sqrt{k}\rfloor\}$ (par exemple). Dans la suite, pour plus de lisibilité, nous omettrons les symboles $\lfloor\rfloor$. Pour tout entier $q \geqslant 1$, la série génératrice de $\mathcal{E}_q$ est donnée par

$$E_q(z) = z\cdot(2kz)^q\cdot(q+1)\hat{B}(z),$$

où, rappelons-le, $\hat{B}(z) = \hat{A}(z) - 2kz$.

Pour tout $q \geqslant 1$, nous allons maintenant partitionner $\mathcal{E}_q$ en deux sous-familles. La famille $\mathcal{E}_{q,1}$ contient tous les arbres de $\mathcal{E}_q$ tels que :

(a) L'ensemble $\mathcal{L}$ des étiquettes des $\sqrt{k}$ premières feuilles de première génération est de cardinal au moins $\sqrt{k}/2$.

(b) Les $q - \sqrt{k}$ feuilles de première génération restantes sont étiquetées par des littéraux dont la négation n'appartient pas à $\mathcal{L}$.

La famille $\mathcal{E}_{q,2}$ contient tous les arbres de $\mathcal{E}^q$ tels que

(a') L'ensemble $\mathcal{L}'$ des étiquettes des $\sqrt{k}$ premières feuilles de première génération est de cardinal au plus $\sqrt{k}/2$.

(b) Les $q - \sqrt{k}$ feuilles de première génération restantes sont étiquetées par des littéraux dont la négation n'appartient pas à $\mathcal{L}'$.

Le lemme suivant vient du fait que tout arbre qui n'est pas une tautologie simple vérifie soit la condition (a), soit la condition (a'), et vérifie toujours la condition (b) :

**Lemme 3.4.2**

*Soit $t$ un arbre qui n'est pas une tautologie simple. Alors, $t \in \cup_{q\geqslant 0}(\mathcal{E}_{q,1} \cup \mathcal{E}_{q,2})$.*

Dès lors, l'ensemble des tautologies simples ayant $q$ feuilles de première génération contient $\mathcal{E}_q \backslash (\mathcal{E}_{q,1} \cup \mathcal{E}_{q,2})$. Minorons la fraction limite de cet ensemble. On introduit la notation suivante : étant données deux séries génératrices $A(z)$ et $B(z)$, on notera $A(z) \prec B(z)$ si, et seulement si, pour tout entier $n \geqslant 0$, $[z^n]A(z) \leqslant [z^n]B(z)$. On note $E_{q,1}(z)$ (resp. $E_{q,2}(z)$) la série génératrice de $\mathcal{E}_{q,1}$ (resp. $\mathcal{E}_{q,2}$). Par la méthode symbolique,

$$E_{q,1}(z) \prec z \cdot (2kz)^{\sqrt{k}} \cdot \left(\left(2k - \frac{\sqrt{k}}{2}\right) z\right)^{q-\sqrt{k}} \cdot (q+1)\hat{B}(z).$$

Nous avons seulement une inégalité car nous n'avons pas restreint les étiquetages des $\sqrt{k}$ feuilles de gauche. De plus, nous interdisons $\frac{\sqrt{k}}{2}$ étiquettes pour les feuilles de droite alors qu'il suffit d'en interdire moins si le cardinal de $\mathcal{L}$ est strictement inférieur à $\frac{\sqrt{k}}{2}$. De même, par des arguments symboliques,

$$E_{q,2}(z) \prec \binom{k}{\sqrt{k}/2} 2^{\sqrt{k}/2} \cdot z \cdot \left(\frac{\sqrt{k}}{2} z\right)^{\sqrt{k}} \cdot (2kz)^{q-\sqrt{k}} \cdot (q+1)\hat{B}(z).$$

Dès lors,

$$\mu_k(\mathcal{E}_q \backslash (\mathcal{E}_{q,1} \cup \mathcal{E}_{q,2})) \geqslant \lim_{z\to\rho_k} \frac{\frac{d}{dz}E^q(z) - \frac{d}{dz}E_1^q(z) - \frac{d}{dz}E_2^q(z)}{A'(z)}.$$

Nous avons

$$\begin{aligned} E^q(z) - E_1^q(z) - E_2^q(z) =& (q+1)z\hat{B}(z) \left[ (2kz)^q - (2kz)^{\sqrt{k}} \left(\left(2k - \frac{\sqrt{k}}{2}\right) z\right)^{q-\sqrt{k}} \right. \\ & \left. - \binom{k}{\sqrt{k}/2} 2^{\sqrt{k}/2} \cdot \left(\frac{\sqrt{k}}{2} z\right)^{\sqrt{k}} \cdot (2kz)^{q-\sqrt{k}} \right], \end{aligned}$$

et donc, comme $\lim_{z\to\rho_k} \frac{\hat{B}'(z)}{A'(z)} = 1/2$,

$$\mu_k(\mathcal{E}^q\backslash(\mathcal{E}_1^q \cup \mathcal{E}_2^q)) \geqslant \frac{q+1}{2}\rho_k\Bigg[(2k\rho_k)^q - (2k\rho_k)^{\sqrt{k}}\left(\left(2k - \frac{\sqrt{k}}{2}\right)\rho_k\right)^{q-\sqrt{k}} - \binom{k}{\sqrt{k}/2}2^{\sqrt{k}/2}\cdot\left(\frac{\sqrt{k}}{2}\rho_k\right)^{\sqrt{k}}\cdot(2k\rho_k)^{q-\sqrt{k}}\Bigg].$$

Regardons les termes de la somme entre crochets un par un : asymptotiquement quand $k$ tend vers $+\infty$,

$$(2k\rho_k)^q \sim \left(1 - \frac{\sqrt{2}}{\sqrt{k}}\right)^q \sim \exp\left(-\frac{q\sqrt{2}}{\sqrt{k}}\right),$$

de plus,

$$\begin{aligned}(2k\rho_k)^{\sqrt{k}}\left(\left(2k - \frac{\sqrt{k}}{2}\right)\rho_k\right)^{q-\sqrt{k}} &\sim \left(1-\sqrt{\frac{2}{k}}\right)\left(\left(1-\frac{1}{4\sqrt{k}}\right)\left(1-\frac{\sqrt{2}}{\sqrt{k}}\right)\right)^{q-\sqrt{k}}\\ &\sim \left(1-\sqrt{\frac{2}{k}}\right)\left(1-\frac{1}{4\sqrt{k}}-\frac{\sqrt{2}}{\sqrt{k}}+\mathcal{O}\left(\frac{1}{k}\right)\right)^{q-\sqrt{k}}\\ &\sim \exp\left(-(1+4\sqrt{2})\frac{q-\sqrt{k}}{4\sqrt{k}}\right).\end{aligned}$$

Enfin, via la formule de Stirling, nous avons

$$\binom{k}{\sqrt{k}/2} \sim (2\mathsf{e})^{\sqrt{k}/2}k^{\frac{\sqrt{k}}{4}-\frac{1}{4}},$$

ce qui implique, après quelques calculs,

$$\binom{k}{\sqrt{k}/2}2^{\sqrt{k}/2}\cdot\left(\frac{\sqrt{k}}{2}\rho_k\right)^{\sqrt{k}}\cdot(2k\rho_k)^{q-\sqrt{k}} = \mathcal{O}\left(\left(\frac{\mathsf{e}}{2\sqrt{k}}\right)^{\sqrt{k}/2}\right).$$

Dès lors,

$$\mu_k(\mathcal{E}_q\backslash(\mathcal{E}_{q,1}\cup\mathcal{E}_{q,2})) \geqslant \frac{(q+1)\rho_k}{2}\mathsf{e}^{-\frac{q\sqrt{2}}{\sqrt{k}}}\left(1-\mathsf{e}^{\frac{q\sqrt{2}}{\sqrt{k}}-(1+4\sqrt{2})\frac{q-\sqrt{k}}{4\sqrt{k}}}\right).$$

Posons $\gamma = \frac{q}{\sqrt{k}}$, alors,

$$\mu_k(\mathcal{E}^q\backslash(\mathcal{E}_1^q\cup\mathcal{E}_2^q)) \geqslant \frac{(q+1)\rho_k}{2}\mathsf{e}^{-\gamma\sqrt{2}}\left(1-\mathsf{e}^{-\gamma/4+\frac{1+4\sqrt{2}}{4}}\right),$$

Ce qui implique, comme l'union considérée est disjointe,

$$\mu_k\left(\bigcup_{q=2\sqrt{k}}^{15\sqrt{k}}(\mathcal{E}^q\backslash(\mathcal{E}_1^q\cup\mathcal{E}_2^q))\right) \geqslant \sum_{q=2\sqrt{k}}^{15\sqrt{k}}\mu_k(\mathcal{E}^q\backslash(\mathcal{E}_1^q\cup\mathcal{E}_2^q)) \geqslant \sum_{q=2\sqrt{k}}^{15\sqrt{k}}\frac{(q+1)\rho_k}{2}c_q,$$

où $c_q = \mathsf{e}^{-\gamma\sqrt{2}}\left(1-\mathsf{e}^{-\gamma/4+\frac{1+4\sqrt{2}}{4}}\right) > 0$. Ainsi,

$$\mu_k\left(\bigcup_{q=2\sqrt{k}}^{15\sqrt{k}}(\mathcal{E}^q\backslash(\mathcal{E}_1^q\cup\mathcal{E}_2^q))\right) \geqslant 14\sqrt{k}\frac{(2\sqrt{k}+1)\rho_k}{2}\min_{q\in\{2\sqrt{k},\ldots,15\sqrt{k}\}}c_q \geqslant cst\cdot k\rho_k,$$

où *cst* est une constante strictement positive. Comme $k\rho_k \sim 1/2$, asymptotiquement quand $k$ tend vers $+\infty$, nous obtenons bien que

$$\mu_k\left(\bigcup_{q=2\sqrt{k}}^{15\sqrt{k}} (\mathcal{E}_q \backslash (\mathcal{E}_{q,1} \cup \mathcal{E}_{q,2}))\right) = \Omega(1),$$

asymptotiquement quand $k$ tend vers $+\infty$. Nous avons donc montré le lemme suivant :

**Lemme 3.4.3**

*Il existe une constante $\alpha$ telle que, pour tout $k \geqslant 1$,*

$$\mu_k(\texttt{Vrai}) \geqslant \mu_k(\mathcal{ST}) \geqslant \alpha > 0.$$

### 3.4.2 Une famille non-négligeable de non-constantes.

Prouvons le lemme suivant :

**Lemme 3.4.4**

*Il existe une constante $\beta$ telle que, pour tout $n \geqslant 0$,*

$$\mu_k(\texttt{Vrai}) \leqslant \beta < \frac{1}{2}.$$

Il s'agit de trouver une famille d'arbres de fraction limite d'ordre $\Theta(1)$ quand $k$ tend vers $+\infty$ et telle que les arbres de cette famille ne calculent ni la fonction constante `Vrai`, ni la fonction constante `Faux` :

**Définition 3.4.5**

*Soit $\check{\mathcal{G}}_q$ la famille des arbres enracinés par un $\vee$ et ayant exactement $q$ feuilles de première génération, étiquetées par exactement $q$ étiquettes différentes $\alpha_1, \ldots, \alpha_q$ (telles qu'une variable et sa négation ne peuvent apparaître toutes deux parmi ces étiquettes) et dont les sous-arbres de taille au moins 3 sont tous des contradictions.*

Remarquons qu'un arbre de $\check{\mathcal{G}}_q$ calcule la fonction $\alpha_1 \vee \ldots \vee \alpha_q$, qui n'est ni une tautologie, ni une contradiction. La fonction génératrice de cette famille est donnée par (rappelons que $T(z)$ est la série génératrice des tautologies)

$$\check{G}_q(z) = \binom{k}{q} 2^q q! z^{q+1} \frac{1}{(1-T(z))^{q+1}},$$

et sa fraction limite est donc donnée par

$$\lim_{z\to\rho_k} \frac{\check{G}'_q(z)}{A'(z)} \sim \binom{k}{q} 2^q q! \rho_k^{q+1} \frac{q+1}{(1-T(\rho_k))^{q+1}} \lim_{z\to\rho_k} \frac{T'(z)}{A'(z)}.$$

La sous-section 3.4.1 nous assure que $\lim_{z\to\rho_k} \frac{T'(z)}{A'(z)} \geqslant \alpha$, car $\mu_k(\texttt{Vrai}) = \lim_{z\to\rho_k} \frac{T'(z)}{A'(z)} \geqslant \alpha$. De plus, nous savons que $0 \leqslant T(\rho_k) \leqslant \hat{B}(\rho_k) < 1$ (car une tautologie ne peut être un arbre de taille 1), ce

qui implique $\frac{q+1}{(1-T(\rho_k))^{q+1}} \geqslant 1$. Dès lors,

$$\begin{aligned}
\lim_{z\to\rho_k} \frac{G'(z)}{A'(z)} &\geqslant \binom{k}{q} 2^q q!(q+1)\rho_k^{q+1}\alpha \\
&\geqslant cst \cdot \frac{k(k-1)\dots(k-q+1)}{k^q}(q+1)(2k)^q\rho_k^{q+1} \\
&\geqslant cst \cdot \prod_{j=1}^{q-1}\left(1-\frac{j}{k}\right)(q+1)(2k)^q\rho_k^{q+1} \\
&\geqslant cst \cdot \left(1-\frac{q-1}{k}\right)^{q-1}(q+1)(2k)^q\rho_k^{q+1} \\
&= cst \cdot \left(1-\mathcal{O}\left(\frac{q^2}{k}\right)\right) q\rho_k \left(1-\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)\right).
\end{aligned}$$

Si l'on suppose $q = \Theta(\sqrt{k})$, asymptotiquement quand $k$ tend vers $+\infty$, alors la borne supérieure ci-dessus est d'ordre $\Theta(\frac{1}{\sqrt{k}})$, et la fraction limite de la famille $\mathcal{G}_q$ est plus grande que $\frac{c_q}{\sqrt{k}}$ où $c_q$ est une constante strictement positive. Considérons la famille $\check{\mathcal{G}} = \bigcup_{q=\sqrt{k}}^{2\sqrt{k}} \check{\mathcal{G}}_q$. Sa fraction limite est donc plus grande que la constante positive $c = \min_{q=\sqrt{k}\dots 2\sqrt{k}} c_q$, ce qui conclut donc la preuve du Lemme 3.4.4, en posant $\beta = \frac{1}{2} - c$.

Nous avons aussi le résultat suivant :

**Théorème 3.4.6**

*La distribution induite sur $\mathcal{F}_k$ des arbres associatifs dont la taille est mesurée en terme de nombre total de nœuds n'exhibe pas d'effet Shannon.*

**Démonstration :** Nous avons montré précédemment que

$$\mu_k\left(\bigcup_{q=2\sqrt{k}}^{15\sqrt{k}} \check{\mathcal{G}}_q\right) \geqslant c > 0.$$

Or, la complexité des fonctions calculée par les arbres de $\check{\mathcal{G}}_q$ est $q+1$, et ce quel que soit $q \geqslant 1$. Donc, la probabilité des fonctions de complexité $\sqrt{k} \leqslant L(f) \leqslant 15\sqrt{k}$ vérifie :

$$\mu_k(\{f \mid \sqrt{k} \leqslant L(f) \leqslant 15\sqrt{k}\}) \geqslant c > 0,$$

ce qui implique l'existence d'une famille de fonctions booléennes de complexité sub-exponentielle en $k$ dont la proportion limite est d'ordre $\Theta(1)$ quand $k$ tend vers $+\infty$, et nie donc l'effet Shannon (cf. Théorème 1.2.9). ■

### 3.4.3 Presque toute tautologie est simple

L'objet de cette partie est de montrer que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple. Cela permet de comparer ce nouveau modèle d'arbres à ceux déjà étudiés dans la littérature et dans ce mémoire. Cette propriété, vraie dans *tous* (ou presque, cf. Chapitre 5) les autres modèles d'arbres est aussi vérifiée ici, même si la preuve que nous en donnons est très différente d'une approche via la théorie des motifs, car la théorie des motifs ne semble pas s'appliquer à ce nouveau modèle.

Montrons que la fraction limite des tautologies qui ne sont pas des tautologies simples tend vers 0, asymptotiquement quand $k$ tend vers $+\infty$ :

**Lemme 3.4.7**

*Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\texttt{Vrai}) \sim \mu_k(\mathcal{ST}).$$

**Démonstration Proof of Theorem 3.2.3 :** Soit $\mathcal{N}$ la famille des tautologies qui ne sont pas simples.

Soit $t \in \mathcal{N}$. Supposons tout d'abord que la racine de $t$ est étiquetée par $\wedge$. Alors, $t$ ne peut avoir aucune feuille de première génération. D'après la Proposition 3.3.1, la fraction limite des arbres associatifs n'ayant aucune feuille de première génération est d'ordre $\frac{1}{k^{3/2}}$. Dès lors, le sous-ensemble des arbres de $\mathcal{N}$ enracinés par un $\wedge$ est négligeable devant l'ensemble des tautologies simples qui a une fraction limite d'ordre $\Theta(1)$ au vu du Lemme 3.4.3.

Soit $t \in \mathcal{N}$ enraciné par un $\vee$ et dont exactement $\ell$ enfants de la racine sont des nœuds internes en lesquels sont enracinés les sous-arbres $A_1, \ldots, A_\ell$ et tels que $q$ littéraux différentes $\alpha_1, \ldots, \alpha_q$ apparaissent parmi les feuilles de première génération. Comme $t$ n'est pas une tautologie simple, une variable et sa négation ne peuvent apparaître simultanément comme étiquettes de deux feuilles de première génération. Pour tout $i \in \{1, \ldots, \ell\}$, les feuilles de première génération de l'arbre $A_i$ sont étiquetées soit par des étiquettes de $\{\alpha_1, \ldots, \alpha_q\}$, soit par de "nouvelles variables" ou leurs négations. Montrons qu'il existe au moins un indice $i \in \{1, \ldots, \ell\}$ tel que $A_i$ a au moins $\ell - 1$ nouvelles variables apparaissant comme étiquettes de ses feuilles de première génération.

Raisonnons par l'absurde et supposons que, pour tout $i \in \{1, \ldots, \ell\}$, au moins $\ell$ feuilles de première génération de $A_i$ soient étiquetées par une nouvelle variable. Considérons le sous-arbre $A_1$ qui est enraciné par $\wedge$. Fixons la valeur d'une des nouvelles variables à `Vrai` ou `Faux` de façon à ce que $A_1$ calcule `Faux` pour cette affectation partielle des variables. Comme $t$ calcule la fonction constante `Vrai` et non la fonction $\alpha_1 \vee \ldots \vee \alpha_q$, et que $t$ est enraciné par un $\vee$, nous savons qu'il existe au moins un sous-arbre $A_{(2)} \in \{A_2, \ldots, A_\ell\}$ qui n'est pas une contradiction pour cette affectation.

Raisonnons par induction. Après l'étape $m - 1 \leqslant \ell$, nous avons affecté les "nouvelles variables" $\nu_1, \nu_2, \ldots, \nu_{m-1}$ et au moins $m-1$ arbres parmi $\{A_1, \ldots, A_\ell\}$ calculent `Faux`. A l'étape $m$, remarquons que l'un des sous-arbres, noté $A_{(m)}$, n'est pas une contradiction pour cette affectation des variables, car sinon $t$ calcule $\alpha_1 \vee \ldots \vee \alpha_q$, ce qui est absurde. L'arbre $A_{(m)}$ est tel que, par hypothèse, au moins $\ell + 1$ nouvelles variables apparaissent comme étiquette d'une feuille de première génération. Nous avons affecté moins de $m - 1 \leqslant \ell$ variables, l'une des nouvelles variables apparaissant comme étiquette d'une feuille de première génération de $A_{(m)}$ n'est donc pas encore affectée, et nous pouvons donc lui choisir une affectation telle que $A_{(m)}$ calcule `Faux`.

Ainsi, après $\ell$ étapes, nous avons trouvé une affectation des "nouvelles variables" telle que tous les $A_1, \ldots, A_\ell$ calculent la fonction `Faux` pour cette affectation partielle. Dès lors, l'arbre total $t$ calcule la fonction $\alpha_1 \vee \ldots \vee \alpha_q$, ce qui est impossible, car, rappelons-le, $t$ est une tautologie.

Nous avons donc démontré par l'absurde qu'il existe au moins un sous-arbre de $\{A_1, \ldots, A_\ell\}$ dont les feuilles de première génération ont leurs étiquettes soit dans $\{\alpha_1, \ldots, \alpha_q\}$, soit dans un ensemble de cardinal plus petit que $\ell$ de nouvelles variables. Cela signifie que $\mathcal{N} \subseteq \bigcup_{q=0}^{k} \bigcup_{\ell=0}^{+\infty} \mathcal{M}^0_{q,\ell,\ell}$ (cf. Section 3.3.2). Si l'on décompose cette union en trois unions distinctes, nous obtenons :

$$\bigcup_{q=0}^{k} \bigcup_{\ell=0}^{+\infty} \mathcal{M}^0_{q,\ell,\ell} = \left( \bigcup_{q=0}^{k^{1/4}} \bigcup_{\ell=0}^{+\infty} \mathcal{M}^0_{q,\ell,\ell} \right) \cup \left( \bigcup_{q=1/4}^{k} \bigcup_{\ell=k^{1/8}}^{+\infty} \mathcal{M}^0_{q,\ell,\ell} \right) \cup \left( \bigcup_{q=1/4}^{k} \bigcup_{\ell=0}^{k^{1/8}} \mathcal{M}^0_{q,\ell,\ell} \right).$$

Le Lemme 3.3.5 nous assure que le premier ensemble de cette union a une fraction limite qui tend vers 0 quand $k$ tend vers $+\infty$, le Lemme 3.3.6 nous assure que le second ensemble de cette union a lui aussi une fraction limite qui tend vers 0 quand $k$ tend vers $+\infty$, et enfin, le Lemme 3.3.4 nous indique que la fraction limite du troisième terme de cette union vérifie :

$$\mu_k \left( \left( \bigcup_{q=1/4}^{k} \bigcup_{\ell=0}^{k^{1/8}} \mathcal{M}^0_{q,\ell,\ell} \right) \right) = \mathcal{O}\left( (k - k^{1/4}) k^{1/8} \frac{1}{k^{3/2}} \right) = \mathcal{O}\left( \frac{1}{k^{3/8}} \right)$$

et tend donc vers 0 quand $k$ tend vers $+\infty$. Nous avons donc montré que l'ensemble des tautologies qui ne sont pas des tautologies simples a une fraction limite qui tend vers 0 quand $k$ tend vers $+\infty$, et est donc négligeable devant l'ensemble des tautologies simples. ■

## 3.5 Fonctions littéral

L'objet de cette partie est d'étudier la complexité des fonctions littéral. En effet, nous avons déjà étudié les tautologies, et donc les contradictions (par symétrie). Pour mieux comprendre le modèle, il est intéressant de regarder les fonctions de complexité 1, même si elles ne sont a priori qu'un cas particulier du cas général. Pour étudier ces fonctions nous allons tout d'abord nous intéresser à la probabilité de l'ensemble des fonctions *plus grandes* qu'une fonction booléenne $f_0$ fixée.

### 3.5.1 Probabilité des fonctions plus grandes qu'une fonction fixée $f_0$

**Définition 3.5.1**

*Soient $f$ et $g$ deux fonctions booléennes à $k$ variables. On dira que $g$ est plus grande que $f$ (et $f$ est plus petite que $g$) et on notera $g \geqslant f$, si, et seulement si, $g(x_1, \ldots, x_k) \leqslant f(x_1, \ldots, x_k)$ pour tout $(x_1, \ldots, x_k) \in \{0,1\}^k$.*

Notre idée est de fixer une fonction booléenne $f_0$, et de calculer la probabilité $\mu_k(\{f \in \mathcal{F}_k \mid f \geqslant f_0\})$. Nous allons d'abord considérer le cas particulier où $f_0 = x_1 \wedge \ldots \wedge x_p$ où $p \geqslant 1$ est un entier, avant d'étudier le cas d'une fonction $f_0$ quelconque. Nous montrons le résultat suivant :

**Proposition 3.5.2**

*Pour toute fonction booléenne $f_0 \in \mathcal{F}_n$, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(f \geqslant f_0) \sim \mu_k(\texttt{Vrai}).$$

**Remarque :** Par symétrie du modèle, nous avons le résultat suivant : pour toute fonction booléenne $f_0 \in \mathcal{F}_n$, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k(f \leqslant f_0) \sim \mu_k(\texttt{Faux}) = \mu_k(\texttt{Vrai}).$$

**Démonstration : Supposons tout d'abord que $f_0 = \gamma_1 \wedge \ldots \wedge \gamma_p$.** Comme la fonction `Vrai` est plus grande que $f_0$, nous avons l'inégalité suivante : $\mu_k(f \geqslant f_0) \geqslant \mu_k(\texttt{Vrai}) \geqslant \alpha > 0$ (cf. Théorème 3.2.3). Soit $t$ un arbre non tautologique calculant une fonction plus grande que $f_0$.

La famille des arbres n'ayant aucune feuille de première génération a une fraction limite d'ordre $1/k\sqrt{2k}$, asymptotiquement quand $k$ tend vers $+\infty$ (cf. Proposition 3.3.1) et est donc négligeable devant $\mu_k(\texttt{Vrai})$. Dès lors, nous pouvons supposer que $t$ a au moins une feuille de première génération.

Si $t$ est enraciné par un $\vee$ :

– Supposons tout d'abord qu'il existe une feuille de première génération étiquetée par l'un des $\{\gamma_1, \ldots, \gamma_p\}$. La famille des arbres ayant au moins une feuille de première génération étiquetée par un $\{\gamma_1, \ldots, \gamma_p\}$ a pour série génératrice

$$G_p(z) = 2kz + 2z \cdot pz \frac{1}{(1 - \hat{B}(z) - (2k-p)z)(1 - \hat{A}(z))} - 2pz^2,$$

et sa fraction limite vérifie donc, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{z\to\rho_k}\frac{G'_p(z)}{A'(z)}\sim p\sqrt{\frac{2}{k}},$$

résultat qui est en accord avec la Proposition 3.3.2. La fraction limite de ces arbres est donc négligeable devant $\mu_k(\texttt{Vrai})$ : nous pouvons négliger cette famille.

– Supposons, au contraire, qu'aucune feuille de première génération de $t$ n'est étiquetée par un littéral de $\{\gamma_1,\dots,\gamma_p\}$. Notons $\{\alpha_1,\dots,\alpha_q\}$ les différents littéraux apparaissant comme étiquettes d'une feuille de première génération. Notons $\ell$ le nombre d'enfants de la racine qui sont des nœuds internes, on notera les sous-arbres enracinés en ces nœuds internes $A_1,\dots,A_\ell$. Comme $t$ n'est pas une tautologie, une variables et sa négation ne peuvent apparaître toutes deux comme étiquettes de feuilles de première génération. Les feuilles de première génération des sous-arbres $A_1,\dots,A_\ell$ sont étiquetées, soit par des littéraux de $\Omega=\{\alpha_1,\dots,\alpha_q,\gamma_1,\dots,\gamma_p\}$ et leurs négations, soit par des "nouvelles variables".
Supposons que pour tout $i\in\{1,\dots,\ell\}$, $A_i$ ait au moins $\ell$ nouvelles variables différentes apparaissant parmi les feuilles de première génération. Dès lors, comme nous l'avons déjà fait dans la Sous-Section 3.4.3, nous pouvons trouver une affectation des variables $\{x_1,\dots,x_k\}\backslash\Omega$ telle que $A_1,\dots,A_\ell$ calculent `Faux` pour cette affectation (rappelons que les $A_i$ sont tous enracinés par $\wedge$ car $t$ est enraciné par $\vee$). Nous pouvons donc ensuite affecter tous les $\alpha_1,\dots,\alpha_q$ à `Faux` et les $\gamma_1,\dots,\gamma_p$ à `Vrai`. Pour cette affectation, l'arbre $t$ calcule `Faux` et $f_0$ vaut `Vrai`, ce qui est absurde car nous avons supposé que $t$ calcule une fonction booléenne plus grande que $f_0$.
Il existe donc $i\in\{1,\dots,\ell\}$ tel que $A_i$ a au plus $\ell-1$ "nouvelles variables" apparaissant comme étiquettes de feuilles de première génération. Cela implique que $t$ appartient à la famille $\bigcup_{q=0}^{k}\bigcup_{r=0}^{\ell-1}\mathcal{M}^p_{q,\ell,r}$.
La fraction limite des arbres non tautologiques calculant une fonction plus grande que $f_0$ est donc plus petite que celle de $\bigcup_{q,\ell\geqslant 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}^p_{q,\ell,r}$, et d'après les résultats de la Section 3.3.2, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k\left(\bigcup_{q,\ell\geqslant 0}\bigcup_{r=0}^{\ell-1}\mathcal{M}^p_{q,\ell,r}\right)\sim\mu_k\left(\left(\bigcup_{q=1/4}^{k}\bigcup_{\ell=0}^{k^{1/8}}\mathcal{M}^p_{q,\ell,\ell}\right)\right)=\mathcal{O}\left((k-k^{1/4})k^{1/8}\frac{1}{k^{3/2}}\right)=\mathcal{O}\left(\frac{1}{k^{3/8}}\right).$$

Cette fraction limite étant négligeable devant celle des tautologies, il ne nous reste plus qu'à étudier le cas où $t$ est enraciné par un $\wedge$ pour en déduire la Proposition 3.5.2 pour le cas particulier où $f_0$ est une union de littéraux.

Supposons que $t$ soit enraciné par un $\wedge$ et calcule une fonction plus grande que $f_0=\gamma_1\wedge\ldots\wedge\gamma_p$. Alors, ses feuilles de première génération doivent être étiquetées par $\{\gamma_1,\dots,\gamma_p\}$, car sinon, il suffit d'affecter une de ces étiquettes à `Faux` pour que $t$ calcule `Faux`, indépendamment de $f_0=\gamma_1\wedge\ldots\wedge\gamma_p$ alors que nous avons supposé que $t$ calcule une fonction plus grande que $f_0$. La famille des arbres dont les feuilles de première génération sont étiquetées par des littéraux de $\{\gamma_1,\dots,\gamma_p\}$ a pour fonction génératrice

$$H_p(z)=2kz+2z\frac{(\hat{B}(z)+pz)^2}{1-\hat{B}(z)-pz}$$

et sa fraction limite est conséquemment d'ordre $1/k\sqrt{2}$, asymptotiquement quand $k$ tend vers $+\infty$. Cette famille est donc négligeable devant la famille des tautologies et nous avons donc bien montré la Proposition 3.5.2 dans le cas particulier où $f_0$ est une union de littéraux.

**Étudions maintenant le cas général : $f_0$ quelconque.** Toute fonction booléenne peut s'écrire sous la forme $f_0=(\gamma_1\wedge\ldots\wedge\gamma_p)\vee g_0$ avec $p\geqslant 1$ un entier, $\{\gamma_1,\dots,\gamma_p\}$ des littéraux et $g_0$ une fonction booléenne. Dès lors, d'après le cas particulier étudié ci-dessus, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k(\texttt{Vrai})\leqslant\mu_k(f\geqslant f_0)\leqslant\mu_k(f\geqslant\gamma_1\wedge\ldots\wedge\gamma_p)\sim\mu_k(\texttt{Vrai})$$

ce qui termine la preuve de la Proposition 3.5.2. ■

### 3.5.2 Probabilité d'une fonction littéral

Nous sommes maintenant en mesure de montrer le Théorème 3.2.4.

**Définition 3.5.3**

*Un arbre associatif booléen est un* **simple-$x$** *s'il est enraciné par un $\vee$ (resp. $\wedge$), si sa racine a deux enfants, l'un étant une feuille étiquetée par $x$ et l'autre étant une tautologie (resp. contradiction). On note $\mathcal{X}$ la famille des simple-$x$.*

**Lemme 3.5.4**

*Asymptotiquement quand $k$ tend vers $+\infty$, la fraction limite des simple-$x$ vérifie*

$$\mu_k(\mathcal{X}) \sim \frac{\mu_k(\texttt{Vrai})}{2k^2}.$$

**Démonstration :** La fonction génératrice des simple-$x$ est donnée par

$$X(z) = 2z^2T(z),$$

où $T(z)$ est la fonction génératrice des tautologies, et donc, sa fraction limite vérifie (cf. Lemme 1.5.1),

$$\mu_k(\mathcal{X}) = \lim_{z\to\rho_k} \frac{X'(z)}{A'(z)} = 2\rho_k^2 \lim_{z\to\rho_k} \frac{T'(z)}{A'(z)}.$$

Comme $\frac{T'(z)}{A'(z)}$ tend vers $\mu_k(\texttt{Vrai})$ quand $z$ tend vers $\rho_k$ et comme $\rho_k \sim 1/2k$ quand $k$ tend vers $+\infty$ (cf. Proposition 3.2.6), nous pouvons conclure la preuve. ■

**Théorème 3.5.5**

*Pour toute fonction littéral $f$, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(f) \sim \mu_k(\mathcal{X}).$$

**Démonstration :** Au vu du Lemme 3.5.4, nous savons que

$$\mu_k(f) \geqslant \frac{\mu_k(\texttt{Vrai})}{2k^2} \geqslant \frac{\alpha}{2k^2}$$

quand $k$ tend vers $+\infty$ (cf. Lemme 3.4.3 pour la définition de $\alpha$). Soit $t$ un arbre calculant la fonction littéral $f \equiv x$. Supposons que cet arbre soit enraciné par $\wedge$ (le cas où la racine est étiquetée par $\vee$ se traite de manière identique). La fraction limite de la famille des arbres $\mathcal{A}_x^{(0)}$ calculant $x$ et n'ayant aucune feuille de première génération ne dépend pas du choix de $x$ parmi les $2k$ littéraux $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$. Dès lors, via la Proposition 3.3.1, comme $\mathcal{A}^{(0)} = \bigcup_{x \text{ littéral}} \mathcal{A}_x^{(0)}$, asymptotiquement quand $k$ tend vers $+\infty$,

$$2k\mu_k(\mathcal{A}_x^{(0)}) \leqslant \mu_k(\mathcal{A}^{(0)}) \sim \frac{1}{k\sqrt{2k}}.$$

Nous en déduisons que la fraction limite des arbres calculant $x$ et n'ayant aucune feuille de première génération est d'ordre $\mathcal{O}(k^{-5/2})$, et est donc négligeable devant la famille des simple-$x$. Nous pouvons donc négliger cette famille et nous concentrer sur la famille des arbres ayant au moins une feuille de première génération.

Notons que les feuilles de première génération de $t$ doivent toutes être étiquetées par $x$, et que, de plus, chaque sous-arbres enraciné en un nœud interne de première génération doit calculer une fonction plus grande que $x$ (au sens de la Définition 3.5.1). Nous en déduisons qu'un arbre calculant

$x$ est presque sûrement, asymptotiquement quand $k$ tend vers $+\infty$, un arbre enraciné par $\wedge$ (resp. $\vee$), ayant au moins une feuille de première génération, dont les feuilles de première génération sont étiquetées par $x$ et dont les sous-arbres enracinée en des nœuds internes de première génération calculent des fonction plus grandes que $x$ (resp. plus petites que $x$).

On notera $L_x(z)$ la fonction génératrice des arbres calculant une fonction plus grande que $x$. Au vu de la Proposition 3.5.2, nous avons, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{z\to\rho_k} \frac{L'_x(z)}{A'(z)} \sim \mu_k(\texttt{Vrai}).$$

Par symétrie, la série génératrice des arbres calculant des fonctions plus petites que $x$ est égale à $L_x(z)$. La série génératrice des arbres non négligeables calculant $x$ est donc donnée par

$$T_x(z) = 2z\frac{z}{1-z}\frac{1}{1-L_x(z)} - 2z^2$$

et sa fraction limite vérifie, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k(\mathcal{T}_x) = \frac{2\rho_k}{1-\rho_k} \lim_{z\to\rho_k} \frac{L'_x(z)}{A'(z)} \sim \frac{\mu_k(\texttt{Vrai})}{2k^2}.$$

Nous en concluons que, asymptotiquement quand $k$ tend vers $+\infty$

$$\mu_k(x) \sim \mu_k(\mathcal{X}) \sim \frac{\mu_k(\texttt{Vrai})}{2k^2}.$$

■

Les preuves développées dans cette Section sont généralisables pour toute conjonction ou disjonction de littéraux : par exemple $((x_1,\dots,x_k) \mapsto x_1 \wedge x_2 \wedge x_3)$. Nous ne détaillons pas ces preuves, mais il est possible de montrer le lemme suivant :

**Lemme 3.5.6**

*Pour tout entier $p \geqslant 2$, pour toute fonction $(f : (x_1,\dots,x_k) \mapsto \alpha_1 \diamond \dots \diamond \alpha_p)$ où $\alpha_1,\dots,\alpha_p$ sont des littéraux deux à deux distincts (ne contenant pas une variable et sa négation) et où $\diamond \in \{\wedge, \vee\}$, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(f) \sim \frac{\mu_k(\texttt{Vrai})}{k^{p+1}} = \frac{\mu_k(\texttt{Vrai})}{k^{L(f)}}.$$

## 3.6 Cas général : arbres minimaux et expansions.

La preuve complète de la Conjecture 3.2.5 n'est pas encore établie. Nous avons pour idée de suivre les étapes utilisées pour le modèle de l'implication dans [FGGG12] : il s'agit de montrer que, asymptotiquement quand $k$ tend vers $+\infty$, presque tout arbre calculant $f$ est une *expansion* d'un arbre minimal de $f$. Pour ce faire, nous calculons tout d'abord la fraction limite des expansions d'arbres minimaux de $f$, puis, nous montrons que les *expansions d'expansions* sont négligeables devant les expansions d'arbres minimaux de $f$, et enfin, nous conjecturons que les arbres calculant $f$, qui ne sont pas expansions (ou issus d'expansions successives) d'un arbre minimal de $f$, mais de ce que nous appellerons un *arbre irréductible*, sont eux aussi négligeables. En pratique, si les preuves des deux premières étapes sont établies et détaillées dans cette partie, le traitement des arbres irréductibles reste ouvert.

### 3.6.1 Expansions

La notion d'expansion est la même que celle introduite dans le traitement des arbres associatifs du Chapitre 2 (cf. Définition 2.5.2). Ceci dit, il semble qu'il suffise dans ce nouveau modèle, de considérer les T-expansions, et non les X-expansions (cf. Définition 2.5.3), cela car les tautologies ont une fraction limite d'ordre $\mathcal{O}(1)$ alors que les arbres calculant un littéral ont une fraction limite d'ordre $\mathcal{O}(1/k^2)$ (soit inférieur de deux ordres de grandeurs et non d'un seul comme dans le cas étudié dans le chapitre précédent.), asymptotiquement quand $k$ tend vers $+\infty$.

Nous reprenons la notion d'expansion définie en Définition 2.5.2. Pour toute famille d'arbres $\mathcal{T}$, on note $E(\mathcal{T})$ l'ensemble des T-expansions d'arbres de $\mathcal{T}$. On notera $E^q(\mathcal{T})$ les arbres obtenus après $q$ T-expansions consécutives à partir d'arbres de $\mathcal{T}$, et $E^*(\mathcal{T}) = \bigcup_{q\geqslant 0} E(\mathcal{T})$.

Remarquons que si l'on fait deux expansions successives en un arbre $t$, et si la seconde expansion est faite en un nœud de l'arbre greffé lors de la première expansion, alors, l'arbre obtenu est une T-expansion de $t$. Tout au long de notre étude, il nous suffit donc de considérer des expansions faites successivement en des nœuds de l'arbre de départ.

**Proposition 3.6.1**

*Asymptotiquement quand $k$ tend vers $+\infty$, la fraction limite de $E^*(\mathcal{M}_f) = \bigcup_{q\geqslant 0} E(\mathcal{T})$, est équivalente à la fraction limite de $E(\mathcal{M}_f)$. De plus,*

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

*Comme tout arbre de $E^*(\mathcal{M}_f)$ calcule $f$, nous avons :*

$$\mu_k(f) \geqslant \mu_k(E^*(\mathcal{M}_f)) = \Theta\left(\frac{1}{k^{L(f)}}\right).$$

**Démonstration :** Soit $\Phi_q(z)$ la série génératrice des arbres obtenus après $q$ expansions successives d'un arbre minimal de $f$, chaque expansion étant faite en un nœud de l'arbre minimal de départ. Étant donné un arbre $t_e$, le nombre de façons différentes de greffer cet arbre dans un arbre minimal $t$ de $f$ est donnée par

$$P_t = \sum_{i \text{ nœud interne de } t} (d(i)+1)$$

où $d(i)$ est l'arité du nœud $i$. Dès lors,

$$P_t = i_t + |t| - 1$$

où $i_t$ est le nombre de nœuds internes de $t$, et $|t|$ la taille de $t$. Comme $t$ est minimal et comme $1 \leqslant i_t \leqslant \lfloor \frac{L(f)}{2} \rfloor$ (la forme binaire est celle qui maximise le nombre de nœuds internes, et un arbre binaire à $m$ feuilles a $m-1$ noeuds internes),

$$L(f) \leqslant P_t \leqslant \frac{3L(f)}{2}.$$

Si $q = 1$,

$$\Phi_1(z) \prec m_f z^{L(f)} \frac{3L(f)}{2} T(z),$$

où $m_f$ est le nombre d'arbres minimaux de $f$, où $T(z)$ est la fonction génératrice des tautologies[3], ce qui implique

$$m_f L(f) \rho_k^{L(f)} \lim_{z\to\rho_k} \frac{T'(z)}{A'(z)} \leqslant \mu_k(E(\mathcal{M}_f) \leqslant m_f \frac{3L(f)}{2} \rho_k^{L(f)} \lim_{z\to\rho_k} \frac{T'(z)}{A'(z)}.$$

3. Rappelons que $A(z) \prec B(z)$ si, et seulement si, pour tout entier $n$, $[z^n]A(z) \leqslant [z^n]B(z)$

Comme, asymptotiquement quand $k$ tend vers $+\infty$, $\mu_k(\texttt{Vrai}) = \Theta(1)$ (cf. Théorème 3.2.3), et comme $\rho_k \sim \frac{1}{2k}$ quand $k$ tend vers $+\infty$, nous obtenons

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

Si l'on fait $q$ expansions successives en des nœuds d'un arbre minimal de $f$, il y a au plus $\lfloor 3/2 L(f) \rfloor$ emplacements différents pour la première, $\lfloor 3/2 L(f) \rfloor + 1$ pour la seconde, et ainsi de suite. Dès lors,

$$\Phi_q(z) \prec m_f z^{L(f)} \binom{\lfloor 3/2 L(f) \rfloor + q}{q} T(z)^q$$

ce qui implique

$$\begin{aligned}\mu_k(E^q(\mathcal{M}_f)) = \lim_{z\to\rho_k} \frac{\Phi_q'(z)}{A'(z)} &\leqslant m_f \rho_k^{L(f)} \binom{\lfloor 3/2 L(f) \rfloor + q}{q} q T(\rho_k)^{q-1} \lim_{z\to\rho_k} \frac{T'(z)}{A'(z)} \\ &\leqslant m_f \rho_k^{L(f)} q \binom{\lfloor 3/2 L(f) \rfloor + q}{q} T(\rho_k)^{q-1} \mu_k(\texttt{Vrai}).\end{aligned}$$

Si $q \geqslant 2$, asymptotiquement quand $k$ tend vers $+\infty$,

$$\mu_k(E^q(\mathcal{M}_f)) \leqslant \beta m_f q \binom{\lfloor 3/2 L(f) \rfloor + q}{q} \rho_k^{L(f)} T(\rho_k)^{q-1},$$

où la constante $\beta$ vérifie $0 < \beta \leqslant \mu_k(\texttt{Vrai})$ (pour tout entier $k$) est définie dans le Théorème 3.2.3. Pour conclure,

$$\begin{aligned}\mu_k(E^{q\geqslant 2}(\mathcal{M}_f)) &\leqslant \beta m_f \rho_k^{L(f)} \sum_{q\geqslant 2} q \binom{\lfloor 3/2 L(f) \rfloor + q}{q} T(\rho_k)^{q-1} \\ &= \beta m_f \rho_k^{L(f)} \sum_{q\geqslant 1} (q+1) \binom{\lfloor 3/2 L(f) \rfloor + (q+1)}{(q+1)} T(\rho_k)^{q} \\ &= m_f \rho_k^{L(f)} T(\rho_k) \sum_{q\geqslant 1} (q+1) \binom{\lfloor 3/2 L(f) \rfloor + (q+1)}{(q+1)} T(\rho_k)^{q-1}.\end{aligned}$$

Comme $(q+1)\binom{\lfloor 3/2 L(f) \rfloor + (q+1)}{(q+1)} = \frac{C+q+1}{q} q \binom{C+q}{q}$, on obtient,

$$\mu_k(E^{q\geqslant 2}(\mathcal{M}_f)) \leqslant (C+2) m_f \rho_k^{L(f)} T(\rho_k) \sum_{q\geqslant 1} q \binom{\lfloor 3/2 L(f) \rfloor + q}{q} T(\rho_k)^{q-1}$$

D'après [GKP94, page 174], pour tout $C \in \mathbb{N}$,

$$\sum_{m\geqslant 0} \binom{C+m}{m} z^m = \frac{1}{(1-z)^{C+1}}. \tag{3.3}$$

Dès lors,

$$\sum_{m\geqslant 1} m \binom{C+m}{m} z^{m-1} = \frac{C+1}{(1-z)^{C+2}},$$

et

$$\mu_k(E^{q\geqslant 2}(\mathcal{M}_f)) \leqslant \beta m_f \rho_k^{L(f)} T(\rho_k) \frac{\lfloor 3/2 L(f) \rfloor + 1}{(1-T(\rho))^{\lfloor 3/2 L(f) \rfloor + 2}}.$$

Enfin, comme $T(\rho_k) \leqslant \hat{B}(\rho_k) \leqslant 1/\sqrt{2k}$ (car une tautologie ne peut être un arbre de taille 1),

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

■

### 3.6.2 Arbres irreductibles

**Définition 3.6.2**

*Soit $t$ un arbre associatif. Si $t$ ne peut être obtenu comme T-expansion (valide) d'un arbre plus petit, alors $t$ est un* **arbre irréductible**. *On notera $\mathcal{I}_f$ l'ensemble des arbres irréductibles calculant $f$ qui ne sont pas des arbres minimaux de $f$.*

Un arbre calculant $f$ est une expansion (ou issu d'une succession d'expansions) soit d'un arbre minimal de $f$, soit d'un arbre irréductible de $f$ :

$$\mu_k(f) \leqslant \mu_k(E^*(\mathcal{M}_f)) + \mu_k(E^*(\mathcal{I}_f)). \tag{3.4}$$

**Conjecture 3.6.3**

*Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(E^*(I_f)) = o\left(\frac{1}{k^{L(f)}}\right).$$

Montrer cette conjecture permettrait de prouver la Conjecture 3.2.5.

## 3.7 Conclusion

Dans ce chapitre nous avons ébauché l'étude d'un nouveau modèle d'arbres associatifs dans lequel la taille d'un arbre est calculée en terme de nœuds, et non plus en terme de feuilles. Ce changement, minime à première vue, met en échec les approches utilisées dans les cas classiques d'arbres aléatoires, notamment la théorie des motifs de Kozik. Dans ce chapitre, nous étudions la probabilité des fonctions constantes, celle des fonctions littéral et présentons les premières étapes d'une étude possible du cas général.

La preuve de la Conjecture 3.2.5 reste ouverte. Nous avons montré qu'un moyen de clore cette preuve est de terminer l'étude des arbres irréductibles en démontrant la Conjecture 3.6.3.

Il est remarquable que, bien que les méthodes de preuve usuelles s'avèrent inefficaces, nous conjecturons un comportement similaire à celui des distributions arborescentes usuelles (cf. Chapitre 2), à savoir un comportement de $\mu_k(f)$ en $\frac{1}{k^{L(f)}}$ quand $k$ tend vers $+\infty$. Ce modèle n'est donc finalement pas si différent du modèle où la taille est mesurée en nombre de feuilles.

# Chapitre 4

# Arbres implicatifs généraux

## 4.1 Des arbres implicatifs non binaires, non plans

L'idée de ce chapitre est la même que celle du Chapitre 2 : il s'agit de définir de nouveaux modèles d'arbres prenant en compte les propriétés logiques inhérentes au connecteurs logiques. Dans ce chapitre, nous nous intéressons aux arbres implicatifs, dont les nœuds internes sont donc tous étiquetés par le connecteur $\rightarrow$. Ce connecteur a la propriété suivante : toute formule implicative s'écrit $(A_1 \rightarrow (A_2 \rightarrow \dots (A_p \rightarrow \alpha)))$ où $A_1, \dots, A_p$ sont des formules implicatives, et l'ordre des *prémisses* $A_1, \dots, A_p$ n'influe pas la fonction booléenne calculée par cette formule. D'un point de vue *arborescent*, un arbre de l'implication binaire plan peut être décrit comme une branche droite de longueur $p$ de nœuds internes étiquetés par $\rightarrow$, dont les fils gauches sont des arbres implicatifs $T_1, \dots, T_p$ et dont la feuille extrême est étiquetée par un littéral positif $\alpha \in \{x_1, \dots, x_k\}$. Les arbres obtenus par permutation des arbres $T_1, \dots, T_p$ calculent tous la même fonction booléenne (cf. Figure 4.1).

Afin de prendre en compte cette propriété, nous représenterons la formule booléenne $(A_1 \rightarrow (A_2 \rightarrow \dots (A_p \rightarrow \alpha)))$ par un arbre non-binaire, non-plan, dont la racine est étiquetée par $\alpha$, et donc les sous-arbres représentent les formules $A_1, \dots, A_p$. Ce modèle est un nouveau modèle prenant en compte les propriétés logiques du connecteur $\rightarrow$. La taille d'un tel arbre implicatif sera le nombre de ses feuilles, et donc le nombre de littéraux intervenant dans la formule implicative associée : cette notion de taille est bien la même que dans le cas binaire plan, et les comparaisons entre ces deux modèles seront donc sensées.

Considérons la distribution uniforme sur les arbres implicatifs généraux de taille $n$ étiquetés sur $k$ variables : nous montrerons que la suite des distributions induites sur $\mathcal{F}_k$ converge vers une distribution limite $\mu_k$ quand $n$ tend vers $+\infty$. L'objet de ce chapitre est d'étudier $\mu_k$ et de comparer cette nouvelle distribution à celle obtenue dans l'étude de arbres implicatifs binaires plans (cf. Section 1.3.2). Nous verrons que les méthodes utilisées dans le cas classique s'adaptent sans difficulté à ce nouveau modèle et que les raisonnements deviennent plus simples de par la structure minimaliste des arbres considérés : l'information superflue du connecteur $\rightarrow$ qui apparaît sur tous les nœuds internes dans le cas binaire plan est ici sous-entendue. Par ailleurs, cette nouvelle distribution a un comportement très similaire à celle du cas classique, même si nous verrons que les deux distributions ne sont pas égales.

La Section 4.2 définit le modèle étudié et énonce le résultat principal de ce chapitre ; elle contient aussi la preuve de l'existence de la distribution asymptotique quand $n$ tend vers $+\infty$ (via le théorème de Drmota-Lalley-Woods). La Section 4.3 est consacrée à l'étude des tautologies dans ce modèle : nous y montrons en particulier que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tauto-

logie est simple. La Section 4.4 est consacrée à la preuve du résultat principal : les méthodes utilisées sont celles introduites dans le cadre du modèle implicatif classique (cf. Fournier et al. [FGGG12]).

## 4.2 Description du modèle et résultat principal

**Définition 4.2.1**

*Une* **formule (ou expression) implicative** *est soit une variable de* $\{x_1, \dots, x_k\}$*, soit une expression de la forme* $\{A_1, A_2, \dots A_p\} \to \alpha$ *où* $A_1, \dots, A_p$ *sont des formules implicatives, et* $\alpha$ *un littéral de* $\{x_1, \dots, x_k\}$*. L'ordre des prémisses* $A_1, \dots, A_p$ *ne change pas la fonction booléenne calculée par cette expression : toutes les expressions* $\{A_{\sigma(1)}, A_{\sigma(2)}, \dots, A_{\sigma(p)}\} \to \alpha$ *où* $\sigma$ *est une permutation de* $\{1, \dots, p\}$ *sont égales. La variable* $\alpha$ *est le* **but** *de cette expression. La* **taille** *d'une formule est le nombre de variables qui y apparaissent, en comptant les éventuelles répétitions.*

**Remarque :** Rappelons qu'une variable est un élément de $\{x_1, \dots, x_k\}$, et qu'un littéral est une variable ou la négation d'une variable. Comme dans ce chapitre, nous ne considérons que des littéraux positifs, les termes variable et littéral représentent le même objet, à savoir un élément de $\{x_1, \dots, x_k\}$.

**Exemple :** Les deux formules $\{(x_1 \to x_2), x_1\} \to x_3$ et $\{x_1, (x_1 \to x_2)\} \to x_3$ représentent la même fonction booléenne $((x_1, \dots, x_k) \mapsto (x_1 \wedge (x_1 \to x_2)) \to x_3)$. Ces deux formules sont de taille 4.

Les formules implicatives peuvent être représentées par des arbres. Considérer des arbres non plans nous permet de représenter par un même arbre les expressions identiques à permutation des prémisses près.

**Définition 4.2.2 (cf. Figure 4.1)**

*Un* **arbre général implicatif** *est un arbre non plan enraciné dont les nœuds sont étiquetés par des variables de* $\{x_1, \dots, x_k\}$*. La* **taille** *d'un arbre général implicatif sera le nombre total de ses nœuds. L'ensemble de tous les arbres généraux implicatifs est noté* $\mathcal{I}$*, et le nombre de ces arbres de taille* $n$ *est noté* $I_{n,k}$*.*

Les arbres généraux implicatifs sont une représentation naturelle des expressions implicatives : l'étiquette de la racine est le but de l'expression représentée par un arbre, et les sous-arbres en sont les prémisses. Les deux notions de taille sont elles aussi cohérentes : un arbre de taille $n$ représente une formule de taille $n$.

**Définition 4.2.3**

*Soit* $f \in \mathcal{F}_k$*. On note* $I_{n,k}(f)$ *le nombre d'arbres implicatifs généraux calculant* $f$ *et*

$$\mu_{n,k}(f) = \frac{I_{n,k}(f)}{I_{n,k}}$$

*la proportion d'arbre de taille* $n$ *calculant* $f$*. On s'intéresse à la limite de cette proportion quand* $n$ *tend vers* $+\infty$ *:*

$$\mu_k(f) = \lim_{n \to +\infty} \mu_{n,k}(f),$$

*et on appelle cette limite, si elle existe,* **probabilité de** $f$*.*

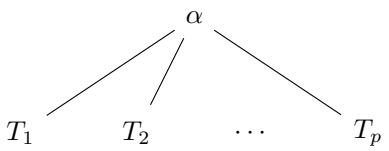



FIGURE 4.1 – *Cet arbre est un arbre général implicatif qui représente l'expression* $A_1 \to (A_2 \to (A_3 \to \ldots (A_p \to \alpha)))$ *où* $A_1, \ldots, A_p$ *sont les expressions représentées par les arbres généraux implicatifs* $T_1, \ldots, T_p$.

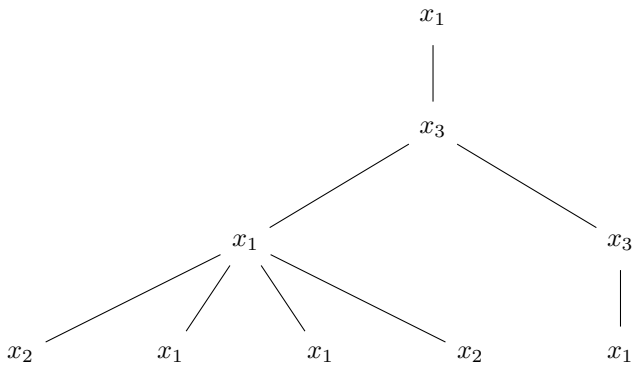



FIGURE 4.2 – *Cet arbre général implicatif représente la fonction booléenne* $((x_1, \ldots, x_k) \mapsto (x_3 \to x_1))$.

**Remarque :** Comme signalé en préliminaire de ce mémoire, le système logique de l'implication n'est pas complet. Une fonction booléenne non expressible dans ce système aura donc une probabilité nulle dans ce modèle. Il nous sera parfois nécessaire de ne considérer que les fonctions expressibles dans le système de l'implication : nous noterons $\mathcal{E}_k$ cet ensemble.

**Lemme 4.2.4**

*Pour toute fonction* $f \in \mathcal{E}_k$, *la probabilité*

$$\mu_k(f) = \lim_{n \to +\infty} \mu_{n,k}(f)$$

*existe et est strictement positive.*

**Démonstration :** La famille des arbres généraux implicatifs est décrite par la spécification suivante :

$$\mathcal{I} = \mathcal{L} \times \texttt{multiset}(\mathcal{I}),$$

où $\mathcal{L}$ est l'ensemble des variables $\{x_1, \ldots, x_k\}$. Par la méthode symbolique, la série génératrice des arbres généraux implicatifs est donc donnée par

$$I(z) = \sum_{n \geqslant 1} I_{n,k} z^n = kz \exp\left( \sum_{i \geqslant 1} \frac{P(z^i)}{i} \right).$$

Pour toute fonction booléenne fixée, posons

$$I_f(z) = \sum_{n \geqslant 1} I_{n,k}(f) z^n.$$

Via la méthode symbolique, nous obtenons

$$I_f(z) = z\mathbb{1}_{\{f \text{ lit}\}} + z\sum_{j=1}^{k}\sum_{\ell=1}^{+\infty}\sum_{\{g_1,\dots,g_\ell\}\to x_j=f}\prod_{p=1}^{\ell}\exp\left(\sum_{i=1}^{+\infty}\frac{I_{g_p}(z^i)}{i}\right),$$

où $\mathbb{1}_{\{f \text{ lit}\}}$ vaut 1 si $f$ est une fonction littéral et 0 sinon, où l'indice $j$ correspond au choix de la variable qui étiquette la racine de l'arbre et l'indice $\ell$ représente le nombre de fonctions booléennes deux à deux distinctes calculées par au moins un sous-arbre de la racine. La somme intérieure se fait donc sur tous les différents choix de $\ell$ fonctions booléennes distinctes vérifiant la condition indiquée. Cette équation étant vérifiée pour toute fonction booléenne $f$ de $\mathcal{F}_k$, l'ensemble des ces équations forme un système d'équations fonctionnelles.

**Remarque :** Le théorème de Drmota-Lalley-Woods nécessite une hypothèse de forte connexité du graphe de dépendance du système d'équations fonctionnelles. Pour que cette hypothèse soit vérifiée, il ne faut considérer que les fonctions booléennes expressibles dans le système de l'implication, i.e. les fonctions de $\mathcal{E}_k$, par définition.

Le théorème de Drmota-Lalley-Woods [Drm09, Lal93, Woo97] (cf. [FS09, page 489]) peut être appliqué à ce système (une vérification des hypothèses dans le cadre du modèle des arbres binaires de l'implication peut être lue dans [FGGG12] ; nous ne détaillons pas cette vérification ici) : les séries génératrices $(I_f(z))_{f\in\mathcal{E}_k}$ ont toutes la même singularité $\eta_k$, et cette singularité est de type racine carrée. Le lemme de transfert de Flajolet et Odlyzsko [FO90] permet donc d'affirmer que, pour toute fonction booléenne de $\mathcal{E}_k$, il existe une constante $c_k(f) > 0$ telle que, asymptotiquement quand $n$ tend vers $+\infty$,

$$I_{n,k}(f) \sim c_k(f) n^{-3/2}\eta_k^{-n}.$$

Notons que, asymptotiquement quand $n$ tend vers $+\infty$,

$$I_{n,k} = \sum_{f\in\mathcal{E}_k} I_{n,k}(f) \sim c_k n^{-3/2}\eta_k^{-n},$$

où $c_k = \sum_{f\in\mathcal{E}_k} c_k(f) > 0$. Dès lors, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mu_{n,k}(f) = \frac{I_{n,k}(f)}{I_{n,k}} \sim \frac{c_k(f)}{c_k} > 0,$$

et $\mu_k(f)$ existe pour tout $f \in \mathcal{E}_k$ et est strictement positive. ■

L'objectif de ce chapitre est d'étudier la probabilité d'une fonction $f$, notamment en fonction de sa complexité. Nous allons montrer le théorème suivant :

**Théorème 4.2.5**

*Soit $f$ une fonction booléenne de $\mathcal{E}_k$ dont le nombre de variables essentielles $E(f)$ ne dépend pas de $k$. Alors, il existe une constante $\lambda_f > 0$ telle que, asymptotiquement quand $k$ tend vers $+\infty$,*

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

## 4.3 Tautologies

La première étape de la preuve du Théorème 4.2.5 est l'étude des tautologies, i.e. de la fonction constante `Vrai`. Pour énumérer les tautologies, nous allons montrer que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est *simple*. Nous suivons les étapes de la preuve développée dans le cadre des arbres binaires plans de l'implication (cf. [FGGZ10]).

**Définition 4.3.1 (cf. Figure 4.3)**

*Une* **tautologie simple** *est un arbre général implicatif dont exactement une prémisse est de taille* 1 *et est étiquetée par la même variable que la racine. Si cette variable est* $x$*, on dira que cette tautologie simple est* **réalisée** *par* $x$*.*

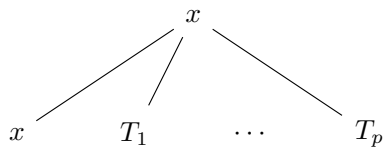



FIGURE 4.3 – *Une tautologie simple réalisée par la variable* $x$*.*

**Définition 4.3.2 (cf. Figure 4.4)**

*Une* **simple non-tautologie** *est un arbre général implicatif dont l'étiquette de la racine est différent des étiquettes de tous ses enfants.*

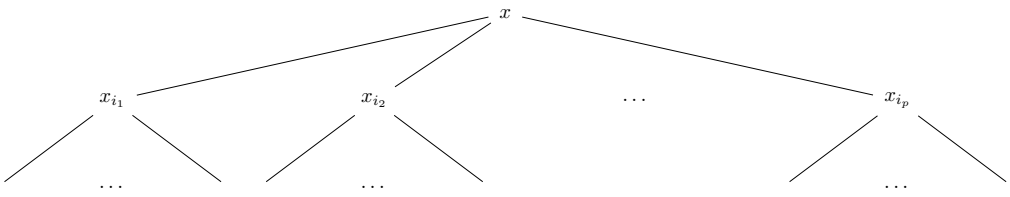



FIGURE 4.4 – *Une simple non-tautologie :* $x \notin \{x_{i_1}, \ldots, x_{i_p}\}$.

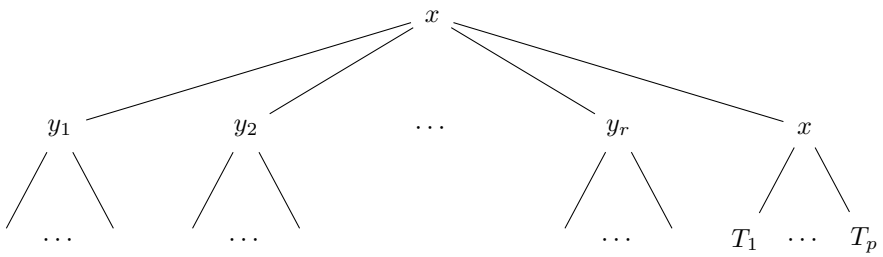
**Définition 4.3.3 (cf. Figure 4.5)**

*Une* **non-tautologie moins simple** *est un arbre général implicatif tel que*
- *exactement un enfant de la racine, noté* $s$*, a la même étiquette* $x$ *que la racine,*
- *le sous-arbre enraciné en* $s$ *est de taille au moins 2,*
- *toute étiquette d'un enfant de* $s$ *est différente de toutes les étiquettes des nœuds à hauteur* 1 *dans l'arbre, et tous les petits-enfants de* $s$ *ont une étiquette distincte de* $x$ *et de l'étiquette de leurs parents respectifs.*

**Remarque :** Comme leurs noms l'indiquent, les tautologies simples calculent la fonction constante `Vrai`, et que les non-tautologies simples ainsi que les non-tautologies moins simples ne calculent pas la fonction `Vrai`.

Nous allons montrer que presque toute formule est, asymptotiquement quand $k$ tend vers $+\infty$, une tautologie simple, une non-tautologie simple ou une non-tautologie moins simple. Ce résultat impliquera que, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est une tautologie simple.

On notera $\mathcal{ST}$ l'ensemble des tautologies simples, $\mathcal{NST}$ and $\mathcal{LNST}$ les ensemble des non-tautologies simples et des non-tautologies moins simples respectivement. De plus, on notera $ST_{n,k}$, $NST_{n,k}$ et $LNST_{n,k}$ respectivement, le nombre de tautologies simples, non-tautologies simples et non-tautologies moins simples de taille $n$ et étiquetées sur $k$ variables.




les $(T_i)_{i=1\ldots p}$ sont des arbres enracinés respectivement par les $w_i \notin \{x, y_1, \ldots, y_r\}$ et dont les nœuds de première génération sont étiquetés par les $z_{ij} \notin \{x, w_i\}$ :

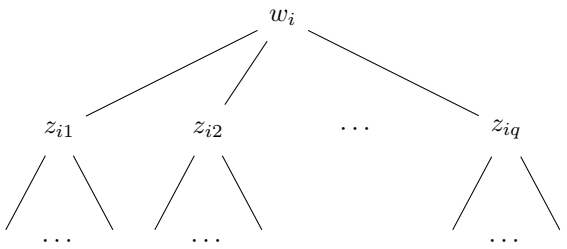



FIGURE 4.5 – *Une non-tautologie moins simple.*

**Théorème 4.3.4**

*Asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est simple. Autrement dit, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\mu_k(\texttt{Vrai}) \sim \lim_{n\to+\infty} \frac{ST_{n,k}}{I_{n,k}}.$$

Pour montrer ce résultat, nous allons montrer que, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{n\to+\infty} \frac{ST_{n,k} + NST_{n,k} + LNST_{n,k}}{I_{n,k}} = 1 - \mathcal{O}\left(\frac{1}{k^2}\right),$$

ce qui impliquera

$$\mu_k(\texttt{Vrai}) = \lim_{n\to+\infty} \frac{ST_{n,k}}{I_{n,k}} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

La proposition suivante nous assure que $\lim_{n\to+\infty} \frac{ST_{n,k}}{I_{n,k}}$ est d'ordre $\frac{1}{k}$ et que les termes en $\mathcal{O}\left(\frac{1}{k^2}\right)$ sont en effet négligeables devant la contribution des tautologies simples.

**Proposition 4.3.5**

*Asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty} \frac{ST_{n,k}}{I_{n,k}} = \frac{1}{\mathtt{e}k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

**Démonstration :** Par définition, l'ensemble des tautologies simples est défini par la spécification suivante :

$$\mathcal{L} \times \mathcal{Z} \times \texttt{multiset}(\mathcal{I} \backslash \mathcal{Z}),$$

où $\mathcal{I}$ et $\mathcal{L}$ sont respectivement l'ensemble des arbres généraux implicatifs et l'ensemble des variables $\{x_1, \dots, x_k\}$, et $\mathcal{Z}$ est l'ensemble constitué d'une seule variable. L'occurrence du $\mathcal{Z}$ dans la spécification correspond à l'unique occurrence de la variable de la racine parmi les nœuds de première génération. Par la méthode symbolique, la fonction génératrice des tautologies simples est donc donnée par

$$ST(z) = kz^2 \exp\left(\sum_{i\geqslant 1} \frac{I(z^i) - z^i}{i}\right) = zI(z) \exp\left(-\sum_{i\geqslant 1} \frac{z^i}{i}\right) = z(1-z)I(z). \tag{4.1}$$

Nous pouvons déduire de cette équation que les fonctions génératrices $ST(z)$ et $I(z)$ ont la même singularité dominante $\eta_k$, et que cette singularité est de type racine carrée. Dès lors,

$$\lim_{n\to+\infty} \frac{ST_{n,k}}{I_{n,k}} = \lim_{z\to\eta_k} \frac{ST'(z)}{I'(z)}.$$

Nous avons

$$\frac{ST'(z)}{I'(z)} = \frac{(1-z)I(z)}{I'(z)} - \frac{zI(z)}{I'(z)} + \frac{z(1-z)I'(z)}{I'(z)}.$$

Comme $\eta_k$ est une singularité de type racine carrée de $I$, nous savons que $I'(z)$ tend vers $+\infty$ quand $z$ tend vers $\eta_k$. Dès lors, les deux premiers termes de cette somme tendent vers 0 quand $z$ tend vers $\eta_k$, et

$$\lim_{z\to\eta_k} \frac{ST'(z)}{I'(z)} = \eta_k(1-\eta_k).$$

A défaut de connaître une formule close pour $\eta_k$, il nous faut déterminer son comportement asymptotique quand $k$ tend vers $+\infty$ : via une application du théorème des fonctions implicites (cf. [Drm09, Theorem 2.19]), $(\eta_k, I(\eta_k))$ vérifie le système d'équations

$$\begin{cases} Y = kXQ(X)\mathsf{e}^Y \\ 1 = kXQ(X)\mathsf{e}^Y \end{cases}$$

où $Q(z) = \exp\left(\sum_{i\geqslant 2} \frac{I(z^i)}{i}\right)$ est une fonction analytique en $\eta_k$. Dès lors, $Y = 1$ et $X = \frac{1}{k\mathsf{e}Q(X)}$. Comme $Q(z) \geqslant 1$ pour tout réel positif $z$, et comme $X = \eta_k$ est un réel positif, nous obtenons $X = \mathcal{O}\left(\frac{1}{k}\right)$. Pour tout entier $i \geqslant 2$,

$$I(X^i) = \sum_{n\geqslant 1} I_{n,k} X^{in} = X^{i-1} \sum_{n\geqslant 1} I_{n,k} X^{i(n-1)+1} \leqslant X^{i-1} \sum_{n\geqslant 1} I_{n,k} X^n = X^{i-1} I(X).$$

Donc, asymptotiquement quand $k$ tend vers $+\infty$,

$$\ln Q(X) = \sum_{i\geqslant 2} \frac{I(X^i)}{i} \leqslant I(X) \sum_{i\geqslant 2} \frac{X^{i-1}}{i} = \frac{Y}{X} \sum_{i\geqslant 2} \frac{X^i}{i} = \frac{1}{X}(-\ln(1-X) - X) \sim X = \mathcal{O}\left(\frac{1}{k}\right).$$

Autrement dit, $Q(X) = 1 + \mathcal{O}\left(\frac{1}{k}\right)$, et $X = \frac{1}{\mathsf{e}k} + \mathcal{O}\left(\frac{1}{k^2}\right)$. En réinjectant ce développement asymptotique dans le système, il est possible d'obtenir

$$\eta_k = \frac{1}{\mathsf{e}k} - \frac{1}{2\mathsf{e}^2k^2} + \mathcal{O}\left(\frac{1}{k^3}\right),$$

développement qui nous sera utile plus tard. Nous obtenons donc finalement, grâce au premier ordre du développement de $\eta_k$, $\lim_{z\to\eta_k} \frac{ST'(z)}{I'(z)} = \eta_k(1-\eta_k) = \frac{1}{\mathsf{e}k} + \mathcal{O}\left(\frac{1}{k^2}\right)$. ■

Résumons dans le lemme suivant les résultats obtenus sur la singularité dominante de $I(z)$ :

**Lemme 4.3.6**

*Soit $I(z) = \sum_{n\geqslant 1} I_{n,k} z^n$ la fonction génératrice des arbres généraux implicatifs, et soit $\eta_k$ sa singularité dominante. Alors,*

$$\eta_k = \frac{1}{\mathsf{e}k} - \frac{1}{2\mathsf{e}^2k^2} + \mathcal{O}\left(\frac{1}{k^3}\right),$$

*et*

$$I(\eta_k) = 1.$$

**Proposition 4.3.7**

*Asymptotiquement quand $k$ tend vers $+\infty$, $\lim_{n\to+\infty} \frac{SNT_{n,k}}{I_{n,k}} = 1 - \frac{2}{k} + \mathcal{O}\left(\frac{1}{k^2}\right)$.*

**Démonstration :** L'ensemble des non-tautologies simples vérifie la spécification suivante :

$$\mathcal{SNT} = \mathcal{L} \times \texttt{multiset}((\mathcal{L}\backslash\mathcal{Z}) \times \texttt{multiset}(\mathcal{P})).$$

Comme, par la méthode symbolique, le terme $(\mathcal{L}\backslash\mathcal{Z}) \times \texttt{multiset}(\mathcal{P})$ donne un terme

$$(k-1)z \exp\left(\sum_{i\geqslant 1} \frac{I(z^i)}{i}\right) = \frac{k-1}{k} I(z)$$

dans la fonction génératrice de $\mathcal{SNT}$, on a :

$$SNT(z) = kz \exp\left(\frac{k-1}{k} \sum_{i\geqslant 1} \frac{I(z^i)}{i}\right) = kz \left(\frac{I(z)}{kz}\right)^{\frac{k-1}{k}} = (kz)^{1/k} I(z)^{\frac{k-1}{k}}.$$

Dès lors,

$$SNT'(z) = \frac{1}{k}(kz)^{\frac{1-k}{k}} I(z)^{\frac{k-1}{k}} + \frac{k-1}{k}(kz)^{1/k} I'(z) I(z)^{\frac{k-1-k}{k}}.$$

Si l'on divise $SNT'(z)$ par $I'(z)$ et si l'on considère la limite quand $z$ tend vers $\eta_k$, le premier terme de la somme tend vers 0. Comme de plus $I(\eta_k) = 1$, nous obtenons

$$\lim_{n\to+\infty} \frac{SNT_{n,k}}{I_{n,k}} = \lim_{z\to\eta_k} \frac{SNT'(z)}{I'(z)} = \frac{k-1}{k}\left(\frac{1}{\mathsf{e}}\right)^{1/k} + \mathcal{O}\left(\frac{1}{k^2}\right) = \left(1 - \frac{1}{k}\right)^2 + \mathcal{O}\left(\frac{1}{k^2}\right) = 1 - \frac{2}{k} + \mathcal{O}\left(\frac{1}{k^2}\right),$$

ce qui conclut la preuve. ■

**Proposition 4.3.8**

*Enfin, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty} \frac{LSNT_{n,k}}{I_{n,k}} = \frac{2 - 1/\mathsf{e}}{k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

**Démonstration :** Si l'on exige qu'au moins deux étiquettes parmi $y_1, \ldots, y_r$ (cf. Figure 4.5) soient égales, cette contrainte supplémentaire rajoute un facteur $\frac{1}{k}$ asymptotiquement quand $k$ tend vers $+\infty$. Dès lors, la famille des non-tautologies moins simples vérifiant cette contrainte est négligeable devant les non-tautologies moins simples ne la vérifiant pas. Cet argument qualitatif pourrait être justifié par le calcul en utilisant la méthode symbolique ainsi que la manipulation des séries génératrices. Nous omettons les détails.

Supposons dès lors que les étiquettes $y_1, \dots, y_r$ sont deux à deux distinctes. On fixe l'entier $r \geqslant 0$ et les variables $\{x, y_1, \dots, y_r\}$. Soit $G(z)$ la fonction génératrice des arbres de la forme des $(T_i)_{i\in\{1,\dots,p\}}$ (cf. Figure 4.5). La racine des $(T_i)_{i\in\{1,\dots,p\}}$ doit être étiquetée par une variable qui n'appartient pas à $\{x, y_1, \dots, y_r\}$ et deux variables sont interdites pour l'étiquetage des nœuds de la première génération : on notera $\mathcal{T}_r$ la famille des arbres vérifiant ces contraintes. La série génératrice de $\mathcal{T}_r$ est donnée par

$$G(z) = (k-r-1)z \exp\left(\sum_{i\geqslant 1} \frac{k-2}{k}\frac{I(z^i)}{i}\right) = (k-r-1)z \exp\left(\sum_{i\geqslant 1} \frac{I(z^i)}{i}\right)^{\frac{k-2}{k}} = (k-r-1)z\left(\frac{I(z)}{kz}\right)^{\frac{k-2}{k}}, \tag{4.2}$$

car $I(z) = kz\sum_{i\geqslant 1}\frac{I(z^i)}{i}$. La famille des non-tautologies moins simples est définie par la spécification suivante :

$$\mathcal{LSNT} = \mathcal{L}\times\mathcal{Z}\times\bigcup_{r\geqslant 0}\binom{k-1}{r}(\mathcal{Z}\times\texttt{multiset}(\mathcal{I}))^r\times(\texttt{multiset}(\mathcal{T}_r)\backslash\{\varnothing\}).$$

Le terme $\mathcal{L}\times\mathcal{Z}$ représente la racine et la feuille de première génération qui a la même étiquette que la racine, et le terme $\binom{k-1}{r}$ compte le nombre de choix possibles pour les variables $y_1, \dots, y_r$. Comme les sous-arbres enracinés en $y_1, \dots, y_r$ sont, de fait, tous différents, le terme $\mathcal{Z}\times\texttt{multiset}(\mathcal{I})$ représente la famille eds arbres dont la racine est déjà étiquetée, ce terme est élevé à la puissance $r$ pour représenter un ensemble (et non un multi-ensemble) de $r$ arbres dont la racine est déjà étiquetée (c'est un ensemble car les racines des arbres de cet ensemble sont distinctes deux à deux). Enfin, l'arbre enraciné en $x$ a pour sous-arbres des arbres de la forme des $(T_i)_{i\in\{1,\dots,p\}}$ (cf. Figure 4.5) mais ne peut être réduit au nœud $x$. Via la méthode symbolique, nous avons donc, comme la fonction génératrice de $\mathcal{Z}\times\texttt{multiset}(\mathcal{I})$ est donnée par $\frac{I(z)}{k}$

$$\begin{aligned} LSNT(z) &= kz^2\sum_{r\geqslant 0}\binom{k-1}{r}\left(\frac{I(z)}{k}\right)^r\left(\exp\left(\sum_{i\geqslant 1}\frac{G(z^i)}{i}\right)-1\right)\\ &= kz^2\sum_{r\geqslant 0}\binom{k-1}{r}\left(\frac{I(z)}{k}\right)^r\left(\exp\left(\sum_{i\geqslant 1}\frac{(k-r-1)z^i}{i}\left(\frac{I(z^i)}{kz^i}\right)^{\frac{k-2}{k}}\right)-1\right). \end{aligned}$$

Notons $Q(z) = \exp\left(\sum_{i\geqslant 2}\frac{(k-r-1)z^i}{i}\left(\frac{I(z^i)}{kz^i}\right)^{\frac{k-2}{k}}\right)$, de sorte que

$$LSNT(z) = kz^2\sum_{r\geqslant 0}\binom{k-1}{r}\left(\frac{I(z)}{k}\right)^r\left(\mathsf{e}^{\frac{I(z)}{kz}}Q(z)-1\right).$$

Dès lors, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{z\to\eta_k}\frac{LSNT'(z)}{I'(z)} \sim k\eta_k^2\sum_{r\geqslant 0}\binom{k-1}{r}\frac{I(\eta)^{r-1}}{k^r}\times \left(\left(\frac{k-2}{k}(k-r-1)\eta_k\left(\frac{I(z)}{k\eta_k}\right)^{\frac{k-2}{k}}+r\right)Q(\eta_k)\mathsf{e}^{(k-r-1)\eta_k\left(\frac{I(\eta_k)}{k\eta_k}\right)^{\frac{k-2}{k}}}-r\right).$$

Il est possible de voir que tronquer cette somme à $\frac{k}{3}$ ne change pas le comportement asymptotique quand $k$ tend vers $+\infty$. Comme $(k-1)_r/k^r \sim 1$ quand $k$ tend vers $+\infty$. De plus, au vu du Lemme 4.3.6, $I(\eta_k) = 1$ et $\eta_k \sim \frac{1}{\mathsf{e}k}$ asymptotiquement quand $k$ tend vers $+\infty$. Dès lors, $Q(\eta_k)\sim 1$, et, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{z\to\eta_k}\frac{LSNT'(z)}{I'(z)} \sim \frac{1}{k\mathsf{e}^2}\sum_{r\geqslant 0}^{\lceil\frac{k}{3}\rceil}\frac{(k-1)_r}{r!k^r}\left(\left(\frac{k-2}{k}(k-r-1)\frac{1/k\mathsf{e}}{(1/\mathsf{e})^{\frac{k-2}{k}}}+r\right)\mathsf{e}^{(k-r-1)\frac{1}{\mathsf{e}k}\mathsf{e}^{\frac{k-2}{k}}}-r\right),$$

ce qui implique

$$\lim_{z\to\eta_k}\frac{LSNT'(z)}{I'(z)}\sim\frac{1}{k\mathtt{e}^2}\sum_{r=0}^{\left[\frac{k}{3}\right]}\frac{1}{r!}\left((r+1)\mathtt{e}-r\right)\sim\frac{2}{k}-\frac{1}{\mathtt{e}k}.$$

■

Il est intéressant de noter que les tautologies simples, non-tautologies simples et non-tautologies moins simples jouent exactement le même rôle que dans le modèle de l'implication classique (cf. [FGGZ07]), et qu'elles suffisent à décrire, asymptotiquement quand $k$ tend vers l'infini, tous les arbres *typiques*.

**Démonstration du Théorème 4.3.4 :** Les Propositions 4.3.5, 4.3.7 et 4.3.8 impliquent que, asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{n\to+\infty}\frac{ST_{n,k}+NST_{n,k}+LNST_{n,k}}{I_{n,k}}=1-\mathcal{O}\left(\frac{1}{k^2}\right),$$

ce qui implique donc

$$\mu_k(\texttt{Vrai})=\lim_{n\to+\infty}\frac{ST_{n,k}}{I_{n,k}}+\mathcal{O}\left(\frac{1}{k^2}\right).$$

■

## 4.4 Probabilité d'une fonction générale

Pour démontrer le Théorème 4.2.5, nous adaptons les idées utilisées dans [FGGG12] et nous définissons les *expansions* d'un arbre général implicatif.

**Définition 4.4.1**

*Étant donné un arbre général implicatif $t$, nous appelons* **expansion** *de $t$ un arbre obtenu en ajoutant à un nœud $\nu$ de $t$ un nouveau sous-arbre $t_e$.*

- *Si cette expansion calcule la même fonction booléenne que $t$, nous dirons que c'est une* **expansion valide**.
- *Si $t_e$ est une tautologie simple, cette expansion est une $T$-**expansion** de $t$. Remarquons que pour tout choix de $\nu$, une $T$-expansion de $t$ est valide.*
- *Si $t_e$ a exactement un sous-arbre de taille 1, si l'unique nœud de ce sous-arbre est étiqueté par $x$, alors cette expansion est une $x$-**prémisse-expansion**. Si $\nu$ a un ancêtre étiqueté par $x$ ou est lui-même étiqueté par $x$, alors cette $x$-prémisse-expansion est valide (cf. Figure 4.6).*
- *Si la racine de $t_e$ est étiquetée par $x$, alors cette expansion est une $x$-**but-expansion**. Si $\nu$ est le parent d'une feuille étiquetée par $x$ ou si $\nu$ a un ancêtre dont l'un des frères est une feuille étiquetée par $x$, alors cette expansion est valide (cf. Figure 4.7).*

Soit $f$ une fonction de $\mathcal{F}_k$ dont le nombre de variables essentielles, $E(f)$ ne dépend pas de $k$. On note $E_t(\mathcal{M}_f)$ (resp. $E_p(\mathcal{M}_f)$, resp. $E_g(\mathcal{M}_f)$) l'ensemble des $T$-expansions (resp. prémisse-expansion, resp. but-expansion) valides d'un arbre minimal de $f$, et $E(\mathcal{M}_f)$ l'union de ces trois ensembles. De plus, pour tout $n\geqslant 1$, on notera $E(\mathcal{M}_f)_{n,k}$ le nombre d'expansions de taille $n$ d'arbres minimaux de $f$.

**Lemme 4.4.2**

*Pour toute fonction booléenne $f$ fixée, il existe une constante $\lambda_f>0$, telle que, asymptotiquement quand $k$ tend vers $+\infty$,*

$$\lim_{n\to+\infty}\frac{E(\mathcal{M}_f)_{n,k}}{I_{n,k}}=\frac{\lambda_f}{k^{L(f)+1}}+\mathcal{O}\left(\frac{1}{k^{L(f)+2}}\right).$$

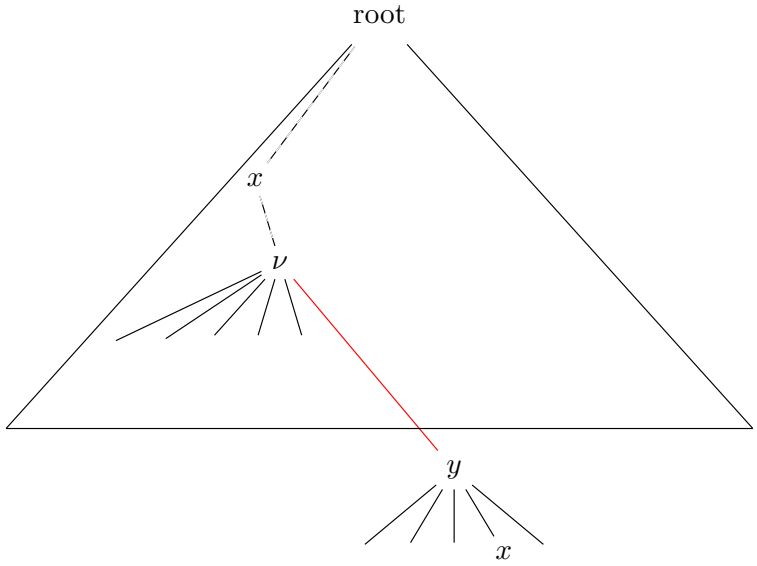



FIGURE 4.6 – *Une $x$-prémisse-expansion valide en $\nu$ (pour tout $y \in \{x_1, \ldots, x_k\}$).*

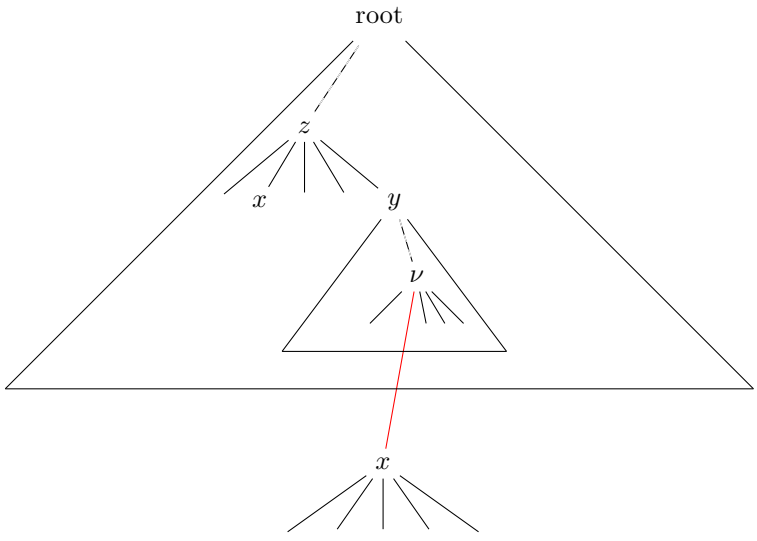



FIGURE 4.7 – *Une $x$-but-expansion valide en $\nu$ (pour tout $y, z \in \{x_1, \ldots, x_k\}$).*

**Démonstration :** Soit $x \in \{x_1, \dots, x_k\}$, on note $\mathcal{E}_x$ l'ensemble des arbres généraux implicatifs dont la racine est étiquetée par $x$, et par $\mathcal{G}_x$ l'ensemble des arbres dont au moins un sous-arbre est une feuille étiquetée par $x$. Toute $x$-prémisse-expansion (resp. $x$-but-expansion) d'un arbre $t$ est obtenue en greffant un arbre de $\mathcal{E}_x$ (resp. $\mathcal{G}_x$) en un nœud $\nu$ de $t$.

Via la méthode symbolique, nous pouvons calculer la fonction génératrice de la famille $\mathcal{E}_x$ pour $x \in \{x_1, \dots, x_k\}$ fixé :

$$E_x(z) = kz^2 \exp \sum_{i \geqslant 1} \frac{I(z^i) - z^i}{i} = z(1-z)I(z) = ST(z),$$

où $ST(z)$, rappelons-le, est la série génératrice des tautologies simples, calculée dans la Section 4.3, Équation (4.1). Nous pouvons en déduire (les détails sont omis) que

$$\lim_{n \to +\infty} \frac{[z^n]E_x(z)}{I_{n,k}} = \frac{1}{\mathsf{e}k} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

De même, la fonction génératrice de la famille $\mathcal{G}_x$ est donnée par

$$G_x(z) = \frac{I(z)}{k},$$

et

$$\lim_{n \to +\infty} \frac{[z^n]G_x(z)}{I_{n,k}} = \frac{1}{n} + \mathcal{O}\left(\frac{1}{k^2}\right).$$

Soit $t$ un arbre. Définissons

$$p(t) = \sum_{x \in \{x_1, \dots, x_k\}} \#\text{sommets de } t \text{ où une } x\text{-prémisse-expansion valide est possible},$$
$$g(t) = \sum_{x \in \{x_1, \dots, x_k\}} \#\text{sommets de } t \text{ où une } x\text{-but-expansion valide est possible}.$$

Dès lors,

$$\begin{aligned}\lim_{n \to +\infty} \frac{E(\mathcal{M}_f)_{n,k}}{I_{n,k}} &= \lim_{n \to +\infty} \sum_{t \in \mathcal{M}_f} \left( p(t) \frac{[z^{n-L(f)}]E_x(z)}{I_{n,k}} + g(t) \frac{[z^{n-L(f)}]G_x(z)}{I_{n,k}} + L(f) \frac{[z^{n-L(f)}]ST(z)}{I_{n,k}} \right) \\ &= \frac{1}{(\mathsf{e}k)^{L(f)}} \left( \lambda_p \frac{1}{\mathsf{e}k} + \lambda_g \frac{1}{k} + L(f) \frac{1}{\mathsf{e}k} \right),\end{aligned}$$

où $\lambda_p = \sum_{t \in \mathcal{M}_f} p(t)$ et $\lambda_g = \sum_{t \in \mathcal{M}_f} g(t)$. Posons

$$\lambda_f = \frac{1}{\mathsf{e}^{L(f)}} \left( \frac{\lambda_p}{\mathsf{e}} + \lambda_g + \frac{L(f)}{\mathsf{e}} \right),$$

alors,

$$\lim_{n \to +\infty} \frac{E(\mathcal{M}_f)_{n,k}}{I_{n,k}} = \frac{\lambda_f}{k^{L(f)+1}} + \mathcal{O}\left(\frac{1}{k^{L(f)+2}}\right).$$

Notons que, dans ce calcul, nous avons compté plusieurs fois les arbres qui peuvent être obtenus à partir de différents arbres minimaux de $f$, via différents types d'expansions. Mais ces arbres vérifient donc deux contraintes, qui chacune introduisent un facteur multiplicatif $1/k$. Cette remarque peut être rendue rigoureuse par le calcul (les détails sont omis ici), et le double-comptage effectué s'avère être négligeable. ■

Le lemme suivant nous assure qu'il suffit de considérer les expansions d'arbres minimaux de $f$ : les expansions d'expansions d'arbres minimaux de $f$ sont négligeables devant les expansions d'arbres minimaux de $f$. Pour tout entier $p \geqslant 1$, nous noterons $E^p(\mathcal{M}_f)$ les arbres obtenus par $p$ expansions successives dans un arbre minimal de $f$.

**Lemme 4.4.3**

*Asymptotiquement quand $k$ tend vers $+\infty$, la fraction limite des arbres obtenus après deux expansions (et non une seule expansion) d'un arbre minimal de $f$ vérifie :*

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

**Démonstration :** Soit $t$ un arbre minimal de $f$. Soit $\hat{t}$ une but-expansion de $t$. On note donc $t_e$ l'arbre rajouté dans $t$ pour obtenir $\hat{t}$. Dès lors, tout arbre obtenu par une expansion en un nœud de $t_e$ est toujours une but-expansion de $t$.

Soit $\hat{t}$ est une prémisse-expansion de $t$ ou une $T$-expansion de $t$. On note donc $t_e$ l'arbre rajouté dans $t$ pour obtenir $\hat{t}$. Si nous faisons une seconde expansion dans $\hat{t}$, si cette expansion est faite en l'unique nœud de $t_e$ dont l'étiquette est la même que celle de la racine de $t_e$ (ou en l'unique nœud réalisant la but-expansion), alors, l'arbre obtenu après les deux extensions n'est peut-être pas dans $E(\mathcal{M}_f)$. Calculons la fonction génératrice de ces arbres obtenus en réalisant une seconde expansion en un nœud précis du premier arbre greffé :

$$N(z) = E_x(z)(E_x(z) + E_g(z) + ST(z)).$$

Dès lors, la proportion limite de cette famille d'arbre est d'ordre $\mathcal{O}\left(\frac{1}{k^{L(f)+2}}\right)$.

Le même raisonnement peut être effectué si la première expansion réalisée est une $T$-expansion. Lorsque nous considérerons des expansions successives dans un arbre minimal de $f$, nous pourrons donc nous restreindre à des expansions qui se feront toutes en des nœuds de l'arbre initial.

La fonction génératrice des arbres obtenus après $k$ expansions dans $k$ nœuds d'un arbre minimal vérifie :

$$G_k(z) \leqslant \frac{L(f)^k}{k!}(E_t(z) + E_p(z) + E_g(z))^k m_f z^{L(f)}.$$

Dès lors, la proportion limite de ces arbres est d'ordre $\mathcal{O}\left(\frac{1}{k^{L(f)+k}}\right)$, pour tout $k \geqslant 2$. ■

Étant donné un arbre $t$, il est possible de simplifier cet arbre en supprimant les arbres qui réalisent une expansion valide. Les arbres qui après simplification appartiennent à $\mathcal{M}_f$ sont donc des éléments de $E(\mathcal{M}_f)$. Malheureusement, certains arbres calculant $f$ ne se simplifient pas jusqu'à obtention d'un arbre minimal de $f$. Nous devons donc contrôler la contribution de ces arbres à $\mu_k(f)$.

**Définition 4.4.4**

*Un* **arbre irréductible** *est un arbre qui ne peut être simplifié par suppression d'expansions valides. Soit $f$ une fonction booléenne. Un* **arbre irréductible de** *$f$ est un arbre irréductible calculant $f$ et qui n'est pas minimal pour $f$. On note $\mathcal{J}_f$ l'ensemble des arbres irréductibles de $f$.*

Un exemple d'arbre irréductible est décrit dans [FGGG12] pour le modèle de l'implication binaire. Ce modèle se traduit immédiatement pour notre modèle puisque les expansions sont une généralisation directe des expansions décrite dans [FGGG12].

La dernière étape de la preuve du Théorème 4.2.5 est de montrer que la contribution des arbres de $E^*(\mathcal{J}_f)$ (i.e. l'ensemble des arbres obtenus par un nombre arbitraire d'expansions successives d'un arbre irréductible de $f$) à la probabilité de $f$ est négligeable, asymptotiquement quand $k$ tend vers $+\infty$. Pour ce faire, nous adapterons des idées de [FGGG12] : nous devons montrer cet analogue du Corollaire 25 de [FGGG12].

**Lemme 4.4.5**

*Soit $\Gamma \subseteq \{x_1, \dots, x_k\}$ un ensemble dont le cardinal ne dépend pas de $k$. Soit $\mathcal{A}_q^p$ l'ensemble des arbres généraux implicatifs qui ont au moins $p$ nœuds de hauteur au plus $q$ et étiquetés par une variable de $\Gamma$. On note $E^*(\mathcal{A}_q^p) = \bigcup_{k\geqslant 0} E^k(\mathcal{A}_q^p)$ et $E^*(z)$ la fonction génératrice de cet ensemble. Alors,*

$$\lim_{n\to+\infty} \frac{[z^n]E^*(z)}{I_{n,k}} = \mathcal{O}\left(\frac{1}{k^p}\right).$$

**Démonstration :** Soit $\mathcal{B}_q^p$ l'ensemble des arbres de hauteur au plus $q$ avec au plus $pq$ nœuds dont au moins $p$ sont étiquetés par une variable de $\Gamma$. Soit $X(\mathcal{B}_q^p)$ l'ensemble des arbres obtenus en rajoutant à chaque nœud interne d'un arbre de $\mathcal{B}_q^p$ un nombre arbitraire de sous-arbres. Notons que $\mathcal{A}_q^p \subseteq X(\mathcal{B}_q^p)$. On notera $\Phi_{q,p}(z)$ la fonction génératrice de l'ensemble $\mathcal{B}_q^p$. Alors,

$$[z^\ell]\Phi_{q,p}(z) \leqslant c_\ell \binom{\ell}{p} \gamma^p k^{\ell-p},$$

où $\gamma$ est le cardinal de $\Gamma$ et $c_\ell$ une constante.

Ajouter un multi-ensemble d'arbres en un nœud d'un arbre de $\mathcal{B}_q^p$ donne une contribution multiplicative $\exp(\sum_{i\geqslant 1} {}^{I(z^i)}/_i) = {}^{P(z)}/_{kz}$ en terme de fonctions génératrices. Dès lors, la fonction génératrice $\Phi_{X_{q,p}}(z)$ de $X(\mathcal{B}_{q,p})$ vérifie

$$[z^n]\Phi_{X_{q,p}}(z) \leqslant [z^n] \sum_{\ell\leqslant pq} c_\ell \binom{\ell}{p} \gamma^p k^{l-p} \frac{I(z)^\ell}{k^\ell} = \frac{1}{k^p} \sum_{\ell\leqslant pq} [z^n] C_\ell I(z)^\ell,$$

où $C_\ell$ est une constante. Comme $I(z)$ et $I^\ell(z)$ ont la même singularité dominante, et comme cette singularité est de type racine carrée pour les deux fonctions, nous en déduisons que $\frac{[z^n]I^\ell(z)}{I_{n,k}}$ converge vers une constante strictement positive quand $n$ tend vers $+\infty$. Dès lors, pour tout $\ell$, il existe une constante $K_\ell$ telle que

$$\frac{[z^n]\Phi_{X_{q,p}}(z)}{I_{n,k}} \leqslant \frac{1}{k^p} \sum_{\ell\leqslant pq} K_\ell,$$

ce qui implique que la proportion limite de l'ensemble $\mathcal{A}_q^p$ est d'ordre $\mathcal{O}\left(\frac{1}{k^p}\right)$, asymptotiquement quand $k$ tend vers $+\infty$. ■

Grâce à ce lemme, les preuves de [FGGG12] peuvent être adaptées littéralement à notre nouveau modèle de l'implication. L'idée est de définir une famille $\mathcal{N}$ comme l'union d'ensembles de la forme $\mathcal{A}_q^p$ qui satisfont tous $p > L(f)+1$. Ainsi, la proportion limite de la famille $\mathcal{N}$ est d'ordre $\mathcal{O}\left(\frac{1}{k^{L(f)+2}}\right)$ et la contribution des arbres de $\mathcal{N}$ à la probabilité de la fonction $f$ est donc négligeable. Il s'agit ensuite de partitionner les arbres irréductibles de $f$ en cinq sous-ensembles selon leur taille et selon le nombre de variables essentielles et inessentielles de $f$ qui apparaissent comme étiquettes de ces arbres. Chacun de ces cinq sous-ensembles sera soit vide, soit inclus dans $\mathcal{N}$. Tous ces arguments sont développés dans [FGGG12], et sont transposables tels quels à notre nouveau modèle. Nous ne détaillons donc pas ici la preuve du lemme suivant, qui clôt la preuve du Théorème 4.2.5 :

**Lemme 4.4.6**

*La proportion limite des expansions d'arbres irréductibles de $f$ vérifie, asymptotiquement quand $k$ tend vers $+\infty$,*

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

**Démonstration du Théorème 4.2.5 :** . L'ensemble des arbres représentant une fonction $f$ fixée est l'union des ensembles suivants : l'ensemble $\mathcal{M}_f$ des arbres minimaux de $f$, l'ensemble $E(\mathcal{M}_f)$ des expansions d'arbres minimaux de $f$, l'ensemble $E^{\geq 2}(\mathcal{M}_f)$ des arbres obtenus après au moins deux expansions d'un arbre minimal de $f$, et l'ensemble $E^*(\mathcal{J}_f)$ des arbres obtenus après un nombre arbitraire d'expansions d'arbres irréductibles de $f$.

Asymptotiquement quand $k$ tend vers $+\infty$, au vu des Lemmes 4.4.3 et 4.4.6, l'ensemble $E(\mathcal{M}_f)$ est le seul à contribuer à la probabilité de $f$. Cette contribution, donnée par le Lemme 4.4.2, conclut donc la preuve. ■

## 4.5 Conclusion

Dans ce modèle des arbres généraux implicatifs, les connecteurs implication qui apparaissaient dans le modèle binaire plan, sans apporter d'information (puisqu'un seul connecteur était autorisé) ont disparu. La taille des arbres généraux reste le nombre de nœuds étiquetés par des variables, tout comme dans le cas binaire plan, et la comparaison entre les deux modèles est donc légitime. Ce chapitre nous aura permis de démontrer que, tout comme dans le cadre et/ou, la prise en compte des propriétés logiques des connecteurs utilisés dans un modèle d'arbres, ici la commutativité des prémisses d'une formule implicative, ne change pas le comportement global en $\frac{1}{k^{L(f)+1}}$ de la distribution induite sur l'espace des fonctions booléennes. Ceci dit, l'étude détaillée de la fonction constante `Vrai` montre que le modèle binaire plan et le modèle général n'induisent pas deux distributions égales sur l'ensemble des fonctions booléennes.

Il est de plus intéressant de noter que les méthodes utilisées dans le cadre binaire et plan sont très facilement transposables au modèle général. Il reste encore à infirmer l'effet Shannon dans ce nouveau modèle : les méthodes utilisées dans [GG10] peuvent-elles être adaptées ?

# Arbres booléens binaires planaires

# Chapitre 5

# Arbre binaire de recherche et arbres aléatoires saturés

## 5.1 Motivations

Comme nous l'avons vu en introduction de cette partie, outre la distribution des arbres de Catalan, issue de la distribution uniforme sur les arbres de taille fixée, Chauvin et al. [CFGG04] ont introduit l'idée d'étudier les distributions issues d'autres arbres aléatoires. Leur première idée a été de considérer l'arbre de Galton-Watson binaire critique qui induit une distribution sur $\mathcal{F}_k$ assez similaire à celle des arbres de Catalan (cf. Section 1.4). L'objectif de ce chapitre est de définir une distribution induite sur $\mathcal{F}_k$ par ce que nous appellerons l'*arbre bourgeonnant*, arbre aléatoire inspiré de l'arbre binaire de recherche (ABR) issu d'une permutation aléatoire (nous faisons référence au livre de Cormen et al. [CLR89] pour une description de cet arbre aléatoire).

Considérons une permutation aléatoire de taille $n$ : cette permutation définit un arbre binaire de recherche aléatoire de taille $n$, dont les nœuds internes sont étiquetés par les entiers de 1 à $n$. Effaçons cet étiquetage et gardons la structure de l'arbre. Remarquons que la **taille** sera exprimé dans ce chapitre en terme de nœuds internes, et non plus en terme de feuilles comme dans les autres modèles. Comme les arbres sont binaires, cela ne fait qu'une différence de 1, mais les calculs s'en verront simplifiés. Étiquetons uniformément cet arbre au hasard, comme Chauvin et al. le font pour l'arbre de Galton-Watson (cf. Section 1.4) : nous obtenons un arbre booléen aléatoire de taille $n$. Cet arbre booléen induit une distribution aléatoire $p_{n,k}$ sur $\mathcal{F}_k$, et nous nous intéressons dans ce chapitre à la distribution asymptotique quand $n$ tend vers $+\infty$, i.e. quand l'arbre bourgeonnant considéré *devient grand*.

L'ABR aléatoire issu d'une permutation aléatoire de taille $n$ a une *forme* très différente de celle d'un arbre de Catalan ou d'un arbre de Galton-Watson : sa hauteur, tout comme son niveau de saturation (i.e. la hauteur de sa feuille la plus proche de la racine) sont d'ordre $\ln n$ quand $n$ tend vers $+\infty$ (cf. Pittel [Pit84] et Devroye [Dev98]), alors que la hauteur d'un arbre de Catalan de taille $n$ est d'ordre $\sqrt{n}$, et son niveau de saturation est d'ordre constant. Il est possible que cette forme *originale* ait une influence sur la distribution induite sur $\mathcal{F}_k$, et nous pouvons espérer un comportement différent de celui de la distribution des arbres de Catalan ou de Galton-Watson.

Nous montrons dans ce chapitre que la distribution induite sur l'ensemble des fonctions booléennes par l'arbre bourgeonnant est *dégénérée*, au sens où, parmi $2^{2^k}$ fonctions, elle ne charge que les fonctions constantes `Vrai` et `Faux`. Nous verrons que ce comportement est très similaire à celui de la distribution induite par les arbres équilibrés (arbres dont toutes les feuilles sont de même génération), étudiée par Fournier et al. [FGG09]. Un arbre équilibré de taille $n$ est de hauteur égale

à son niveau de saturation, et ces deux caractères sont d'ordre $\ln n$, de même que pour les arbres équilibrés. Cette remarque amène a une conjecture reliant le niveau de saturation d'un arbre aléatoire et le caractère *dégénéré* de la distribution induite sur $\mathcal{F}_k$. Nous prouvons en fin de chapitre un méta-théorème résolvant cette conjecture.

Le chapitre est organisé comme suit. La suite de cette introduction est consacrée à la définition de l'arbre bourgeonnant et à l'énoncé des résultats principaux (dans le modèle et/ou et dans le modèle de l'implication). La Section 5.2 est consacrée à la preuve du résultat principal dans le modèle et/ou, et ce via deux méthodes : l'une par combinatoire analytique et la seconde par plongement en temps continu. Dans la section 5.3 nous nous concentrons sur le modèle et/ou et étudions des modèles dans lequel l'étiquetage de l'arbre bourgeonnant est biaisé : que se passe-t-il par exemple si la probabilité d'étiqueter un nœud interne par $\wedge$ devient $q \in [0,1]$ quelconque et non plus $\frac{1}{2}$, et ce pour tout nœud interne ? L'étude de ces extensions permet la comparaison de la distribution de l'arbre bourgeonnant avec celle des arbres équilibrés et donne naissance à une conjecture reliant niveau de saturation d'un arbre aléatoire et dégénérescence de la distribution induite par cet arbre sur $\mathcal{F}_k$. Nous prouvons cette conjecture en Section 5.5, mais étudions tout d'abord en Section 5.4 les tautologies dans le cadre du modèle de l'implication : est-il vrai que presque toute tautologie est simple asymptotiquement quand $k$ tend vers $+\infty$, comme dans les autres modèles ?

### 5.1.1 Définition de l'arbre bourgeonnant

Dans ce chapitre, nous étudierons en détail la distribution induite sur l'espace des fonctions booléennes à $k$ variables $\mathcal{F}_k$ par le modèle de l'arbre bourgeonnant. L'arbre binaire de recherche (ABR) permet en informatique de trier des données en les comparant deux à deux. Lorsque les données que l'on insère dans l'arbre binaire de recherche sont aléatoires (i.e. quand c'est une suite i.i.d. de variables aléatoires), l'arbre sous-jacent non étiqueté croît selon un processus aléatoire facilement décrit : *les feuilles de cet arbre bourgeonnent uniformément* (une description détaillée de l'ABR aléatoire peut être lue dans [CLR89, page 254]). L'arbre bourgeonnant, défini comme suit, est cet arbre aléatoire sous-jacent.

**Définition 5.1.1**

*L'***arbre bourgeonnant** $(\mathcal{T}_i)_{i\in\mathbb{N}}$ *est une suite d'arbres aléatoires définie comme suit :*
- $\mathcal{T}_0$ *est l'arbre qui ne contient qu'un seul nœud, sa racine ;*
- *étant donné* $\mathcal{T}_i$*, choisissons uniformément au hasard une de ses feuilles, transformons-la en un nœud interne et donnons lui deux enfants. L'arbre ainsi obtenu est noté* $\mathcal{T}_{i+1}$*.*

*Pour tout* $n \in \mathbb{N}$*, l'arbre aléatoire* $\mathcal{T}_n$ *est appelé l'***arbre bourgeonnant de taille** $n$*.*

Tout comme dans le cas des arbres de Catalan ou de Galton-Watson, après avoir défini un arbre aléatoire non étiqueté, nous étiquetons aléatoirement et uniformément cet arbre (comme décrit en Section 1.4). Après l'étiquetage uniforme de l'arbre bourgeonnant $\mathcal{T}_n$ de taille $n$ avec $k$ variables, nous obtenons un arbre booléen aléatoire noté $\mathcal{T}_{n,k}$, que nous appellerons aussi arbre bourgeonnant de taille $n$ pour plus de simplicité. Nous noterons $\mathbb{P}_{n,k}$ la loi de $\mathcal{T}_{n,k}$.

**Définition 5.1.2**

*La distribution induite par* $\mathbb{P}_{n,k}$ *sur* $\mathcal{F}_k$ *via* $\Phi$ *(cf. Définition 1.2.5) sera notée* $p_{n,k}$ *: pour toute fonction booléenne* $f$ *de* $\mathcal{F}_k$*,*

$$p_{n,k}(f) = \mathbb{P}(\mathcal{T}_{n,k} \text{ calcule } f).$$

L'objectif de ce chapitre est d'étudier la loi de l'arbre bourgeonnant : la suite des lois $(p_{n,k})_{n\geqslant 0}$ converge-t-elle ? Si oui, que peut-on dire de sa loi limite ? Observe-t-on un effet Shannon (cf. Théo-

rème 1.2.9) ?

## 5.2 La distribution de l'arbre bourgeonnant

### 5.2.1 Existence

Les deux théorèmes ci-dessous, qui seront prouvés dans la suite de cette section établissent l'existence d'une probabilité limite des $(p_{n,k})_{n\geqslant 0}$ quand $n$ tend vers $+\infty$. Cette probabilité limite, notée $p_k$ car elle dépend toujours du nombre de variables utilisées pour l'étiquetage, est appelée la *distribution de l'arbre bourgeonnant*. Ces deux théorèmes montrent par ailleurs que cette distribution limite sur $\mathcal{F}_k$ est étonnamment simple : asymptotiquement quand $n$ tend vers $+\infty$, la fonction calculée par l'arbre bourgeonnant est presque sûrement constante !

**Définition 5.2.1**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$, on notera $\delta_f$ la distribution de probabilité telle que $\delta_f(f) = 1$.*

La norme que nous utiliserons pour démontrer la convergence sur l'espace des distributions de probabilité sur $\mathcal{F}_k$ est la suivante :

**Définition 5.2.2**

*Définissons la norme infinie sur l'espace des mesures signées sur $\mathcal{F}_k$ comme suit :*

$$\|p\|_\infty = \sup_{f\in\mathcal{F}_k} |p(f)|.$$

Dans ce chapitre, $k$ est fixé : choisir une autre norme que la norme infinie ne changerait pas les ordres de grandeur, car toutes les normes sont équivalentes sur $\mathcal{F}_k$.

**Théorème 5.2.3 (Arbre bourgeonnant - Système et/ou)**

*Si l'arbre bourgeonnant est étiqueté uniformément au hasard dans le système logique et/ou, alors,*

$$p_{n,k} \to p_k = \frac{1}{2}\delta_{\texttt{Vrai}} + \frac{1}{2}\delta_{\texttt{Faux}} \text{ quand } n \to +\infty.$$

*De plus,*

$$\|p_{n,k} - p_k\|_\infty = \mathcal{O}\left(\frac{1}{\ln n}\right) \text{ quand } n \to +\infty.$$

Dans la Section 5.3, nous étendrons ce théorème à un cadre un peu plus large : l'arbre bourgeonnant $\mathcal{T}_{n,k}$ sera défini à partir de l'arbre non étiqueté via un étiquetage non uniforme. En effet, que dire par exemple si le connecteur $\wedge$ apparaît en chaque nœud avec une probabilité $q \in [0,1]$ ? Ou si la variable $x_1$ apparaît avec plus forte probabilité que $x_2$ ?

**Théorème 5.2.4 (Arbre bourgeonnant - Modèle de l'implication)**

*Si l'arbre bourgeonnant est étiqueté uniformément au hasard dans le système logique de l'implication, alors,*

$$p_{n,k} \to p_k = \delta_{\texttt{Vrai}} \text{ quand } n \to +\infty.$$

*De plus,*

$$\|p_{n,k} - p_k\|_\infty = \mathcal{O}\left(\frac{1}{\ln n}\right) \text{ quand } n \to +\infty.$$

Il est intéressant de remarquer que la distribution limite diffère quand on change le système logique. Cela vient du fait que la fonction `Faux` ne peut être calculée dans le système de l'implication. En réalité, ces deux théorèmes peuvent être résumés comme suit : la distribution de l'arbre bourgeonnant ne charge que les fonctions constantes qui peuvent être calculées dans le système logique choisi. Une question naturelle se pose : que se passerait-il dans un système logique dans lequel ni `Vrai` ni `Faux` ne peuvent être calculées ? Un tel système est étudié dans le paragraphe 5.3.3, et nous verrons que dans ce cas, la distribution limite ne charge qu'un petit nombre de fonctions appelées *fonctions seuil*.

Les démonstrations des Théorèmes 5.2.3 et 5.2.4 sont presque identiques. C'est pourquoi nous ne développerons que la preuve dans le cadre du système logique et/ou, qui est par ailleurs la plus compliquée des deux, puisqu'elle fait intervenir deux connecteurs logiques au lieu d'un seul dans le cadre de l'implication. De plus, nous présenterons deux preuves du Théorème 5.2.3 : une première preuve par combinatoire analytique, idée la plus naturelle au vu des méthodes utilisées dans le cas des arbres de Catalan ou de l'arbre de Galton-Watson, et une seconde preuve via plongement de l'arbre bourgeonnant en temps continu (méthode parfois appelée *poissonisation*), spécifique à ce modèle de l'arbre bourgeonnant.

Le suite de cette section se concentre donc sur le système logique et/ou, ainsi que la Section 5.3 qui présentera des résultats spécifiques à ce système logique. La Section 5.4 concernera quant à elle exclusivement le système de l'implication dans lequel nous étudierons plus spécifiquement les arbres tautologiques (i.e. représentant la fonction constante `Vrai`).

### 5.2.2 Preuve par combinatoire analytique

**Définition 5.2.5**

*Soit $f$ une fonction booléenne à $k$ variables ($f \in \mathcal{F}_k$), sa* **fonction génératrice** *est donnée par :*

$$\phi_f(z) = \sum_{n=0}^{+\infty} p_{n,k}(f) z^n,$$

*où $p_{n,k}(f)$ est la probabilité que l'arbre bourgeonnant $\mathcal{T}_{n,k}$ de taille $n$ calcule $f$ (cf. Définition 5.1.2).*

Il est bien connu que l'arbre bourgeonnant vérifie la propriété d'auto-similarité suivante :

**Lemme 5.2.6**

*Pour tout $n \geqslant 1$, les deux sous-arbres de $\mathcal{T}_n$, l'arbre bourgeonnant de taille $n$, sont eux-même des arbres bourgeonnants, et le sous-arbre gauche (resp. sous-arbre droit) de $\mathcal{T}_n$, noté $\mathcal{L}_n$ est de taille $m$ avec probabilité $1/n$, pour tout $m \in \{0, \dots, n-1\}$.*

**Démonstration :** Raisonnons par récurrence. Si $n = 1$, alors le sous-arbre gauche est de taille nulle. Soit $n \geqslant 1$. Supposons que la la taille du sous arbre gauche de $\mathcal{T}_n$ suive une loi uniforme sur $\{0, \dots, n-1\}$. Rappelons que l'arbre bourgeonnant de taille $n+1$ est obtenu à partir de $\mathcal{T}_n$, en choisissant une feuille $\lambda$ uniformément au hasard pour la faire pousser en lui donnant deux fils. Soit $i \in \{0, n\}$. Il n'y a que deux façons d'obtenir un sous-arbre gauche de taille $i$ dans $\mathcal{T}_{n+1}$ :

- le sous-arbre gauche est de taille $i-1$ dans $\mathcal{T}_n$ et $\lambda$ a été choisi dans ce sous-arbre, ou
- le sous-arbre gauche est de taille $i$ dans $\mathcal{T}_n$ et $\lambda$ a été choisi dans le sous-arbre droit.

Les calculs suivants se font donc en conditionnant par rapport à la taille de $\mathcal{L}_n$ : pour tout $i \in \{1, \dots, n-$

1} :

$$\mathbb{P}(|\mathcal{L}_{n+1}| = i) = \left(1 - \frac{i+1}{n+1}\right)\mathbb{P}(|\mathcal{L}_n| = i) + \frac{i}{n+1}\mathbb{P}(|\mathcal{L}_n| = i-1)$$
$$= \frac{n-i}{n+1}\cdot\frac{1}{n} + \frac{i}{n+1}\cdot\frac{1}{n} = \frac{1}{n+1}.$$

De plus,

$$\mathbb{P}(|\mathcal{L}_{n+1}| = 0) = \frac{n}{n+1}\mathbb{P}(|\mathcal{L}_n| = 0) = \frac{1}{n+1},$$

et,

$$\mathbb{P}(|\mathcal{L}_{n+1}| = n) = \frac{n}{n+1}\mathbb{P}(|\mathcal{L}_n| = n-1) = \frac{1}{n+1}.$$

Donc, la taille du sous-arbre gauche $\mathcal{L}_{n+1}$ de $\mathcal{T}_{n+1,k}$ suit la loi uniforme sur $\{0,\dots,n\}$. ■

Grâce à cette propriété récursive de l'arbre bourgeonnant, nous pouvons démontrer le lemme suivant, première étape de la preuve du Théorème 5.2.3.

**Lemme 5.2.7**

*Pour toute fonction booléenne $f \in \mathcal{F}_k$, la fonction génératrice $\phi_f(z)$ définie dans la Définition 5.2.5 vérifie :*

$$2\phi_f'(z) = \sum_{g\wedge h=f}\phi_g(z)\phi_h(z) + \sum_{g\vee h=f}\phi_g(z)\phi_h(z),$$

$$\phi_f(0) = \frac{1}{2k}\mathbb{1}_{\{f \text{ lit}\}},$$

*où $\mathbb{1}_{\{f \text{ lit}\}} = 1$ si et seulement si $f$ est une fonction littéral.*

**Démonstration :** L'arbre bourgeonnant de taille $n$, $\mathcal{T}_{n,k}$, calcule $f$ si, et seulement si

- $n = 0$, $f = \alpha$ est une fonction littéral, et la racine (et unique nœud) de $\mathcal{T}_{0,k}$ étiquetée par le littéral $\alpha$ ; ou
- $n \geqslant 1$, le sous arbre-gauche de $\mathcal{T}_{n,k}$ calcule une fonction booléenne $g$, le sous-arbre droit de $\mathcal{T}_{n,k}$ calcule une fonction booléenne $h$, la racine de $\mathcal{T}_{n,k}$ est étiquetée par $\diamond \in \{\wedge, \vee\}$ et $f = g \diamond h$.

Via le Lemme 5.2.6, en conditionnant par rapport à la taille du sous-arbre gauche de $\mathcal{T}_{n+1,k}$, pour tout $n \geqslant 0$,

$$p_{n+1,k}(f) = \frac{1}{2}\sum_{g\wedge h=f}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h) + \frac{1}{2}\sum_{g\vee h=f}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h). \tag{5.1}$$

Il ne reste plus qu'à multiplier (5.1) par $(n+1)z^n$ et à sommer sur $n \geqslant 0$ pour obtenir le Lemme 5.2.7. ■

D'après le Lemme 5.2.7, le vecteur de séries génératrices $(\phi_f(z))_{f\in\mathcal{F}_k}$ vérifie un système de $2^{2^k}$ équations différentielles. Il est intéressant de noter que dans le cadre de l'étude de la distribution des arbres de Catalan (cf. 1.3.1 + [CFGG04]), un tel système d'équations fonctionnelles apparaît également, à la différence que ce système est alors algébrique, et non différentiel. Cette différence est notable puisque le théorème de Drmota-Lalley-Woods [Drm97, Lal93, Woo97] (voir [FS09, page 489]), qui permet de conclure immédiatement à l'existence d'une loi limite dans le cadre d'un système algébrique, n'a pas d'équivalent à ce jour pour des systèmes d'équations différentielles. Une approche différente est donc nécessaire.

Avant tout, il peut être utile de remarquer que l'étiquetage uniforme (cf. Section 1.4) de l'arbre bourgeonnant induit des symétries : entre les deux connecteurs $\wedge$ et $\vee$, ainsi qu'entre une variable et sa négation. Par exemple, la fonction constante `Vrai` est calculée par $\mathcal{T}_{n,k}$ avec la même probabilité

que la fonction constante `Faux`. Plus généralement, une fonction $f$ et sa négation $\bar{f}$ ont même probabilité d'être calculées par $\mathcal{T}_{n,k}$, et ce pour tout $n \in \mathbb{N}$. Ainsi, une fonction booléenne et sa négation ont la même fonction génératrice :

$$\forall f \in \mathcal{F}_k, \forall z \in \mathbb{C}, \phi_f(z) = \phi_{\bar{f}}(z).$$

**Proposition 5.2.8**

*Il existe une constante $\kappa > 0$ telle que, pour tout $x \in [0,1)$,*

$$\frac{1}{2(1-x)} - \frac{1/\kappa}{2(1-x)\left(1/\kappa + \ln\left(\frac{1}{1-x}\right)\right)} \leqslant \phi_T(x) \leqslant \frac{1}{2(1-x)}.$$

**Démonstration :** Posons

$$\phi_S(z) = \sum_{f \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_f(z).$$

Il est à noter que

$$\phi_S + 2\phi_V = \frac{1}{1-z},$$

et, au vu du Lemme 5.2.7,

$$\begin{aligned} 2\phi_S'(z) &= \sum_{f \notin \{\texttt{Vrai},\texttt{Faux}\}} \sum_{g \wedge h = f} \phi_g(z)\phi_h(z) + \sum_{f \notin \{\texttt{Vrai},\texttt{Faux}\}} \sum_{g \vee h = f} \phi_g(z)\phi_h(z) \\ &\leqslant 2 \sum_{g,h \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_g(z)\phi_h(z) \\ &\quad - 2 \sum_{g \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_g(z)\phi_{\bar{g}}(z) \\ &\quad + 4\phi_V(z) \sum_{g \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_g(z), \end{aligned}$$

car une fonction non constante est

- la conjonction ou la disjonction de deux fonctions non constantes (premier terme ci-dessus),
- mais chaque $g \vee \bar{g}$ ou $g \wedge \bar{g}$ est une fonction constante (second terme ci-dessus),
- ou la conjonction d'une fonction non constante avec `Vrai` (ou la disjonction d'une fonction non constante avec `Faux`).

L'équation ci-dessus est une inégalité et non une égalité car le terme de droite inclus, par exemple, les termes associés aux conjonctions de la forme $x \wedge (\bar{x} \wedge y) \equiv$ `Faux` (où $x$ et $y$ sont deux variables distinctes). Au final,

$$\begin{aligned} 2\phi'(z) &\leqslant 2 \sum_{g,h \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_g(z)\phi_h(z) - 2 \sum_{f \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_f(z)\phi_{\bar{f}}(z) + 4\phi_S(z)\phi_V(z) \\ &\leqslant 2\phi_S(z)^2 + 4\phi_S(z)\phi_V(z) - 2\kappa\phi_S(z)^2, \end{aligned}$$

ou $\kappa > 0$ est une constante telle que

$$\sum_{f \notin \{\texttt{Vrai},\texttt{Faux}\}} \phi_f(z)^2 \geqslant 2\kappa\phi_S(z)^2.$$

Une telle constante $\kappa$ existe car la fonction $(x \mapsto x^2)$ est une fonction convexe sur $\mathbb{R}$. Dès lors,

$$\phi_S'(z) + \kappa\phi_S(z)^2 \leqslant 2\phi_S(z)\phi_V(z) + \phi_S(z)^2 = \frac{\phi_S(z)}{1-z}.$$

Soit $Y(z)$ la solution de $Y'(z)+\kappa Y(z)^2 = \frac{Y(z)}{1-z}$ telle que $Y(0) = 1$. Cette équation différentielle est une equation de Bernoulli. Résolvons-là par un changement de variable standard. Soient $W(z) = \frac{1}{Y(z)}$ et $\psi(z) = \frac{1}{\phi_S(z)}$. Ces deux fonctions vérifient

$$W'(z) + \frac{W(z)}{1-z} = \kappa$$

et

$$\psi'(z) + \frac{\psi(z)}{1-z} \geqslant \kappa,$$

avec pour condition initiale $W(0) = \psi(0) = 1$. Dès lors, via le lemme de Grönwall, pour tout $x \in [0,1)$,

$$\psi(x) \geqslant W(x).$$

De plus,

$$W(x) = \kappa(1-x)\left({}^1\!/\!_\kappa + \ln\left(\frac{1}{1-x}\right)\right),$$

ce qui implique, pour tout $x \in [0,1)$,

$$\psi_S(x) \leqslant \frac{{}^1\!/\!_\kappa}{(1-x)\left({}^1\!/\!_\kappa + \ln\left(\frac{1}{1-x}\right)\right)}.$$

■

Comme précisé, ces résultats ne sont démontrés pour $x \in [0,1)$. Cette restriction nous empêche donc d'appliquer le lemme de transfert de Flajolet et Odlyzko [FO90] (voir aussi [FS09, page 389]) qui nécessite une analyse asymptotique des fonctions génératrices dans $\mathbb{C}$. Cependant, la Proposition 5.2.8 nous permet d'appliquer un théorème taubérien (voir [Har49, page 155]) et d'obtenir ainsi le comportement asymptotique des sommes partielles des coefficients $p_{n,k}(\texttt{Vrai})$ de la fonction génératrice $\phi_V(z)$. Asymptotiquement quand $n$ tend vers $+\infty$,

$$\sum_{i=1}^{n}\left(\frac{1}{2} - p_{i,k}(\texttt{Vrai})\right) = \sum_{i=1}^{n}\left(\frac{1}{2} - p_{i,k}(\texttt{Faux})\right) = \mathcal{O}\left(\frac{n}{\ln n}\right). \tag{5.2}$$

Nous n'avons plus qu'à en déduire le comportement asymptotique des coefficients eux-mêmes. Au vu de l'Équation (5.1),

$$\begin{aligned} p_{n+1,k}(\texttt{Vrai}) &= \frac{1}{2}\sum_{g\wedge h=\texttt{Vrai}}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h) + \frac{1}{2}\sum_{g\vee h=\texttt{Vrai}}\sum_{i=0}^{n}\frac{1}{n+1}p_{i,k}(g)p_{n-i,k}(h) \\ &\geqslant \frac{1}{2(n+1)}\sum_{i=1}^{n}p_{i,k}(\texttt{Vrai})p_{n-i,k}(\texttt{Vrai}) \\ &\quad + \frac{1}{n+1}\sum_{i=1}^{n}p_{i,k}(\texttt{Vrai})(1-p_{n-i,k}(\texttt{Vrai})) + \frac{1}{2(n+1)}\sum_{i=1}^{n}p_{i,k}(\texttt{Vrai})p_{n-i,k}(\texttt{Vrai}) \end{aligned}$$

où le premier terme de cette somme est la probabilité de calculer `Vrai` quand la racine de $\mathcal{T}_{n,k}$ est étiquetée par le connecteur $\wedge$, et où la somme des deux autres termes est plus petite que cette même probabilité quand la racine de $\mathcal{T}_{n,k}$ est étiquetée par le connecteur $\vee$. L'Équation (5.2) implique donc que :

$$\frac{1}{2} - p_{n+1,k}(\texttt{Vrai}) \leqslant \frac{{}^1\!/\!_2}{n+1} + \frac{1}{n+1}\sum_{i=1}^{n}\left(\frac{1}{2} - p_{i,k}(\texttt{Vrai})\right) = \mathcal{O}\left(\frac{1}{\ln n}\right), \text{ quand } n \to +\infty.$$

### 5.2.3 Preuve probabiliste

Cette seconde preuve du Théorème 5.2.3 utilise le plongement en temps continu, ou *poissonisation*, de l'arbre binaire de recherche, décrit par exemple dans [Pit84]. Plonger des processus discrets en temps continu est assez standard en probabilités. Cette méthode est notamment utilisée pour étudier les urnes de Pólya, comme introduit dans le livre de Athreya et Ney [AN72], et comme détaillé dans la Partie II du présent mémoire. L'analogue en temps continu de l'arbre bourgeonnant est appelé *arbre de Yule* et est défini comme suit, à l'aide *d'horloges exponentielles* de paramètre 1.

**Définition 5.2.9 (Pittel [Pit84])**

*Un* **arbre de Yule** *est un processus à temps continu d'arbres binaires $(\mathcal{Y}_t)_{t\geqslant 0}$ défini comme suit :*

- *$\mathcal{Y}_0$ est l'arbre de taille zéro ;*
- *chaque feuille de $\mathcal{Y}_t$ se transforme en nœud interne parent de deux feuilles au bout d'un temps aléatoire de loi exponentielle de paramètre 1, et ce indépendamment des autres feuilles.*

On dira que chaque feuille de l'arbre est équipée d'une *horloge exponentielle* de paramètre 1, et que toutes les horloges exponentielles de l'arbre sont indépendantes. Grâce aux propriétés de la loi exponentielle qui a été choisie pour ces horloges, notamment sa propriété *d'absence de mémoire*, il est aisé de relier l'arbre de Yule et l'arbre bourgeonnant par ce que nous appellerons une *connexion*. Pour cela, nous avons besoin de définir la suite $(\tau_n)_{n\geqslant 0}$, suite des dates aléatoires auxquelles l'arbre de Yule a *poussé*.

**Définition 5.2.10**

*On note $\tau_n$, et on appelle* $\boldsymbol{n^{\text{ème}}}$ **temps de saut**, *la date à laquelle l'arbre de Yule atteint la taille $n$ :*

$$\tau_n = \inf\{t \geqslant 0, |\mathcal{Y}_t| = n\}.$$

Si l'on considère le processus de Yule pris à ces dates $(\tau_n)_{n\geqslant 0}$, on peut reconnaître le processus de l'arbre bourgeonnant. Le choix de la loi exponentielle nous assure en effet que l'horloge qui sonne en premier parmi $n$ horloges est choisie uniformément parmi ces $n$ horloges. Autrement dit, pour passer de $\mathcal{Y}_{\tau_n}$ à $\mathcal{Y}_{\tau_{n+1}}$, nous avons choisi une feuille au hasard et l'avons transformée en nœud interne. Nous reconnaissons là le processus de croissance de l'arbre bourgeonnant. Nous avons donc la connexion suivante :

$$(\mathcal{Y}_{\tau_n})_{n\geqslant 0} \overset{(loi)}{=} (\mathcal{T}_n)_{n\geqslant 0}, \tag{5.3}$$

où, de plus, la suite des $(\tau_n)_{n\geqslant 0}$ est indépendante de la suite $(\mathcal{Y}_{\tau_n})_{n\geqslant 0}$.

Nous avons défini aisément le plongement en temps continu de l'arbre bourgeonnant non étiqueté ; occupons-nous maintenant de son étiquetage.

**Définition 5.2.11**

*Un* **arbre de Yule étiqueté** *est un processus en temps continu $(\mathcal{Y}_{t,k})_{t\geqslant 0}$ d'arbres binaires étiquetés, défini comme suit :*

- *considérons l'arbre de Yule $\mathcal{Y}_t$ non étiqueté au temps $t$,*
- *étiquetons cet arbre uniformément (cf. 1.4).*

*L'arbre aléatoire obtenu est noté $\mathcal{Y}_{t,k}$, et on notera sa loi $\mathtt{P}_{t,k}$.*

De manière naturelle, l'arbre du Yule étiqueté au temps $t$ représente une fonction booléenne aléatoire dont on notera la loi $\mathfrak{p}_{t,k}$ :

| | Arbre non étiqueté | Arbre étiqueté | Loi de l'arbre étiqueté | Loi induite sur $\mathcal{F}_k$ | Distribution asymptotique |
|---|---|---|---|---|---|
| Temps discret | $\mathcal{T}_n$ | $\mathcal{T}_{n,k}$ | $\mathbb{P}_{n,k}$ | $p_{n,k}$ | $p_k$ |
| Temps continu | $\mathcal{Y}_t$ | $\mathcal{Y}_{t,k}$ | $\mathtt{P}_{t,k}$ | $\mathfrak{p}_{t,k}$ | |

FIGURE 5.1 – *Ce tableau est un résumé des différentes notations issues des deux preuves du Théorème 5.2.3, il pourra être utile lors de leur lecture.*

**Définition 5.2.12**

*Pour toute fonction booléenne $f$,*

$$\mathfrak{p}_{t,k}(f) = \mathbb{P}(\mathcal{Y}_{t,k} \text{ computes } f).$$

La connexion entre l'arbre de Yule et l'arbre bourgeonnant (cf. Équation(5.3)) suggère une connexion entre les lois $p_{n,k}$ et $\mathfrak{p}_{t,k}$ induites sur l'ensemble des fonctions booléennes. C'est pourquoi la suite de cette partie sera consacrée à l'étude de $\mathfrak{p}_{t,k}$, quand $t$ tend vers $+\infty$, puis au transfert des résultats obtenus en temps continu vers le modèle en temps discret. Le modèle en temps continu s'avère en effet plus simple à étudier que le modèle en temps discret, tout simplement grâce à l'indépendance que le plongement introduit : dans l'arbre de Yule, les deux sous-arbres de chaque nœud interne sont indépendants.

Notre objectif est donc de montrer que la distribution $\mathfrak{p}_{t,k}$ converge, quand $t$ tend vers $+\infty$ vers $p_k = \frac{1}{2}\delta_{\mathtt{Vrai}} + \frac{1}{2}\delta_{\mathtt{Faux}}$. Notre stratégie s'inspire d'un article sur des arbres booléens *équilibrés* (ou *triangulaires*) [FGG09]. Nous détaillerons dans la suite (cf. Section 5.3, Section 5.5) la ressemblance entre le modèle de l'arbre bourgeonnant et celui des arbres équilibrés, et montrerons comment ces deux modèles peuvent être unifiés en un résultat plus général. Dans la présente preuve, nous fixons deux affectations des variables, $a, b \in \{0,1\}^k$, et calculons la probabilité que leurs images par la fonction booléenne aléatoire calculée par l'arbre de Yule soient distinctes. Nous montrerons que cette probabilité tend vers 0 quand $t$ tend vers $+\infty$ pour tout choix de $a$ et $b$, et donc que $\mathfrak{p}_{t,k}$ converge bien vers $p_k$.

Soient $a = (a_1, \ldots, a_k)$ et $b = (b_1, \ldots, b_k)$ deux éléments distincts de $\{0,1\}^k$. Soient $\alpha$ et $\beta$ deux éléments de $\{0,1\}$ et $f \in \mathcal{F}_k$ une fonction booléenne aléatoire de loi $\mathfrak{p}_{t,k}$. Pour tout $t \geqslant 0$, on note

$$\mathfrak{p}_{t,k}^{\alpha\beta}(a,b) = \mathfrak{p}_{t,k}\left(f(a) = \alpha \text{ et } f(b) = \beta\right).$$

De même, pour tout entier $n \geqslant 0$, nous noterons

$$\mathbb{P}_n^{\alpha\beta}(a,b) = p_{n,k}\left(f(a) = \alpha \text{ et } f(b) = \beta\right).$$

**Fait :** Tout comme dans le cas discret, au regard des symétries du modèle, et notamment des symétries de l'étiquetage uniforme, la probabilité que l'arbre de Yule au temps $t$ calcule $f$ est égale à la probabilité qu'il calcule sa négation $\bar{f}$. Nous avons donc les relations suivantes : pour tout $a, b \in \{0,1\}^k$,

$$\begin{aligned} \mathfrak{p}_{t,k}^{01}(a,b) &= \mathfrak{p}_{t,k}^{10}(a,b) \\ \mathfrak{p}_{t,k}^{00}(a,b) &= \mathfrak{p}_{t,k}^{11}(a,b) \\ \mathfrak{p}_{t,k}^{10}(a,b) + \mathfrak{p}_{t,k}^{00}(a,b) &= 1/2. \end{aligned} \tag{5.4}$$

Lorsque la précision n'est pas nécessaire, nous noterons $\mathfrak{p}_t^{\alpha\beta}$ eu lieu de $\mathfrak{p}_{t,k}^{\alpha\beta}(a,b)$. Calculons $\mathfrak{p}_t^{10}$ en conditionnant par la date à laquelle l'horloge de la racine se déclenche, i.e. au premier temps de saut (cf. Définition 5.2.10). Cette date suit par définition une loi exponentielle de paramètre 1, et

$$\mathfrak{p}_t^{10} = \sum_{i=1}^{k} \frac{\mathsf{e}^{-t}}{2k} \mathbb{1}_{\{a_i \neq b_i\}} + \frac{1}{2}\int_0^t \left(\mathfrak{p}_{t-s}^{11}\mathfrak{p}_{t-s}^{10} + \mathfrak{p}_{t-s}^{10}(\mathfrak{p}_{t-s}^{11} + \mathfrak{p}_{t-s}^{10}) + \mathfrak{p}_{t-s}^{10}(\mathfrak{p}_{t-s}^{00} + \mathfrak{p}_{t-s}^{10}) + \mathfrak{p}_{t-s}^{00}\mathfrak{p}_{t-s}^{10}\right) \mathsf{e}^{-s} ds.$$

Le premier terme de cette somme est égal à la probabilité que l'horloge de la racine n'ait pas sonné avant la date $t$ et $f(a) = 1$ et $f(b) = 0$, le second terme est la probabilité que l'horloge ait sonné avant la date $t$ et $f(a) = 1$ et $f(b) = 0$. Si l'horloge a sonné avant la date $t$, les valeurs de $f(a)$ et $f(b)$ dépendent alors directement de l'étiquette de la racine ($\wedge$ ou $\vee$) et des valeurs en $a$ et $b$ des deux fonctions calculées par les deux sous-arbres de l'arbre de Yule au temps $t$ ; c'est ce que représente l'intégrande du second terme. Cette équation peut être simplifiée en utilisant les relations de symétries (cf. Équation (5.4)) :

$$\mathfrak{p}_t^{10} = \frac{\mathsf{e}^{-t}}{2k} c_{a,b} + \mathsf{e}^{-t}\int_0^t \left(\mathfrak{p}_s^{10} - (\mathfrak{p}_s^{10})^2\right) \mathsf{e}^{s} ds$$

où $c_{a,b} = \sum_{i=1}^{k} \mathbb{1}_{\{a_i \neq b_i\}}$ est une constante qui ne dépend que de $a$ et $b$. Notons, pour simplifier, $\pi_{a,b}(t) = \mathfrak{p}_t^{10}(a,b)$, et étudions-la en tant que fonction de $t$ :

$$\mathsf{e}^{t}\pi_{a,b}(t) = \frac{c_{a,b}}{2k} + \int_0^t \left(\pi_{a,b}(s) - \pi_{a,b}(s)^2\right) \mathsf{e}^{s} ds. \tag{5.5}$$

Grâce à cette équation, nous pouvons désormais démontrer la proposition suivante :

**Proposition 5.2.13**

- *Si $a \neq b$ alors $\pi_{a,b}(t) = \frac{1}{t+t_{a,b}}$ avec $t_{a,b} = \frac{2k}{c_{a,b}}$.*
- *Si $a = b$, alors $\pi_{a,a}(t)$ est la fonction constante égale à zéro.*

*Donc, pour tout $a, b \in \{0,1\}^k$, $\pi_{a,b}(t) = \mathfrak{p}_{t,k}^{10}(a,b)$ tend vers 0 quand $t \to +\infty$.*

**Démonstration :** L'Équation (5.5) implique que $\pi_{a,b}$ est dérivable et qu'elle vérifie l'équation différentielle $\pi'_{a,b} + \pi_{a,b}^2 = 0$.

- Supposons $a \neq b$. Alors il existe $i_0 \in \{1 \ldots, k\}$ tel que $a_{i_0} \neq b_{i_0}$, et $c_{a,b} \geqslant \mathbb{1}_{\{a_{i_0} \neq b_{i_0}\}} = 1$ et $\pi_{a,b}(0) = \frac{c_{a,b}}{2k} > 0$. Dès lors, $\pi_{a,b}(t) = \frac{1}{t+t_{a,b}}$ avec $t_{a,b} = \frac{2k}{c_{a,b}}$.
- Supposons $a = b$, alors $\pi_{a,a}(0) = 0$ et $\pi_{a,a}(t) = 0$ pour tout $t \geqslant 0$ car une affectation $a$ des variables $(x_1, \ldots, x_k)$ ne peut avoir deux images différentes par $f$. ∎

Enfin, pour obtenir la convergence de $\mathfrak{p}_{t,k}$ quand $t$ tend vers $+\infty$, nous n'avons plus qu'à noter que

$$\begin{aligned}
\mathfrak{p}_{t,k}(\mathcal{F}_k \backslash \{\texttt{Vrai}, \texttt{Faux}\}) &\leqslant \sum_{(a,b), a\neq b} \mathfrak{p}_{t,k}(f(a) = 1 \text{ et } f(b) = 0) \\
&\leqslant 2^k(2^k - 1) \sup_{(a,b), a\neq b} \mathfrak{p}_{t,k}^{10}(a,b) \\
&\leqslant \frac{2^k(2^k-1)}{t + t_{min}},
\end{aligned}$$

où $t_{min} = \inf_{(a,b),a\neq b}\{\frac{2k}{c_{a,b}}\} = 2k$. Dès lors, pour toute fonction booléenne non constante, $f \notin \{\mathtt{Vrai}, \mathtt{Faux}\}$, $\mathfrak{p}_{t,k}(f)$ converge vers zéro quand $t$ tend vers $+\infty$, ce qui implique directement que $\mathfrak{p}_{t,k}$ converge quand $t$ tend vers $+\infty$ vers la distribution $p_k = \frac{1}{2}\delta_{\mathtt{Vrai}} + \frac{1}{2}\delta_{\mathtt{Faux}}$. De plus, nous connaissons la vitesse de convergence qui est d'ordre $1/t$ :

$$\|\mathfrak{p}_{t,k} - p_k\|_\infty = \sup_{f\in\mathcal{F}_k} |\mathfrak{p}_{t,k}(f) - p_k(f)| \leqslant \frac{2^k(2^k-1)}{t+2k}, \qquad \text{pour tout } t \geqslant 0. \tag{5.6}$$

Il ne nous reste plus qu'à traduire ce résultat en temps discret via la connexion (5.3), et ainsi démontrer le Théorème 5.2.3 dans son intégralité.

**Proposition 5.2.14**

*Asymptotiquement quand $n$ tend vers $+\infty$,*

$$\|p_{n,k} - p_k\|_\infty = \mathcal{O}\left(\frac{1}{\ln n}\right).$$

La démonstration de cette proposition s'appuie sur le résultat suivant, démontré par exemple dans le livre d'Athreya et Ney [AN72].

**Proposition 5.2.15 ([AN72])**

*Pour tout $n \geqslant 1$, pour tout $t \geqslant 0$, les* temps de saut *du processus de l'arbre de Yule (cf. Définition 5.2.10) vérifient*

$$\mathbb{P}(\tau_n \leqslant t) = \mathbb{P}(n(t) \geqslant n) = (1-\mathsf{e}^{-t})^n,$$

*où $n(t)$ est le nombre de feuille de l'arbre bourgeonnant au temps $t$.*

**Démonstration de la Proposition 5.2.14 :** Pour tout $t \geqslant 0$, nous noterons $\eta_t$ la fonction booléenne aléatoire calculée par l'arbre de Yule étiqueté $\mathcal{Y}_{t,k}$. Rappelons par ailleurs que l'on note $f_n$ la fonction booléenne aléatoire calculée par l'arbre bourgeonnant de taille $n$.

Rappelons que $n(t)$ est le nombre de nœuds internes de l'arbre $\mathcal{Y}_t$, sa loi est décrite par la Proposition 5.2.15. Pour tout choix de $a, b \in \{0,1\}^k$,

$$\begin{aligned}\mathfrak{p}^{10}_{t,k}(a,b) &= \sum_{n\geqslant 0} \mathbb{P}(\eta_t(a)=1 \text{ et } \eta_t(b)=0 \,|\, n(t)=n)\mathbb{P}(n(t)=n)\\ &= \sum_{n\geqslant 0} \mathbb{P}(\eta_{\tau_n}(a)=1 \text{ et } \eta_{\tau_n}(b)=0 \,|\, n(t)=n)\mathbb{P}(n(t)=n).\end{aligned}$$

Le plongement en temps continu nous assure que les variables aléatoire $n(t)$ et $\eta_{\tau_n}$ sont indépendantes pour tout $n \geqslant 0$. Comme $n(t)$ suit une loi géométrique de paramètre $\mathsf{e}^{-t}$, nous obtenons :

$$\begin{aligned}\mathfrak{p}^{10}_{t,k}(a,b) &= \sum_{n\geqslant 0} \mathbb{P}(\eta_{\tau_n}(a)=1 \text{ et } \eta_{\tau_n}(b)=0)\mathsf{e}^{-t}(1-\mathsf{e}^{-t})^n\\ &= \mathsf{e}^{-t}\sum_{n\geqslant 0} \mathbb{P}(f_n(a)=1 \text{ et } f_n(b)=0)(1-\mathsf{e}^{-t})^n\\ &= \mathsf{e}^{-t}\sum_{n\geqslant 0} \mathbb{P}^{10}_{n,k}(a,b)(1-\mathsf{e}^{-t})^n.\end{aligned}$$

Pour tout $t \geqslant 0$, au vu de la Proposition 5.2.13,

$$\mathsf{e}^{-t}\sum_{n\geqslant 0} \mathbb{P}^{10}_{n}(a,b)(1-\mathsf{e}^{-t})^n = \frac{1}{t+t_{a,b}}.$$

Soit $z = 1 - \mathrm{e}^{-t}$ : pour tout $z \in [0,1[$, on note $\varphi_{a,b}(z) = \sum_{n\geqslant 0} \mathbb{P}_n^{10}(a,b) z^n$. Dès lors,

$$\varphi_{a,b}(z) = \frac{1}{(1-z)\left(t_{a,b} + \ln \frac{1}{1-z}\right)}. \tag{5.7}$$

Grâce à un théorème taubérien (voir par exemple [FS09, page 435]), nous en déduisons que, asymptotiquement quand $n$ tend vers $+\infty$, la somme partielle des $n$ premiers coefficients de la série génératrice $\varphi_{a,b}$ est d'ordre $\frac{n}{\ln n}$ : pour tout $a \neq b \in \{0,1\}^k$, asymptotiquement quand $n$ tend vers $+\infty$,

$$\sum_{m=0}^{n} \mathbb{P}_m^{10}(a,b) \sim \frac{n}{\ln n}.$$

Dès lors, pour tout $n$ assez grand,

$$\sum_{m=0}^{n} \mathbb{P}_m^{10}(a,b) \leqslant 2\frac{n}{\ln n},$$

et

$$\sum_{m=0}^{n} p_{m,k}(f \notin \{\mathtt{Vrai}, \mathtt{Faux}\}) \leqslant \sum_{a\neq b\in\{0,1\}^k} \sum_{m=0}^{n} \mathbb{P}_m^{10}(a,b) \leqslant 2^k(2^k-1)\cdot 2\frac{n}{\ln n}.$$

Nous en déduisons

$$\sum_{m=0}^{n} p_{m,k}(\mathtt{Vrai}) = \sum_{m=0}^{n} \frac{1}{2}(1 - p_{m,k}(f \notin \{\mathtt{Vrai}, \mathtt{Faux}\})) \geqslant \frac{n}{2} - 2^k(2^k-1)\frac{n}{\ln n}.$$

Finalement, au vu de l'Équation (5.1) appliquée à $f \equiv \mathtt{Vrai}$,

$$\begin{aligned}
p_{n+1,k}(\mathtt{Vrai}) &= \frac{1}{2} \sum_{g\wedge h = \mathtt{Vrai}} \sum_{i=0}^{n} \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) + \frac{1}{2} \sum_{g\vee h = \mathtt{Vrai}} \sum_{i=0}^{n} \frac{1}{n+1} p_{i,k}(g) p_{n-i,k}(h) \\
&\geqslant \frac{1}{2(n+1)} \sum_{i=0}^{n} p_{i,k}(\mathtt{Vrai}) p_{n-i,k}(\mathtt{Vrai}) \\
&+ \frac{1}{n+1} \sum_{i=0}^{n} p_{i,k}(\mathtt{Vrai})(1 - p_{n-i,k}(\mathtt{Vrai})) + \frac{1}{2(n+1)} \sum_{i=0}^{n} p_{i,k}(\mathtt{Vrai}) p_{n-i,k}(\mathtt{Vrai}) \\
&= \frac{1}{n+1} \sum_{i=0}^{n} p_{i,k}(\mathtt{Vrai}) \\
&\geqslant \frac{n}{n+1}\left(\frac{1}{2} - 2^k(2^k-1)\frac{1}{\ln n}\right).
\end{aligned}$$

Dès lors,

$$\frac{1}{2} - p_{n+1,k}(\mathtt{Vrai}) \leqslant \frac{1}{2} - \frac{n}{n+1}\left(\frac{1}{2} - 2^k(2^k-1)\frac{1}{\ln n}\right) = \mathcal{O}\left(\frac{1}{\ln n}\right),$$

ce qui conclut la preuve de la Proposition 5.2.15. ■

Le Théorème 5.2.3 est donc démontré dans son intégralité, grâce à cette approche en temps continu. Cette dernière preuve s'avère être plus efficace que l'approche par combinatoire analytique, c'est pourquoi la méthode de plongement en temps continu sera privilégiée dans la suite de l'étude de la distribution de l'arbre bourgeonnant (cf. Section 5.3).

## 5.3 Extensions dans le système logique et/ou

Dans cette section, nous nous intéressons exclusivement au système logique et/ou et nous modifions l'étiquetage aléatoire de l'arbre bourgeonnant, étiquetage qui était jusqu'à maintenant uniforme

(cf. Section 1.4). Nous modifierons tout d'abord l'étiquetage des feuilles de l'arbre bourgeonnant : comment se comporte la distribution de l'arbre bourgeonnant si, par exemple, la variable $x_1$ apparaît plus souvent que la variable $x_2$ ? Puis nous jouerons aussi sur la distribution des connecteurs $\wedge$ and $\vee$ en chaque nœud interne. Tous ces résultats, et leurs preuves, seront à regarder en miroir avec ceux établis dans l'article de Fournier et al. [FGG09] sur les arbres équilibrés. La ressemblance entre ces deux modèles sera à l'origine d'un théorème plus général dans la Section 5.5.

### 5.3.1 Biaiser la loi des littéraux

Définissons un nouvel étiquetage aléatoire :

**Définition 5.3.1 (Étiquetage aléatoire - variables non symétriques)**

*Étant donné un arbre binaire planaire, étiquetons-le au hasard comme suit :*
- *chaque nœud interne est étiqueté par $\wedge$ avec probabilité $1/2$ ou par $\vee$ avec probabilité $1/2$,*
- *on définit $\nu$ une distribution de probabilité sur l'ensemble des littéraux $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ telle que pour tout $i \in \{1, \dots, k\}$, $\nu(x_i) = \nu(\bar{x}_i)$ ; chaque feuille de l'arbre est étiquetée par un littéral choisi selon cette loi de probabilité ;*
- *chaque nœud, interne ou externe, est étiqueté indépendamment des autres.*

Considérons l'arbre bourgeonnant étiqueté selon ce nouvel étiquetage aléatoire. Ce nouvel arbre booléen aléatoire induit une distribution sur l'ensemble des fonctions booléennes, distribution que nous noterons aussi $p_{n,k}$, même si ce n'est pas la même que celle étudiée précédemment. Cependant, il est facile de voir, via les preuves développées ci-dessus, que cette nouvelle distribution $p_{n,k}$ converge elle aussi vers la distribution asymptotique $p_k = \frac{1}{2}\delta_{\texttt{Vrai}} + \frac{1}{2}\delta_{\texttt{Faux}}$. Cela vient du fait que la symétrie entre une fonction et sa négation est conservée puisque la symétrie entre $\wedge$ et $\vee$ l'est, ainsi que celle entre tout littéral et sa négation. Plus précisément, dans la preuve par plongement en temps continu (cf. Sous-section 5.2.3), seule la constante $c_{a,b}$ est modifiée dans l'Équation (5.5), constante qui disparaît ensuite par dérivation, et qui n'influe donc pas sur le résultat final. Cette nouvelle loi vérifie donc aussi le Théorème 5.2.3.

### 5.3.2 Biaiser la loi des connecteurs

En plus du biais introduit sur la distribution des littéraux sur les feuilles de l'arbre, biaisons la distribution des connecteurs logiques :

**Définition 5.3.2 (Étiquetage aléatoire - variables et connecteurs non symétriques)**

*Étant donné un arbre binaire planaire, étiquetons-le au hasard comme suit :*
- *chaque nœud interne est étiqueté par $\wedge$ avec probabilité $\varpi$ ou par $\vee$ avec probabilité $1 - \varpi$,*
- *on définit $\nu$ une distribution de probabilité sur l'ensemble des littéraux $\{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$ telle que pour tout $i \in \{1, \dots, k\}$, $\nu(x_i) = \nu(\bar{x}_i)$ ; chaque feuille de l'arbre est étiquetée par un littéral choisi selon cette loi de probabilité ;*
- *chaque nœud, interne ou externe, est étiqueté indépendamment des autres.*

L'arbre bourgeonnant étiqueté aléatoirement selon ce nouvel étiquetage induit donc une nouvelle loi sur l'ensemble des fonctions booléennes, toujours notée $p_{n,k}$ pour plus de clarté. Le cas $\varpi = 1/2$ est déjà traité dans la Sous-section 5.3.1, et le résultat suivant généralise le Théorème 5.2.3 au cas $\varpi \neq 1/2$ :

**Théorème 5.3.3**

*La suite de distribution $(p_{n,k})_{n\geqslant 0}$ définie via l'arbre bourgeonnant étiqueté aléatoirement selon la Définition 5.3.2 vérifie :*

- *si $\varpi > 1/2$, alors $p_{n,k} \to \delta_{\texttt{Faux}}$ quand $n$ tend vers $+\infty$ ;*
- *si $\varpi < 1/2$, alors $p_{n,k} \to \delta_{\texttt{Vrai}}$ quand $n$ tend vers $+\infty$.*

*De plus, la vitesse de convergence est d'ordre $\mathcal{O}\left(\frac{1}{n^{|2\varpi-1|}}\right)$ dans les deux cas.*

Quelle que soit la valeur du paramètre $\varpi$, la distribution asymptotique ne charge que les fonctions constantes. Si la proportion de $\wedge$ (resp. de $\vee$) est plus grande, alors la seule fonction chargée est la fonction constante égale à `Faux` (resp. `Vrai`). Le cas $\varpi = 1/2$ s'avère donc être un cas critique. Il est d'ailleurs intéressant de noter que la vitesse de convergence est bien plus lente (logarithmique) lorsque nous sommes dans le cas critique $\varpi = 1/2$ (cf. Théorème 5.2.3) alors qu'elle devient polynomiale dans les cas non critiques $\varpi \neq 1/2$.

**Démonstration :** Les cas $\varpi < 1/2$ et $\varpi > 1/2$ sont symétriques et donnent donc lieu à deux preuves presque identiques. C'est pourquoi nous ne traitons ici que le cas $\varpi > 1/2$, donc le cas où la proportion de connecteurs $\wedge$ est plus importante que celle de $\vee$. Nous développons ici la preuve par plongement en temps continu.

Pour tout $t \geqslant 0$, étiquetons l'arbre de Yule $\mathcal{Y}_{t,k}$ selon l'étiquetage aléatoire décrit en Définition 5.3.2 : cet arbre booléen aléatoire induit donc une distribution de probabilité sur $\mathcal{F}_k$ notée $\mathfrak{p}_{t,k}$. Tout comme précédemment, nous allons montrer que cette distribution converge vers la distribution asymptotique $\delta_{\texttt{Faux}}$ avant de traduire ce résultat en temps discret via la connexion (5.3). Pour cela, fixons $a = (a_1, \dots, a_k) \in \{0,1\}^k$ une affectation des $k$ variables. Nous allons montrer que la probabilité que la fonction booléenne calculée par $\mathcal{Y}_{t,k}$ vaille 1 en $a$ tend vers 0 quand $t$ tend vers $+\infty$. Notons $\pi_a(t) = \mathfrak{p}_{t,k}(f(a) = 1)$.

Si l'on calcule cette probabilité selon que l'horloge de la racine a sonné avant ou après la date $t$, on obtient :

$$\pi_a(t) = \mathtt{e}^{-t} \sum_{i=1}^{k} (\nu(x_i)\mathbb{1}_{\{a_i=1\}} + \nu(\bar{x}_i)\mathbb{1}_{\{a_i=0\}}) + \int_0^t \left[\varpi\, \pi_a(t-s)^2 + (1-\varpi)(2\pi_a(t-s) - \pi_a(t-s)^2)\right] \mathtt{e}^{-s} ds,$$

qui implique

$$\mathtt{e}^{t}\pi_a(t) = \frac{1}{2} + \int_0^t \left((2\varpi - 1)\pi_a(s)^2 + 2(1-\varpi)\pi_a(s)\right) e^s ds.$$

Dérivons par rapport à $t$. On a $\pi_a + \pi_a' = (2\varpi - 1)\pi_a^2 + 2(1-\varpi)\pi_a$, ce qui implique $\pi_a' = (2\varpi - 1)(\pi_a^2 - \pi_a)$. La fonction $\pi_a$ vérifie donc $\pi_a(t) = \frac{1}{\mathtt{e}^{(2\varpi-1)t}+1}$ car $\pi_a(0) = 1/2$. Nous pouvons donc en déduire

$$\mathfrak{p}_{t,k}(\mathcal{F}_k \backslash \{\texttt{Faux}\}) \leqslant \sum_a \pi_a(t) \leqslant 2^k \left(\frac{1}{\mathtt{e}^{(2\varpi-1)t}+1}\right).$$

Comme $\varpi > 1/2$, on a $\lim_{t\to+\infty} \mathfrak{p}_{t,k}(\mathcal{F}_k \backslash \{\texttt{Faux}\}) = 0$ et via des arguments similaires à ceux développés dans la démonstration du Théorème 5.2.3 (cf. Proposition 5.2.14), nous pouvons traduire ce résultat en temps discret :

$$\|p_{n,k} - \delta_{\texttt{Faux}}\|_\infty = \mathcal{O}\left(\frac{1}{n^{2\varpi-1}}\right).$$

■

### 5.3.3 Autoriser seulement les littéraux positifs

Ce dernier modèle d'étiquetage est inspiré de l'article de Fournier et al. [FGG09] concernant les arbres équilibrés. Il s'agit de n'autoriser que les littéraux positifs dans l'étiquetage des feuilles. Bien entendu, le système logique constitués des connecteurs $\wedge$ et $\vee$ et des seuls littéraux positifs n'est plus

complet : par exemple, ni la fonction `Vrai` ni la fonction `Faux` ne peuvent être représentées dans ce système. Dans les précédents systèmes logiques, seules les fonctions constantes étaient chargées par la distribution de l'arbre bourgeonnant : il est donc difficile de conjecturer comment la distribution se comportera dans un système logique qui ne calcule pas les fonctions constantes.

**Définition 5.3.4 (Étiquetage aléatoire - littéraux positifs)**

*Étant donné un arbre binaire planaire, étiquetons-le au hasard comme suit :*
- *chaque nœud interne est étiqueté par $\wedge$ avec probabilité $\varpi$ ou par $\vee$ avec probabilité $1-\varpi$,*
- *on définit $\nu$ une distribution de probabilité sur l'ensemble des littéraux positifs $\{x_1,\dots,x_k\}$ ; chaque feuille de l'arbre est étiquetée par un littéral choisi selon cette loi de probabilité ;*
- *chaque nœud, interne ou externe, est étiqueté indépendamment des autres.*

L'arbre bourgeonnant de taille $n$ étiqueté aléatoirement selon ce procédé induit donc une nouvelle loi, toujours notée $p_{n,k}$, sur l'ensemble des fonctions booléennes à $k$ variables $\mathcal{F}_k$. Nous allons établir le résultat asymptotique suivant :

**Théorème 5.3.5**

*La suite des distributions de probabilité $(p_{n,k})_{n\geqslant 0}$ vérifie :*
- *si $\varpi > 1/2$, alors $p_{n,k} \to \delta_{x_1\wedge\ldots\wedge x_k}$ ;*
- *si $\varpi < 1/2$, alors $p_{n,k} \to \delta_{x_1\vee\ldots\vee x_k}$.*

*Dans les deux cas, la vitesse de convergence est d'ordre $\mathcal{O}\left(\frac{1}{n^{|2\varpi-1|}}\right)$.*

Encore une fois, la distribution de l'arbre bourgeonnant se concentre sur une seule fonction booléenne et la vitesse de convergence est la même que dans le Théorème 5.3.3. Le cas critique $\varpi = 1/2$ n'est pas traité dans ce théorème : son comportement est plus compliqué et sera traité dans le Théorème 5.3.8.

**Démonstration :** Supposons par exemple que $\varpi > 1/2$. Nous développons encore une fois une approche par plongement en temps continu, et considérons la distribution $\mathfrak{p}_{t,k}$ induite sur $\mathcal{F}_k$ par l'arbre de Yule au temps $t$ étiqueté selon la Définition 5.3.4. Soit $a=(a_1,\dots,a_k)$ une affectation des $k$ variables, on définit $\pi_a(t) = \mathfrak{p}_{t,k}(f(a)=1)$.

Si $a=(1,\dots,1)$ alors $\pi_a(0) = \sum_{i=1}^k \nu(x_i)\mathbb{1}_{\{a_i=1\}} = 1$ et $\pi_a(t)=1$. Donc $\mathfrak{p}_{t,k}(f(1,\dots,1)=1)=1$. Si $a=(0,\dots,0)$, $\pi_a(t)=0$ pour tout $t\geqslant 0$. Enfin, si $a\neq(1,\dots,1)$ et $a\neq(0,\dots,0)$, par un calcul similaire à celui de la preuve du Théorème 5.3.3, nous obtenons

$$\pi_a(t) = \mathfrak{p}_{t,k}(f(a)=1) = \frac{1}{c_a \mathsf{e}^{(2\varpi-1)t}+1} \text{ pour tout } t\geqslant 0,$$

où $c_a$ est une constante non nulle car $\pi_a(0)\neq 1$. Comme $\varpi > \frac{1}{2}$, on a $\lim_{t\to+\infty}\mathfrak{p}_{t,k}(f(a)=1) = 0$. Nous pouvons désormais, via des arguments similaires à ceux développés dans la preuve de la Proposition 5.2.14, traduire ce résultat en temps discret : la suite des distributions $(p_{n,k})_{n\geqslant 0}$ converge quand $n$ tend vers $+\infty$ existe et donne probabilité 1 à la fonction $((x_1,\dots,x_k)\mapsto x_1\wedge\ldots\wedge x_k)$. ■

Afin de compléter l'étude de ce dernier étiquetage aléatoire, nous devons examiner le cas critique $\varpi = 1/2$. Il s'avère que ce dernier cas sera le plus complexe de tous, puisqu'il fera intervenir des fonctions un peu plus variées que les fonctions constantes ou les unions et conjonctions du Théorème 5.3.5 : les *fonctions seuil*. Ces fonctions, introduites par exemple dans [Ser04], apparaissent aussi dans l'étude des arbres booléens équilibrés [FGG09]. Voici leur définition :

**Définition 5.3.6 ([FGG09])**

*Soit $a=(a_1,\dots,a_k)\in\{0,1\}^k$ une affectation des variables. Le* **poids** *de $a$ relativement à la distribution $\nu$ sur les variables booléennes $\{x_1,\dots,x_k\}$ est le scalaire $\omega_\nu(a) = \nu(x_1)a_1 + \ldots +$*

$\nu(x_k)a_k$.

**Définition 5.3.7 ([FGG09])**

*Quel que soit $\theta \geqslant 0$, on définit la fonction seuil $f_{\nu,\theta}$ associée au réel $\theta$, relativement à $\nu$ comme suit : $\forall (a_1, \ldots, a_k) \in \{0,1\}^k$,*

$$f(a_1, \ldots, a_k) = 1 \Leftrightarrow \omega_\nu(a) \geqslant \theta.$$

**Théorème 5.3.8**

*Trions les éléments de $\{0,1\}^k$ par ordre de poids croissant (cf. Définition 5.3.6) :*

$$\omega_\nu(a^{(1)}) \leqslant \omega_\nu(a^{(2)}) \leqslant \ldots \leqslant \omega_\nu(a^{(2^k)}).$$

*Alors,*

$$p_{n,k} \longrightarrow \sum_{j=2}^{2^k} \left(\omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)})\right) \delta_{f_{\nu,\omega_\nu(a^{(j)})}} \text{ quand } n \to +\infty.$$

Autrement dit, $p_{n,k}$ tend vers la distribution aléatoire $p_k$ telle que

$$p_k(f_{\nu,\omega_\nu(a^{(j)})}) = \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)}) \text{ pour tout } j \in \{2, \ldots, 2^k\}.$$

Si $f \notin \{f_{\nu,\omega_\nu(a^{(j)})}\}_{j=2..2^k}$, alors $p_k(f) = 0$. Cette nouvelle distribution de l'arbre bourgeonnant ne charge que $2^k - 1$ fonctions seuils. On peut donc dire que cette distribution est encore *dégénérée* au sens où elle ne charge qu'un petit nombre de fonctions parmi les $2^{2^k}$ fonctions booléennes à $k$ variables. Ce comportement est toujours très éloigné de ce que l'on observait dans le cas des arbres de Catalan ou de Galton-Watson (cf. Théorèmes 1.3.1 et 1.3.8).

**Démonstration :** Nous développons encore une fois une approche par plongement en temps continu. Il est intéressant de lire cette preuve en parallèle de celle de la [FGG09, Proposition 7] car les ressemblances sont flagrantes.

Soient $a = (a_1, \ldots, a_k)$ et $b = (b_1, \ldots, b_k)$ deux affectations des $k$ variables, et soient $\alpha, \beta \in \{0,1\}$. Pour tout $t \geqslant 0$, soit

$$\pi_{\alpha\beta}(t) = \mathfrak{p}_{t,k}(f(a) = \alpha \text{ et } f(b) = \beta).$$

Calculons $\pi_{10}(t)$ selon que le premier temps de saut (cf. Définition 5.2.10) soit plus grand ou plus petit que $t$ :

$$\pi_{10}(t) = \mathsf{e}^{-t} \sum_{i=1}^{k} a_i(1-b_i)\nu(x_i) + \int_0^t \frac{1}{2}\big[\pi_{11}(t-s)\pi_{10}(t-s) + \pi_{10}(t-s)(\pi_{10}(t-s) + \pi_{11}(t-s)) + \pi_{10}(t-s)(\pi_{10}(t-s) + \pi_{00}(t-s)) + \pi_{10}(t-s)\pi_{00}(t-s)\big]\,\mathsf{e}^{-s}ds.$$

Après simplifications, nous obtenons

$$\pi_{10}(t)\mathsf{e}^t = \sum_{i=1}^{k} a_i(1-b_i)\nu(x_i) + \int_0^t \left(\pi_{10}(s)^2 + \pi_{10}(s)\pi_{11}(s) + \pi_{10}(s)\pi_{00}(s)\right)\mathsf{e}^s ds,$$

qui, d'après $\pi_{11}+\pi_{10}+\pi_{01}+\pi_{00} = 1$, et après dérivation, donne l'équation différentielle $\pi'_{10} = -\pi_{10}\pi_{01}$. Le même calcul peut être fait pour les trois autres fonctions $\pi_{00}$, $\pi_{01}$ et $\pi_{11}$, établissant ainsi le système

$$\left\{\begin{array}{rcl} \pi'_{10} &=& -\pi_{10}\pi_{01}; \\ \pi'_{01} &=& -\pi_{10}\pi_{01}; \\ \pi'_{11} &=& \pi_{10}\pi_{01}; \\ \pi'_{00} &=& \pi_{10}\pi_{01}. \end{array}\right. \tag{5.8}$$

Au vu de (5.8), nous pouvons affirmer que $\pi_{10}(t)$ et $\pi_{01}(t)$ sont deux fonctions décroissantes, et donc convergentes quand $t$ tend vers $+\infty$, comme elles sont aussi positives. De même, $\pi_{11}$ et $\pi_{00}$ sont croissantes et bornées supérieurement par 1, donc convergentes. On notera donc $l_{\alpha\beta} = \lim_{t\to+\infty} \pi_{\alpha\beta}(t)$ pour tout choix de $\alpha, \beta \in \{0,1\}$.

Comme, pour tout $\alpha, \beta \in \{0,1\}$ la fonction $\pi_{\alpha\beta}$ est monotone et converge quand $t \to +\infty$, sa dérivée tend vers 0 quand $t$ tend vers $+\infty$. Ainsi, nous pouvons passer à la limite dans le Système (5.8) et obtenir :

$$l_{10}l_{01} = 0. \tag{5.9}$$

De plus, d'après le Système (5.8), la fonction $\pi_{10} - \pi_{01}$ est constante, et donc,

$$l_{10} - l_{01} = \pi_{10}(0) - \pi_{01}(0) = \omega_\nu(a) - \omega_\nu(b). \tag{5.10}$$

Ainsi, si $\omega_\nu(a) \geqslant \omega_\nu(b)$, alors, au vu des Équations (5.9) and (5.10), $l_{01} = 0$. Autrement dit, $\mathfrak{p}_{t,k}(f(a) = 0 \text{ and } f(b) = 1)$ converge vers 0 quand $t$ tend vers $+\infty$. Cela veut dire que, s'il existe $a$ et $b$ tels que $\omega_\nu(a) \geqslant \omega_\nu(b)$, $f(a) = 0$ et $f(b) = 1$, alors $p_{n,k}(f)$ tend vers 0 quand $n$ tend vers $+\infty$. Ainsi, les seules fonctions booléennes chargées par $p_{n,k}$ quand $n$ tend vers $+\infty$ sont les fonctions telles que : pour tout $a, b$ tels que $\omega_\nu(a) \geqslant \omega_\nu(b)$, on a $f(a) \geqslant f(b)$. Ces fonctions sont donc des fonctions seuil : si la distribution asymptotique des $p_{n,k}$ existe quand $n$ tend vers $+\infty$, alors son support est inclus dans l'ensemble des fonctions seuil relativement à la distribution $\nu$.

Les calculs effectués dans la preuve du Théorème 5.3.5 peuvent être refaits pour le cas $\varpi = 1/2$, et ils permettent de montrer que $\mathfrak{p}_{t,k}(f(a) = 1)$ est une fonction constante pour tout $a \in \{0,1\}^k$. Ainsi $\mathfrak{p}_{t,k}(f(a) = 1) = \omega_\nu(a)$ et, pour tout $j \in \{2, \ldots, 2^k\}$,

$$p_{n,k}(f_{\nu,\omega_\nu(a^{(1)})}) + \ldots + p_{n,k}(f_{\nu,\omega_\nu(a^{(j)})}) \to \omega_\nu(a^{(j)}),$$

et $p_{n,k}(f_{\nu,\omega_\nu(a^{(j)})}) \to \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)})$ quand $n \to +\infty$. Finalement, l'égalité $\sum_{j=2}^{2^k} \omega_\nu(a^{(j)}) - \omega_\nu(a^{(j-1)}) = 1$ permet de conclure la preuve du Théorème 5.3.8. ■

Les Théorèmes 5.3.3, 5.3.5 et 5.3.8 vérifiés par le modèle de l'arbre bourgeonnant, mais aussi par le modèle des arbres équilibrés (cf. [FGG09]) ouvrent donc la porte à une réflexion plus générale : pourquoi ces deux modèles d'arbres induisent-ils une distribution *dégénérée* sur $\mathcal{F}_k$, alors que les arbres de Catalan et de Galton-Watson induisent un comportement tout à fait différent de leurs distributions induites ? C'est à cette question que nous répondrons dans la Section 5.5. Avant de s'intéresser à cette question, nous allons recentrer notre étude sur le modèle de l'implication, toujours en vue d'une comparaison plus complète entre les différents modèles arborescents cités.

## 5.4 Étude des tautologies dans le système de l'implication

Dans cette Section, nous nous intéressons uniquement au modèle de l'implication : les nœuds internes des arbres considérés sont tous étiquetés par le connecteur $\to$. Le première sous-section concerne le modèle de l'implication *classique* (tel qu'il est défini en Section 1.3.2), alors que la seconde sous-section concerne une généralisation où les littéraux positifs et négatifs sont autorisés pour l'étiquetage des feuilles.

Nous nous intéressons dans cette partie aux tautologies. Pour cette étude, la distribution qui nous intéresse n'est donc plus la distribution induite sur les fonctions booléennes $p_{n,k}$ (cf. Définition 5.1.2), mais directement la distribution sur les arbres booléens $\mathbb{P}_{n,k}$ (cf. Définition 5.1.1). Dans les modèles des arbres de Catalan, et des arbres de Galton-Watson, presque toute tautologie est *simple*, asymptotiquement quand $n$ (la taille des arbres) tend vers $+\infty$ (cf. [FGGZ07]). Cette propriété est-elle vraie pour l'arbre bourgeonnant ?

### 5.4.1 Tautologies simples

Rappelons la définition d'une tautologie simple :

**Définition 5.4.1 ([FGGZ07])**

*Dans le modèle de l'implication, toute expression booléenne s'écrit sous la forme* $A_1 \to (A_2 \to \ldots (A_p \to \alpha))$*, où les* $(A_i)_{i=1,\ldots,p}$ *sont des expressions booléennes. Les sous-arbres représentant* $A_1, \ldots, A_p$ *sont appelés* **prémisses** *de l'arbre booléen, et* $\alpha$ *est son* **but**. *Une* **tautologie simple** *est un arbre booléen dont au moins l'une des prémisses est réduite à une feuille étiquetée par le littéral* $\alpha$ *(cf. Figure 5.2).*

*On notera* $ST_{n,k}$ *l'ensemble des tautologies simples de taille* $n$ *sur* $k$ *variables.*

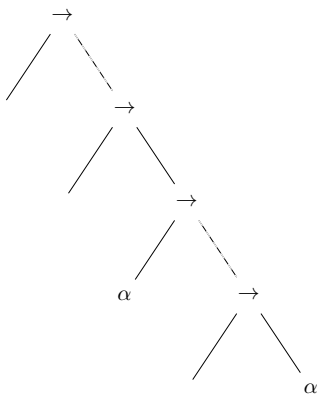



FIGURE 5.2 – *Une tautologie simple dans le système de l'implication.*

Rappelons que la distribution des arbres de Catalan ainsi que celle de l'arbre de Galton-Watson vérifient les Théorèmes 1.3.11 et 1.4.3 : en est-il de même pour la distribution de l'arbre bourgeonnant ? Le théorème suivant affirme que, du point de vue des tautologies, le comportement de l'arbre bourgeonnant diffère encore de celui des arbres de Catalan ou de l'arbre de Galton-Watson :

**Théorème 5.4.2**

$$\mathbb{P}_{n,k}(ST_{n,k}) \xrightarrow{n\to+\infty} 1 - \mathsf{e}^{-1/k} \sim 1/k,$$

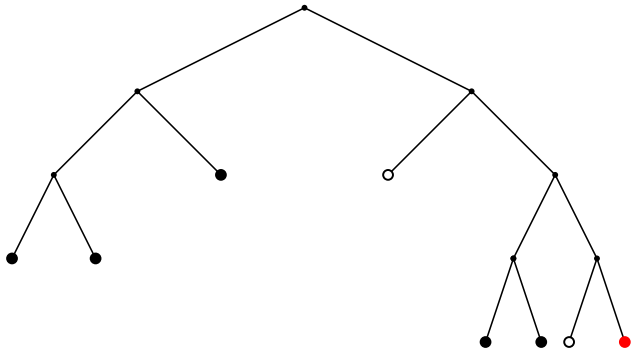
*asymptotiquement quand* $k$ *tend vers* $+\infty$.

Comme la probabilité de `Vrai` est 1 sous la distribution de l'arbre bourgeonnant (cf. Théorème 5.2.4), le Théorème 5.4.2 nous permet donc de conclure qu'asymptotiquement quand $k$ tend vers $+\infty$, *presqu'aucune tautologie n'est simple*. Nous avons donc là aussi un comportement très différent de celui observé pour les distributions des arbres de Catalan et de Galton-Watson.

La preuve de ce théorème s'appuie l'étude d'une urne de Pólya. Une introduction aux urnes de Pólya peut être lue en deuxième partie de ce mémoire (cf. Partie II) : nous utiliserons ici une approche par combinatoire analytique (cf. [FGP05] pour une introduction à cette méthode qui ne sera pas développées dans la Partie II du mémoire).

**Démonstration :** La preuve est faite en deux étapes : calculons tout d'abord la distribution du nombre $\ell_n$ de prémisses de l'arbre bourgeonnant de taille $n$ qui sont réduites à une feuille : nous appellerons ces prémisses des *feuilles sympas*. Calculons la probabilité qu'au moins une prémisse sympa soit étiquetée par le même littéral que le but de l'arbre bourgeonnant. Nous séparons le calcul en deux phases : une étude de la *forme* de l'arbre bourgeonnant, puis une étude de l'étiquetage.

FIGURE 5.3 – *L'arbre binaire ci dessous est colorié en accord avec l'urne décrivant les feuilles sympas : les feuilles blanches sont les feuilles sympas et la feuille rouge est la feuille but.*



C'est la première étape, celle de l'étude de la *forme* de l'arbre qui peut être modélisée via une urne de Pólya à trois couleurs. Nous définirons l'urne de telle sorte qu'elle contienne $n$ boules à l'étape $n$. Ces $n$ boules représenteront les $n$ feuilles de l'arbre bourgeonnant $\mathcal{T}_n$. Les boules blanches représenteront les feuilles sympas, il n'y aura qu'une seule boule rouge dans l'urne, qui représentera la feuille but de l'arbre, et les boules noires représenterons les autres feuilles (cf. Figure5.3). A l'étape 0, l'urne contient une unique boule rouge. A l'étape $n$, nous tirons au hasard une boule dans l'urne (donc une feuille dans $\mathcal{T}_n$) :

- si cette boule est rouge, alors, on la remet dans l'urne et on y ajoute une boule blanche (i.e. une feuille sympa) ;
- si cette boule est blanche, alors on la retire de l'urne et on ajoute deux boules noires dans l'urne ;
- si cette boule est noire, alors on la remet dans l'urne et on y ajoute une seconde boule noire.

La matrice de remplacement de l'urne (cf Partie II) est donc donnée par :

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Cette urne a été étudiée par Morcrette [Mor13] qui a prouvé que le nombre de boules blanches dans l'urne à l'étape $n$ a par ailleurs la même distribution que le nombre de points fixes dans une permutation uniforme de $\mathfrak{S}_n$, et vérifie : pour tout $n \geqslant 1$, pour tout $m \leqslant n+1$ :

$$\mathbb{P}_{n,k}(\ell_n = m) = \frac{1}{m!}\left(\mathsf{e}^{-1} - \sum_{j \geqslant n+1-m} \frac{(-1)^j}{j!}\right) = \sum_{j=0}^{n-m} \frac{(-1)^j}{j!}. \tag{5.11}$$

Passons maintenant à l'étude de l'étiquetage : calculons $\mathbb{P}_{n,k}(ST_{n,k})$ en conditionnant par le nombre de feuilles sympas de $\mathcal{T}_n$. Comme $\left(1 - \left(1 - \frac{1}{k}\right)^m\right)$ est la probabilité qu'au moins l'une des $m$ feuilles sympas soit étiquetée par la même étiquette que le but de $\mathcal{T}_{n,k}$, on a :

$$\mathbb{P}_{n,k}(ST_{n,k}) = \sum_{m=1}^{n} \mathbb{P}_{n,k}(\ell_n = m)\left(1 - \left(1 - \frac{1}{k}\right)^m\right).$$

Soit $c = \left(1 - \frac{1}{k}\right)$,

$$\mathbb{P}_{n,k}(ST_{n,k}) = \sum_{m=1}^{n} \frac{(1-c^m)}{m!} \sum_{j=0}^{n-m} \frac{(-1)^j}{j!} (1-c^m)$$
$$= \sum_{s=1}^{n} \sum_{j=0}^{s-1} \frac{\left(1-c^{s-j}\right)}{(s-j)!} \frac{(-1)^j}{j!}.$$

Cette série converge vers

$$\sum_{s=1}^{\infty} \sum_{j=0}^{s-1} \frac{\left(1-c^{s-j}\right)}{s-j!} \frac{(-1)^j}{j!} = \left(\sum_{s=1}^{\infty} \frac{(1-c^s)}{s!}\right) \left(\sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\right) = (\mathsf{e} - \mathsf{e}^c)\mathsf{e}^{-1} = 1 - \mathsf{e}^{c-1}.$$

Puis, comme $c = \left(1 - \frac{1}{k}\right)$, le Théorème 5.4.2 est démontré. ■

### 5.4.2 Autoriser les littéraux positifs et négatifs

Dans la littérature (cf. [FGGZ10]), nous pouvons trouver une étude des tautologies simples dans un modèle de l'implication modifié : où les littéraux négatifs sont autorisés dans l'étiquetage des feuilles. Dans cette sous-section, nous étudions les tautologies dans ce nouveau système logique selon la distribution de l'arbre bourgeonnant. Dans ce modèle, l'arbre bourgeonnant $\mathcal{T}_n$ est étiqueté uniformément au hasard avec le connecteur $\to$ et les littéraux $\{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$ (cf. Section 1.4). L'arbre étiqueté obtenu est noté $\mathcal{T}_{n,k}$, sa loi est de nouveau notée $\mathbb{P}_{n,k}$, et la distribution induite sur l'espace des fonctions booléennes $\mathcal{F}_k$ est notée $p_{n,k}$.

Tout comme cela a été fait dans la Section 5.2, nous pouvons montrer que la distribution $p_{n,k}$ converge vers la distribution limite $\delta_{\texttt{Vrai}}$ quand $n$ tend vers $+\infty$. La preuve est omise car elle est très similaire à la preuve développée dans la Section 5.2.

Dans ce système logique, pour les modèles de Catalan et de Galton-Watson, asymptotiquement quand $k$ tend vers $+\infty$, presque toute tautologie est *simple*. Mais dans ce cas, les tautologies simples peuvent être de deux types : les tautologies de premier type sont celles définies en Définition 1.3.10 et en Figure 5.2, et celle de second type sont définies comme suit :

**Définition 5.4.3 ([FGGZ10])**

*Une* **tautologie simple de second type** *est une expression booléenne dans laquelle deux feuilles sympas sont étiquetées par un littéral et sa négation (cf. Figure 5.4). On notera $ST^1_{n,k}$ (resp. $ST^2_{n,k}$) l'ensemble des tautologies de premier type (resp. de second type).*

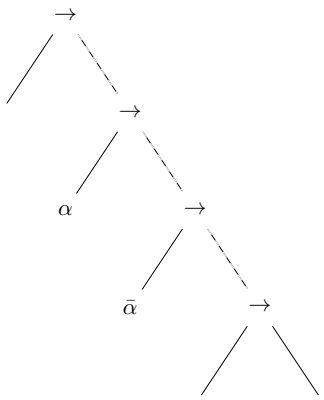
Encore une fois, dans le modèle de l'arbre bourgeonnant, *presqu'aucune tautologie n'est simple quand $k$ tend vers $+\infty$*. Plus que pour le résultat, l'introduction de ce système logique de l'implication avec littéraux négatifs est intéressante pour la preuve qui met en œuvre des méthodes d'allocation aléatoire.

**Théorème 5.4.4**

*Asymptotiquement quand $k$ tend vers $+\infty$,*
- $\mathbb{P}_{n,k}(ST^1_{n,k}) \xrightarrow{n\to+\infty} 1 - \mathsf{e}^{-1/2k} \sim \frac{1}{2k}$, *et*
- $\mathbb{P}_{n,k}(ST^2_{n,k}) \xrightarrow{n\to+\infty} 1 - \frac{1}{\mathsf{e}}(2\mathsf{e}^{1/2k} - 1)^k \sim \frac{1}{4k}$.

Comme la probabilité de `Vrai` est 1 selon la distribution de l'arbre bourgeonnant (cf. Théorème 5.2.4), on a bien que *presqu'aucune tautologie n'est simple* quand $k$ tend vers $+\infty$. Comme

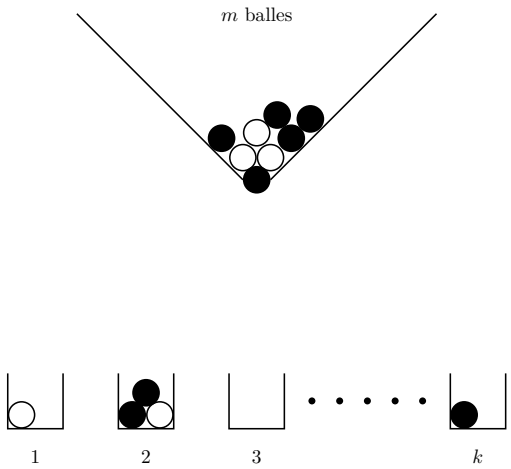
FIGURE 5.4 – *Tautologie simple de second type*




indiqué précédemment, la preuve de ce résultat repose sur une modélisation via un problème d'allocation aléatoire. Cette preuve, ainsi que celle du Théorème 5.4.2, montrent que l'étude de l'arbre bourgeonnant est liée à des problèmes de combinatoire très variés. Les problèmes d'allocation sont introduits par exemple dans le livre de Johnson et Kotz [JK97] : nous utiliserons ici une approche par combinatoire analytique (cf [Gar02] pour une revue de littérature sur cette approche).

**Démonstration :** La première affirmation du Théorème 5.4.4 peut être prouvée de la même façon que dans la preuve du Théorème 5.4.2 : nous ne détaillons pas cette preuve. Par ailleurs, la preuve de la seconde affirmation peut être démontrée en deux étapes comme dans la preuve du Théorème 5.4.2. Comme l'Équation (5.11) ne dépend que de la *forme* de l'arbre bourgeonnant, elle est encore valable dans ce nouveau modèle de l'implication. Supposons que $\mathcal{T}_{n,k}$ a $m$ feuilles sympas. Nous n'avons donc plus qu'à calculer la probabilité que deux feuilles sympas au moins parmi $m$ soient étiquetées par un littéral et sa négation. Modélisons ce problème par un problème d'allocation aléatoire. Nous devons distribuer $m$ boules (i.e. $m$ feuilles sympas) dans $k$ urnes (i.e. $k$ variables booléennes). Les boules peuvent être noires ou blanches (littéral négatif ou positif) avec probabilité $1/2$, indépendamment les unes des autres. Le probabilité qu'au moins une urne contienne au moins une boule blanche et une boule noire est égale à la probabilité que $\mathcal{T}_{n,k}$ soit une tautologie de second type (cf. Figure 5.5).

FIGURE 5.5 – *Problème d'allocation associé à l'étude des tautologies de second type.*




Nous allons résoudre ce problème par combinatoire analytique (cf. [Gar02] pour une revue des

ces méthodes). Considérons une urne isolée parmi les $k$ urnes du problème. La fonction génératrice $\mathsf{e}^{x+y}$ énumère le nombre de compositions de l'urne, où $x$ marque le nombre de boules blanches (resp. d'occurences positives de la variable booléenne associée à l'urne considérée) et $y$ le nombre de boules noires (resp. le nombre d'occurences négatives de la variable). On peut décomposer cette fonction génératrice comme suit :

$$\mathsf{e}^{x+y} = (\mathsf{e}^x - 1)(\mathsf{e}^y - 1) + (\mathsf{e}^x - 1) + (\mathsf{e}^y - 1) + 1,$$

où le premier terme représente les compositions de l'urne faisant intervenir au moins une boule blanche et une boule noire, le second terme (resp. le troisième terme) représente les compositions ne faisant intervenir que des boules blanches (resp. noires) et le quatrième terme représente l'urne vide. Introduisons une nouvelle variable complexe $z$ qui marquera les allocations qui réalisent une tautologie simple. La fonction génératrice des différentes compositions de l'urne devient

$$z(\mathsf{e}^x - 1)(\mathsf{e}^y - 1) + (\mathsf{e}^x - 1) + (\mathsf{e}^y - 1) + 1.$$

Oublions maintenant la différence entre les variables $x$ et $y$, et notons-les $t$. La description des différentes compositions d'une urne isolée devient donc

$$z(\mathsf{e}^t - 1)^2 + 2\mathsf{e}^t - 1.$$

Dès lors, si $\alpha_{r,m}$ est le nombre de façons de distribuer $m$ boules dans $k$ urnes de façon à réaliser $r$ fois une tautologie simple de second type (i.e. la tautologie simple est réalisée pour $r$ couples de feuilles sympas), alors

$$\Phi(t,z) := \sum_{r,m} \alpha_{r,m} z^r \frac{t^m}{m!} = (z(\mathsf{e}^t - 1)^2 + 2\mathsf{e}^t - 1)^k.$$

Comme $\Phi(t,0)$ est la fonction génératrice des allocations de $m$ boules dans les $k$ urnes qui *ne* réalisent *pas* une tautologie de second type (mais qui peuvent en réaliser une de premier type), on a :

$$\mathbb{P}_{n,k}(\overline{ST^2_{n,k}} | f_n = m) = \frac{[\frac{t^m}{m!}]\Phi(t,0)}{(2k)^m} = \frac{\alpha_{0,m}}{(2k)^m},$$

ce qui, au vu de l'Équation (5.11), donne

$$\mathbb{P}_{n,k}(\overline{ST^2_{n,k}}) = \underbrace{\mathsf{e}^{-1} \sum_{m=0}^{n} \frac{\alpha_{0,m}}{m!\,(2k)^m}}_{Q_n} - \underbrace{\sum_{m=0}^{n} \frac{\alpha_{0,m}}{m!\,(2k)^m} \sum_{j \geqslant n+1-m} \frac{(-1)^j}{j!}}_{R_n}.$$

Il est facile de voir que $R_n$ tend vers 0 quand $n$ tend vers $+\infty$, et que

$$Q_n \to \frac{1}{\mathsf{e}} \Phi\left(\frac{1}{2k}, 0\right) \text{ quand } n \to +\infty.$$

Donc,

$$\mathbb{P}_{n,k}(\overline{ST^2_{n,k}}) \xrightarrow{n \to +\infty} \frac{1}{\mathsf{e}} \Phi\left(\frac{1}{2k}, 0\right) = \frac{1}{\mathsf{e}} (2\mathsf{e}^{1/2k} - 1)^k \sim 1 - \frac{1}{4k},$$

quand $k$ tend vers $+\infty$. ■

## 5.5 Généralisation : arbres saturés

L'étude de la distribution de l'arbre bourgeonnant réalisée dans ce chapitre montre comment ce nouveau modèle est très différent des modèles de Catalan et de Galton-Watson : en effet, la distribution obtenue est dégénérée, aussi bien dans le système logique et/ou que dans celui de l'implication ;

et, si nous nous concentrons sur les tautologies simples dans le modèle de l'implication, nous avons montré que, asymptotiquement quand $k$ tend vers $+\infty$, il existe une proportion non négligeable de tautologies non simples parmi les tautologies. Nous avons même montré qu'asymptotiquement quand $k$ tend vers $+\infty$, presqu'aucune tautologie n'est simple. A contrario, la distribution de l'arbre bourgeonnant semble très proche de celle des arbres équilibrés : pouvons-nous montrer un méta-théorème nous assurant que *tout arbre aléatoire de la "même forme" qu'un arbre équilibré induit une distribution dégénérée sur* $\mathcal{F}_k$ ? Quelle est la bonne notion de *"forme"* à considérer ?

L'objet de cette Section est d'énoncer et de démontrer le méta-théorème annoncé ci-dessus. Il s'avère que la bonne notion de *forme* est le niveau de saturation d'un arbre aléatoire, et que la condition nécessaire et suffisante pour que la distribution induite sur $\mathcal{F}_k$ soit dégénérée est que le niveau de saturation tende vers $+\infty$ en probabilité quand la taille de l'arbre aléatoire tend vers $+\infty$. La démonstration de ce méta-théorème met en œuvre les idées de la preuve par plogement en temps continu du Théorème 5.2.3.

**Définition 5.5.1**

*On appelle* **niveau de saturation** *d'un arbre la hauteur de sa feuille la plus proche de sa racine.*

Introduisons tout d'abord une nouvelle notation : étant donné un arbre booléen $\tau$, nous noterons $f_\tau$ la fonction booléenne représentée par $\tau$.

Rappelons que $\mathcal{T}$ est l'ensemble des arbres binaires plans, et soit $\mathcal{T}_\sigma$ l'ensemble de ces arbres ayant un niveau de saturation supérieur ou égal à $\sigma$, pour tout $\sigma \geqslant 0$. Remarquons que $\mathcal{T}_0 = \mathcal{T}$.

Étant donné un arbre $t$, étiquetons-le uniformément au hasard dans le système et/ou (cf. Définition 1.2.6), et notons $\hat{t}$ l'arbre booléen aléatoire obtenu, et $f_{\hat{t}}$ la fonction booléenne aléatoire qu'il calcule : c'est une fonction booléenne de $\mathcal{F}_k$.

Soient $a$ et $b$ deux éléments distincts de $\{0,1\}^k$, et soient $\alpha, \beta \in \{0,1\}$. Notons, pour tout arbre $t$,

$$\mathbb{P}_t^{\alpha\beta}(a,b) = \mathbb{P}(f_{\hat{t}}(a) = \alpha \text{ and } f_{\hat{t}}(b) = \beta),$$

et

$$S_\sigma^{\alpha\beta}(a,b) = \sup\{\mathbb{P}_t^{\alpha,\beta}(a,b), t \in \mathcal{T}_\sigma\}.$$

Si cet abus de notation est sans ambiguïté, nous écrirons $\mathbb{P}_t^{\alpha\beta}$ au lieu de $\mathbb{P}_t^{\alpha\beta}(a,b)$ et $S_\sigma^{\alpha\beta}$ au lieu de $S_\sigma^{\alpha\beta}(a,b)$.

Jusqu'à maintenant, l'aléa du modèle vient de l'étiquetage des arbres : la structure est pour l'instant déterministe. Nous souhaitons désormais rendre aléatoire cette structure. Nous aurons dans la suite deux niveaux d'aléa, celui sur la structure, et celui sur l'étiquetage. Soit $(T_n)_{n\geqslant 0}$ une suite d'arbres binaires plans aléatoires : nous nous intéressons au comportement de $f_{\hat{T}_n}$, et notre objectif est de démontrer le théorème suivant :

**Théorème 5.5.2**

*Soit $(T_n)_{n\geqslant 0}$ une suite d'arbre binaires plans aléatoires (non étiquetés). Asymptotiquement quand $n$ tend vers $+\infty$,*

$$\mathbb{P}(f_{\hat{T}_n} = \texttt{Vrai}) = \mathbb{P}(f_{\hat{T}_n} = \texttt{Faux}) \to \frac{1}{2}, \tag{5.12}$$

*si, et seulement si, le niveau de saturation $s_n$ de $T_n$ vérifie*

$$\lim_{n\to+\infty} s_n = +\infty \text{ en probabilité.}$$

Des cas particuliers de ce théorème ont déjà été cités dans ce mémoire :

- Si, pour tout $n \geqslant 0$, $T_n$ est presque sûrement égal à l'arbre équilibré de hauteur $n$ (i.e. l'unique arbre ayant toutes ses feuilles à hauteur $n$), alors, il est démontré dans [FGG09] que l'Équation (5.12) est bien vérifiée.
- Si la suite $(T_n)_{n\geqslant 0}$ est l'arbre bourgeonnant (non étiqueté), alors, il est montré dans ce chapitre que l'Équation (5.12) est bien vérifiée (cf. Théorème 5.2.3).
- Si, pour tout $n \geqslant 0$, l'arbre $T_n$ suit la loi uniforme parmi les arbres de taille $n$, c'est alors le modèle de Catalan (cf. Section 1.3.1) et il est démontré dans [CFGG04, Koz08] que l'Équation (5.12) n'est pas vérifiée.
- Si la suite $(T_n)_{n\geqslant 0}$ est le processus de Galton-Watson critique (remarquons que, presque sûrement, pour $n$ assez grand, $T_{n+1} = T_n$ car le processus de Galton-Watson critique s'éteint presque sûrement), alors, c'est le modèle de Galton-Watson (cf. Section 1.4) et il est démontré que l'Équation (5.12) n'est pas vérifiée.

Pour démontrer l'équivalence énoncée dans le Théorème 5.5.2, nous montrons séparément les deux implications dans les Propositions 5.5.4 et 5.5.6. Il nous faut aussi séparer l'étude des deux niveaux d'aléa, c'est pourquoi nous établirons tout d'abord deux lemmes qui ne traiteront que l'aléa provenant de l'étiquetage aléatoire.

**Lemme 5.5.3**

*Pour tout $a \neq b \in \{0,1\}^k$, asymptotiquement quand $\sigma$ tend vers $+\infty$,*

$$1/2 - S_\sigma^{11}(a,b) = 1/2 - S_\sigma^{00}(a,b) \sim \frac{1}{\sigma},$$

*et, pour tout $\varepsilon > 0$,*

$$S_\sigma^{10}(a,b) = S_\sigma^{01}(a,b) = \mathcal{O}(1/\sigma^{1-\varepsilon}).$$

**Démonstration :** Remarquons tout d'abord que les symétries de notre modèle d'étiquetage aléatoire impliquent $\mathbb{P}_t^{10} = \mathbb{P}_t^{01}$ et $\mathbb{P}_t^{11} = \mathbb{P}_t^{00}$ pour tout arbre $t$ : les probabilités de $\wedge$ et $\vee$ sont en effet égales, ainsi que les probabilités d'une variable et de sa négation. Nous pouvons donc conclure que, pour tout arbre $t$, et pour tout fonction booléenne $f$,

$$\mathbb{P}(f_{\hat{t}} = f) = \mathbb{P}(f_{\hat{t}} = \bar{f}).$$

De plus, pour tout arbre $t$, $\mathbb{P}_t^{10} + \mathbb{P}_t^{11} = \frac{1}{2}$.

Soit $t$ un arbre de $\mathcal{T}_\sigma$. Son sous-arbre gauche $t_\ell$ et son sous-arbre droit $t_r$ sont tous deux des éléments de $\mathcal{T}_{\sigma-1}$, ce qui implique :

$$\begin{aligned}\mathbb{P}_t^{10} &= \frac{1}{2}\left(\mathbb{P}_{t_\ell}^{10}(\mathbb{P}_{t_r}^{10} + \mathbb{P}_{t_r}^{11}) + \mathbb{P}_{t_\ell}^{11}\mathbb{P}_{t_r}^{10}\right) + \frac{1}{2}\left(\mathbb{P}_{t_\ell}^{10}(\mathbb{P}_{t_r}^{10} + \mathbb{P}_{t_r}^{00}) + \mathbb{P}_{t_\ell}^{00}\mathbb{P}_{t_r}^{10}\right)\\ &= \frac{1}{2}\mathbb{P}_{t_\ell}^{10} + \frac{1}{2}\mathbb{P}_{t_r}^{10} - \mathbb{P}_{t_\ell}^{10}\mathbb{P}_{t_r}^{10}. \qquad (5.13)\end{aligned}$$

En effet, le facteur $\frac{1}{2}$ est égal à la probabilité que la racine de $t$ soit étiquetée par $\wedge$ ou par $\vee$, le premier terme de la somme suppose que la racine est étiquetée par $\wedge$ et le second terme par $\vee$ ; il suffit de remarquer que les étiquettes des deux sous-arbres sont indépendantes pour établir les égalités ci-dessus. Comme $\mathbb{P}_{t_\ell}^{10} \leqslant S_{\sigma-1}^{10}$ et $\mathbb{P}_{t_r}^{10} \leqslant S_{\sigma-1}^{10}$, nous avons

$$S_\sigma^{10} \leqslant S_{\sigma-1}^{10},$$

ce qui implique que $(S_\sigma^{10})_{\sigma\geqslant 0}$ est convergente quand $\sigma$ tend vers $+\infty$. Un calcul similaire donne

$$\begin{aligned}\mathbb{P}_t^{11} &= \frac{1}{2}\mathbb{P}_{t_\ell}^{11}\mathbb{P}_{t_r}^{11} + \frac{1}{2}(\mathbb{P}_{t_\ell}^{11} + \mathbb{P}_{t_\ell}^{10}(\mathbb{P}_{t_r}^{01} + \mathbb{P}_{t_r}^{11}) + \mathbb{P}_{t_\ell}^{01}(\mathbb{P}_{t_r}^{10} + \mathbb{P}_{t_r}^{11}) + \mathbb{P}_{t_\ell}^{00}\mathbb{P}_{t_r}^{11})\\ &= \frac{1}{4} + \mathbb{P}_{t_\ell}^{11}\mathbb{P}_{t_r}^{11}\end{aligned}$$

ce qui implique

$$\frac{1}{4} + (S^{11}_{\sigma-1})^2 = S^{11}_{\sigma}.$$

Posons $v_\sigma = \frac{1}{2} - S^{11}_\sigma$. Dès lors, pour tout $\sigma \geqslant 0$, $v_{\sigma+1} = f(v_\sigma)$ où $f(x) = x - x^2$. Notons que 0 est l'unique point fixe de $f$ et que $[0,1]$ est stable par $f$. De plus, $f'(0) = 1$ et $f''(0) = -2$. Par un résultat standard sur les suites récursives : asymptotiquement quand $\sigma$ tend vers $+\infty$,

$$v_\sigma \sim -\frac{2}{\sigma f''(0)} = \frac{1}{\sigma}.$$

Dès lors, la suite $(S^{11}_\sigma)_{\sigma\geqslant 0}$ converge vers $\frac{1}{2}$ quand $\sigma$ tend vers $+\infty$, et

$$\frac{1}{2} - S^{11}_\sigma \sim \frac{1}{\sigma} \text{ quand } \sigma \to +\infty.$$

Nous en déduisons

$$S^{11}_\sigma = \frac{1}{2} - \frac{1}{\sigma} + o\left(\frac{1}{\sigma}\right).$$

En particulier, pour tout $\varepsilon > 0$, il existe $\sigma_\varepsilon \geqslant 0$ tel que, pour tout $\sigma \geqslant \sigma_\varepsilon$,

$$S^{11}_\sigma \leqslant \frac{1}{2} - \frac{1-\varepsilon}{\sigma}.$$

Au vu de l'Équation (5.13), nous avons

$$\begin{aligned}
\mathbb{P}^{10}_t &= \frac{1}{2}\mathbb{P}^{10}_{t_\ell} + \frac{1}{2}\mathbb{P}^{10}_{t_r} - \mathbb{P}^{10}_{t_\ell}\mathbb{P}^{10}_{t_r} \\
&= \frac{1}{2}\mathbb{P}^{10}_{t_\ell} + \frac{1}{2}\mathbb{P}^{10}_{t_r} - \left(\frac{1}{2} - \mathbb{P}^{11}_{t_\ell}\right)\left(\frac{1}{2} - \mathbb{P}^{11}_{t_r}\right) \\
&= \frac{1}{2}\mathbb{P}^{10}_{t_\ell} + \mathbb{P}^{11}_{t_\ell}\mathbb{P}^{10}_{t_r} \\
&\leqslant \frac{1}{2}S^{10}_{\sigma-1} + \left(\frac{1}{2} - \frac{1-\varepsilon}{\sigma-1}\right) S^{10}_{\sigma-1},
\end{aligned}$$

pour tout $\sigma \geqslant \sigma_\varepsilon$. Dès lors, pour tout $\sigma \geqslant \sigma_\varepsilon$,

$$S^{10}_\sigma \leqslant S^{10}_{\sigma-1}\left(1 - \frac{1-\varepsilon}{\sigma-1}\right).$$

Via la formule de Stirling $\prod_{k=1}^{n}\left(1 - \frac{1-\varepsilon}{n}\right) \sim \frac{n^{-(1-\varepsilon)}}{\Gamma(\varepsilon)}$ quand $n$ tend vers $+\infty$, il existe une constante $c_\varepsilon > 0$ telle que

$$S^{10}_\sigma \leqslant S^{10}_{\sigma_\varepsilon} \prod_{k=\sigma_\varepsilon}^{\sigma-1}\left(1 - \frac{1-\varepsilon}{k}\right) \leqslant \frac{c_\varepsilon}{(\sigma-1)^{1-\varepsilon}}.$$

Nous avons donc montré que, pour tout $\varepsilon > 0$, pour tout choix de $a \neq b$,

$$S^{10}_\sigma(a,b) = \mathcal{O}\left(\frac{1}{\sigma^{1-\varepsilon}}\right).$$

■

Nous sommes désormais prêts à démontrer la première implication du Théorème 5.5.2 : si la suite des niveaux de saturation des $(T_n)_{n\geqslant 0}$ tend vers $+\infty$ en probabilité, alors la loi induite sur $\mathcal{F}_k$ est dégénérée, au sens où elle ne charge que les deux fonctions constantes `Vrai` et `Faux`.

**Proposition 5.5.4**

*Soit $(T_n)_{n\geqslant 0}$ une suite d'arbres binaires plans aléatoires (non étiquetés) dont la suite $(s_n)_{n\geqslant 0}$*

*des niveaux de saturation tend vers $+\infty$ en probabilité. Dès lors, pour tout choix de $a, b \in \{0,1\}^k$ tels que $a \neq b$,*

$$\mathbb{P}^{10}_{T_n}(a,b) \to 0$$

*presque sûrement quand $n$ tend vers $+\infty$, ce qui implique*

$$\mathbb{P}(f_{\hat{T}_n} = \texttt{Vrai}) = \mathbb{P}(f_{\hat{T}_n} = \texttt{Faux}) \to \frac{1}{2} \text{ quand } n \to \infty.$$

*De plus, asymptotiquement quand $n$ tend vers $+\infty$,*

$$\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) \to 0.$$

**Démonstration :** Par hypothèse, pour tout $\sigma > 0$, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mathbb{P}(s_n \geqslant \sigma) \to 1.$$

Autrement dit, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mathbb{P}(T_n \notin \mathcal{T}_\sigma) \to 0.$$

Donc, pour tout $\sigma > 0$, au vu du Lemme 5.5.3,

$$\begin{aligned}
\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) &= \mathbb{P}(\exists a, b \in \{0,1\}^k | f_{\hat{T}_n}(a) \neq f_{\hat{T}_n}(b)) \\
&\leqslant \sum_{a \neq b} \mathbb{P}(f_{\hat{T}_n}(a) \neq f_{\hat{T}_n}(b)) \\
&\leqslant \sum_{a \neq b} \left( \mathbb{P}(f_{\hat{T}_n}(a) \neq f_{\hat{T}_n}(b) \text{ et } T_n \in \mathcal{T}_\sigma) + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right) \\
&= \sum_{a \neq b} \left( \sum_{t \in \mathcal{T}_\sigma} \mathbb{P}(f_{\hat{t}}(a) \neq f_{\hat{t}}(b) \text{ et } T_n = t) + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right),
\end{aligned}$$

car $\mathcal{T}_\sigma$ est dénombrable. De plus, comme la suite $(T_n)_{n \geqslant 0}$ est indépendante de son étiquetage,

$$\begin{aligned}
\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) &= \sum_{a \neq b} \left( \sum_{t \in \mathcal{T}_\sigma} \mathbb{P}(f_{\hat{t}}(a) \neq f_{\hat{t}}(b)) \mathbb{P}(T_n \in \mathcal{T}_\sigma) + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right) \\
&\leqslant \sum_{a \neq b} \left( (S^{10}_\sigma(a,b) + S^{01}_\sigma(a,b)) \mathbb{P}(T_n \in \mathcal{T}_\sigma) + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right) \\
&\leqslant \sum_{a \neq b} \left( 2 S^{10}_\sigma(a,b) + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right).
\end{aligned}$$

Le Lemme 5.5.3 appliqué à $\varepsilon = \frac{1}{2}$ nous assure qu'il existe une constante $c > 0$ telle que, pour tout $n \geqslant 0$,

$$\begin{aligned}
\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) &\leqslant \sum_{a \neq b} \left( 2 \frac{c_{a,b}}{\sigma^{1/2}} + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right) \\
&\leqslant 2^k (2^k - 1) \left( \frac{cst}{\sigma^{1/2}} + \mathbb{P}(T_n \notin \mathcal{T}_\sigma) \right).
\end{aligned}$$

Comme $\mathbb{P}(T_n \notin \mathcal{T}_\sigma)$ tend vers 0 quand $n$ tend vers $+\infty$, nous obtenons que, pour tout $\sigma > 0$, il existe $m \geqslant 0$, tel que, pour tout $n \geqslant m$,

$$\mathbb{P}(T_n \notin \mathcal{T}_\sigma) \leqslant \frac{c}{\sqrt{\sigma}},$$

ce qui implique

$$\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) \leqslant \frac{2cst}{\sigma^{1/2}}.$$

■

Il nous reste à montrer la seconde implication du Théorème 5.5.2 : si la suite $(s_n)_{n\geqslant 0}$ des niveaux de saturation des $(T_n)_{n\geqslant 0}$ ne tend pas vers $+\infty$ en probabilité, alors, la distribution induite sur $\mathcal{F}_k$ n'est pas dégénérée. Étudions tout d'abord l'influence de l'aléa de l'étiquetage via le lemme suivant :

**Lemme 5.5.5**

*Soit $\mathcal{A}_\sigma$ l'ensemble des arbres binaires plans (non étiquetés) de niveau de saturation $\sigma$, et soit*

$$I_\sigma = \inf_{t\in\mathcal{A}_\sigma} \mathbb{P}(f_{\hat{t}} \notin \{\texttt{Vrai}, \texttt{Faux}\}).$$

*Pour tout $\sigma > 0$,*

$$I_\sigma \geqslant \frac{1}{2^\sigma}.$$

**Démonstration :** Soit $t$ un arbre de $\mathcal{A}_\sigma$ et $\hat{t}$ sa version après étiquetage aléatoire. On note $x$ l'étiquette de $\ell$, l'une de ses feuilles de hauteur $\sigma$, $\diamond_1, \ldots, \diamond_\sigma$ les connecteurs étiquetant les ancêtres de $\ell$ ($\diamond_1$ sera l'étiquette de la racine et $\diamond_\sigma$ l'étiquette du parent de $\ell$), et par $g_1, \ldots, g_k$ les fonctions booléennes aléatoires calculées par les sous-arbres composés des descendants respectifs des ancêtres de $\ell$ (numérotés de haut en bas). Dès lors, la fonction calculée par $\hat{t}$ est donnée par

$$f_{\hat{t}} = (g_1 \diamond_1 (g_2 \diamond_2 (\ldots (g_\sigma \diamond_\sigma x)))).$$

Soit $j$ le minimum des $k \in \{1, \ldots, \sigma\}$ tels que $g_k$ admette $x$ comme variable essentielle. Si un tel $k$ n'existe pas, on posera $j = \sigma$. Soit $\rho = g_{j+1} \diamond_{j+1} (\ldots (g_\sigma \diamond_\sigma x))$. Notons que l'on peut choisir $\diamond_j$ tel que $g_j \diamond_j \rho$ admette $x$ pour variable essentielle : si $g_j \not\equiv 0$, alors $x \wedge g_j$ admet $x$ comme variable essentielle et si $g_j \equiv 0$, alors $x \vee g_j \equiv x$ admet $x$ comme variable essentielle. Par induction, nous pouvons ainsi choisir $\diamond_{j-1}, \ldots, \diamond_1$ tels que $f_{\hat{t}}$ admet $x$ comme variable essentielle, et la probabilité, conditionnellement au reste des étiquettes de l'arbres (et ce pour tout conditionnement), que $\diamond_1, \ldots, \diamond_j$ soient en effet égaux à ce choix est égale à $\frac{1}{2^j}$.

En conclusion, si l'on note $f_{\hat{t}}(x = 1)$ (resp. $f_{\hat{t}}(x = 0)$) la restriction de $f_{\hat{t}}$ au sous-ensemble de $\{0, 1\}^k$ où $x = 1$ (resp. $x = 0$),

$$\mathbb{P}(f_{\hat{t}}(x = 1) \neq f_{\hat{t}}(x = 0)) \geqslant \frac{1}{2^\sigma},$$

ce qui implique

$$\mathbb{P}(f_{\hat{t}} \neq \texttt{Vrai} \text{ et } f_{\hat{t}} \neq \texttt{Faux}) \geqslant \frac{1}{2^\sigma}.$$

Cette dernière inégalité est valable pour tout arbre $t \in \mathcal{A}_\sigma$ : il ne reste plus qu'à prendre l'infimum pour montrer le Lemme 5.5.5. ■

**Proposition 5.5.6**

*Supposons que la suite $(s_n)_{n\geqslant 0}$ des niveaux de saturation des $(T_n)_{n\geqslant 0}$ ne tende pas vers $+\infty$ en probabilité. Alors, il existe $\alpha > 0$ tel que, pour tout $m \geqslant 0$, il existe $n \geqslant m$ tel que,*

$$\mathbb{P}(f_{\hat{T}_n} = \texttt{Vrai}) = \mathbb{P}(f_{\hat{T}_n} = \texttt{Faux}) < \frac{1}{2} - \alpha.$$

**Démonstration :** Par hypothèse, il existe $\sigma > 0$ et $\varepsilon > 0$ tels que, pour tout $m \in \mathbb{N}$, il existe $n \geqslant m$ tel que

$$\mathbb{P}(s_n \geqslant \sigma) < 1 - \varepsilon.$$

Dès lors, au vu du Lemme 5.5.5,

$$\begin{aligned}\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) &\geqslant \mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\} \text{ et } s_n < \sigma) \\ &= \sum_{m=0}^{\sigma-1} \mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\} \text{ et } s_n = m) \\ &= \sum_{m=0}^{\sigma-1} \sum_{t \in \mathcal{A}_m} \mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\} \text{ et } T_n = t).\end{aligned}$$

Via l'indépendance entre la suite des arbres aléatoires $(T_n)_{n\geqslant 0}$ et leur étiquetage aléatoire, nous obtenons :

$$\begin{aligned}\mathbb{P}(f_{\hat{T}_n} \notin \{\texttt{Vrai}, \texttt{Faux}\}) &= \sum_{m=0}^{\sigma-1} \sum_{t \in \mathcal{A}_m} \mathbb{P}(f_{\hat{t}} \notin \{\texttt{Vrai}, \texttt{Faux}\}) \mathbb{P}(T_n = t) \\ &\geqslant \sum_{m=0}^{\sigma-1} \sum_{t \in \mathcal{A}_m} I_m \mathbb{P}(T_n = t),\end{aligned}$$

car $I_m = \inf_{t \in \mathcal{A}_m} \mathbb{P}(f_{\hat{t}} \notin \{\texttt{Vrai}, \texttt{Faux}\})$. Nous en déduisons

$$\begin{aligned}\mathbb{P}(f_{\hat{T}_n} \notin \{True, False\}) &\geqslant \sum_{m=0}^{\sigma-1} \frac{1}{2^m} \mathbb{P}(s_n = m) \\ &\geqslant \frac{1}{2^{\sigma-1}} \mathbb{P}(s_n < \sigma) \\ &\geqslant \frac{\varepsilon}{2^{\sigma-1}},\end{aligned}$$

par hypothèse. En posant $\alpha = \frac{\varepsilon}{2^\sigma}$, nous concluons la preuve de la Proposition 5.5.6 et donc celle du Théorème 5.5.2. ■

## 5.6 Conclusion

Nous avons montré dans ce chapitre le lien étroit qui relie le niveau de saturation d'une famille d'arbres booléens aléatoires à la distribution qu'elle induit sur l'ensemble des fonction booléennes à $k$ variables. Ce résultat a été conjecturé suite à l'étude de la nouvelle distribution sur $\mathcal{F}_k$, dite de l'arbre bourgeonnant, distribution qui s'avère très similaire à celle induite par les arbres équilibrés.

Il est important d'insister sur les méthodes utilisées pour l'étude du modèle de l'arbre bourgeonnant : nous avons présenté une approche par combinatoire analytique qui était plus naturelle au vu de la littérature sur les arbres booléens aléatoires, et une approche par plongement en temps continu, spécifique à l'arbre bourgeonnant dont le plongement en temps continu, l'arbre de Yule, est standard.

L'idée clef de la preuve du méta-théorème reliant niveau de saturation d'une suite d'arbres aléatoires et dégénérescence de la distribution qu'elle induit sur $\mathcal{F}_k$ est de séparer l'aléa sur la forme des arbres aléatoires de l'aléa sur leur étiquetage. Il est intéressant de noter que, une fois cette séparation faite, la démonstration met en œuvre des idées similaires à celles de la preuve par plongement en temps continu pour l'arbre bourgeonnant.

Pour compléter ce résultat général sur les distributions induites sur $\mathcal{F}_k$ par des modèles d'arbres booléens, il faudrait étudier plus en détail le cas non dégénéré : avons-nous toujours dans ce cas une probabilité favorisant les fonctions de faible complexité, avec un comportement en $\frac{1}{k^{L(f)}}$, asymptotiquement quand $k$ tend vers $+\infty$ ?

Par ailleurs, pouvons-nous trouver une approche générale qui incluerait aussi les modèles d'arbres non binaires (par exemple ceux du Chapitre 2) ? Par exemple, que dire des *généralisations* non binaires de l'arbre binaire de recherche tels les arbres croissants [BFS92] ou le *balanced probability model* [CDM93] ? A priori, ces arbres "équilibrés" devraient induire des distributions dégénérées, tout comme le modèle de l'arbre bourgeonnant étudié dans le présent chapitre. Est-il possible montrer un tel résultat en toute généralité ?

# Chapitre 6

# Arbres aléatoires étiquetés par un ensemble non borné de variables

## 6.1 Introduction

Dans tous les chapitres de la première sous-partie **Arbres booléens généraux** (cf. Chapitres 2, 3 et 4), ainsi que dans le modèle des arbres de Catalan étudié dans la littérature, une double limite est opérée. La taille $n$ des arbres considérés tend vers $+\infty$ afin de définir la distribution asymptotique longuement étudiée ensuite. Cette distribution, asymptotique, généralement notée $\mu_k$ dépend du nombre $k$ de variables utilisées pour l'étiquetage des arbres booléens. Les seuls résultats décrits dans la littérature sont ensuite obtenus asymptotiquement quand $k$ tend vers l'infini. Mais cette double-limite, ordonnée (on ne peut, a priori échanger les deux limites), ne biaise-t-elle pas les résultats obtenus ?

L'ambition de ce chapitre est d'échanger les deux limites, voire de laisser $n$ et $k$ tendre vers $+\infty$ ensemble, en exprimant $k$ comme fonction de $n$.

Une première idée due à Genitrini et al. [GKZ07, GK12], afin d'*échanger* les deux limites est de considérer des arbres booléens dont les étiquettes sont prises dans un ensemble infini $\{x_i\}_{i\geqslant 1}$ de variables (et leurs négations). La famille de ces arbres booléens n'est plus une classe combinatoire : il y a une nombre infini de tels arbres de taille $n$ (dans ce chapitre, la taille des arbres sera le nombre de feuilles). Nous définissons donc une relation d'équivalence sur les arbres booléens de façon à ce qu'il y ait un nombre fini de classes d'équivalences d'arbres de taille $n$. Cette notion d'équivalence induit une notion d'équivalence sur l'espace des fonctions booléennes.

Nous définissons $\mathbb{P}_n\langle f\rangle$, la probabilité d'une classe d'équivalence de fonctions $\langle f\rangle$ (où $f$ est une fonction booléenne), comme la proportion de classes d'équivalence d'arbres de taille de taille $n$ calculant une fonction de $\langle f\rangle$ parmi les classes d'équivalence d'arbres de taille $n$. Quel est le comportement de $\mathbb{P}_n\langle f\rangle$ quand $n$ tend vers $+\infty$ ?

Nous remarquerons que ce modèle où les deux limites (sur $k$ puis sur $n$) ont été *échangées* est en fait équivalent à un modèle où $k = n$, et nous généraliserons nos résultats pour toute suite $(k_n)_{n\geqslant 1}$ : nous supposerons qu'un arbre peut être étiquetée par au plus $k_n$ variables différentes où $(k_n)_{n\geqslant 1}$ sera une suite croissante et tendant vers $+\infty$ quand $n$ tend vers $+\infty$. Il est intéressant de remarquer que comme un arbre de taille $n$ a $n$ feuilles, $1 \leqslant k_n \leqslant n$, pour tout entier $n \geqslant 1$. Ce nouveau modèle nous permet donc de laisser $n$ et $k$ tendre vers $+\infty$ simultanément et nous verrons comment le comportement de la distribution $\mathbb{P}_n$ dépend de la dynamique de $k_n$ en fonction de $n$.

Dans la Section 6.2, nous définirons notre nouveau modèle, et notamment la notion de classe d'équivalence d'arbres et de fonctions booléennes, et nous énoncerons notre résultat principal. Les

Sections 6.3 et 6.4 sont le cœur technique de notre étude. Dans un premier temps (Section 6.3), nous étudions en détails le nombre de classes d'équivalences d'arbres de taille $n$, et nous y voyons clairement le rôle de la suite $k_n$. Nous remarquons un phénomène de *saturation* si la suite $(k_n)_{n\geqslant 1}$ vérifie $k_n \geqslant \frac{n}{\ln n}$ pour tout $n$ assez grand. Dans un second temps (Section 6.4), nous généralisons la théorie des motifs de Kozik [Koz08] à ce nouveau modèle de classes d'équivalence. Enfin, la Section 6.5 applique ces résultats techniques à l'étude de la distribution de probabilité induite sur l'ensemble des classes d'équivalence de fonctions booléennes : nous démontrons ici notre résultat principal.

## 6.2 Définition du modèle

### 6.2.1 Relations d'équivalence

Dans ce chapitre, les arbres et/ou sont des arbres binaires plans dont les nœuds internes sont étiquetés par $\wedge$ ou $\vee$, et dont les feuilles sont étiquetées par une variable de $\{x_i\}_{i\in\mathbb{N}}$ ou sa négation. La taille d'un arbre est le nombre de ses feuilles : bien entendu, il existe un nombre infini de tels arbres de taille $n$. Il est donc impossible de définir la distribution uniforme sur l'ensemble des arbres de taille $n$ comme dans le cas classique. Pour remédier à ce caractère infini, nous définissons des classes d'équivalence : nous dirons que deux arbres sont équivalents s'ils sont *identiques à renumérotation des variables près*. En plus de pouvoir ainsi définir un analogue de la distribution des arbres de Catalan sur l'ensemble des fonctions booléennes à un nombre non borné de variables, la famille des classes d'équivalence d'arbres et/ou formera une classe combinatoire, sur laquelle nous pourrons donc appliquer les méthodes de combinatoire analytique.

Une telle notion d'équivalence d'arbres est déjà définie dans Genitrini et al. [GKZ07, GK12], dans un cadre légèrement différent car le système logique étudié dans ces deux articles ne prend pas en compte les littéraux négatifs. Rappelons que nous appelons **squelette** un arbre binaire plan dont les noeuds internes sont étiquetés par des connecteurs et dont les feuilles sont non-étiquetées (cf. Définition 1.3.3). Rappelons aussi qu'une variable est un élément de $\{x_i\}_{i\geqslant 1}$ alors qu'un littéral est une variable ou sa négation, donc un élément de $\{x_i, \bar{x}_i\}_{i\geqslant 1}$. On dit qu'un nœud est étiqueté par la variable $x_i$ s'il est étiqueté par le littéral $x_i$ ou par le littéral $\bar{x}_i$. Ainsi, deux nœuds peuvent être étiquetés par la même variable mais par deux littéraux différents.

**Définition 6.2.1 (cf. Figure 6.1)**

*Deux arbres et/ou $t_1$ et $t_2$ sont équivalents si, et seulement si,*

*(i) leurs squelettes sont égaux,*

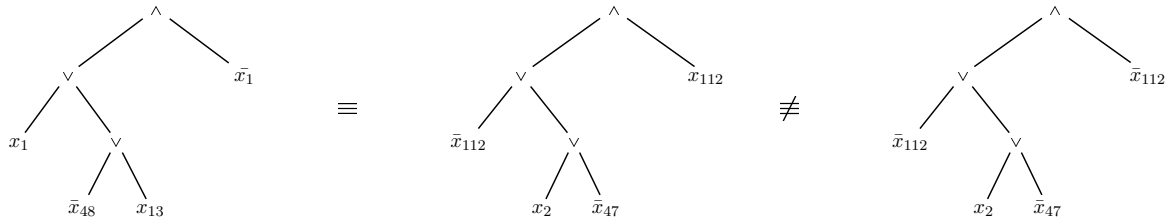
*(ii) deux feuilles sont étiquetées par la même variable dans $t_1$ si, et seulement si, elles sont étiquetées par la même variable dans $t_2$, et*

*(iii) deux feuilles sont étiquetées par le même littéral dans $t_1$ si, et seulement si, elles sont étiquetées par le même littéral dans $t_2$.*

**Remarque :** Cette remarque n'est pas redondante. Imaginons que nous supprimions la condition $(ii)$. Dès lors, l'expression $x_1 \wedge \bar{x}_1$ devient équivalente à l'expression $x_1 \wedge x_2$, ce qui n'est pas le cas. Par ailleurs, imaginions que nous supprimions la condition $(iii)$, alors, l'expression $x_1 \wedge \bar{x}_1$ devient équivalente à $x_1 \wedge x_1$, ce qui n'est pas le cas non-plus. Les conditions $(ii)$ et $(iii)$ ne sont donc pas identiques.

Cette relation d'équivalence induit (via $\Phi$, cf. Définition 1.2.5) une relation d'équivalence sur

FIGURE 6.1 – *Les deux arbres de gauche sont équivalents alors que les deux arbres de droite ne le sont pas.*




l'ensemble $\mathcal{F}_\infty$ des fonctions booléennes (cf. Définition 1.2.1). Par exemple, la fonction `Vrai` tout comme la fonction `Faux` sont les uniques éléments de leurs classes d'équivalence respectives et toutes les fonctions littéral sont équivalentes. Il est important de remarquer que les fonctions booléennes d'une même classe d'équivalence ont la même complexité et le même nombre de variables essentielles. On note $\langle f \rangle$ la classe d'équivalence de la fonction booléenne $f$.

### 6.2.2 Distribution de probabilité sur l'ensemble des classes d'équivalence de fonctions booléennes

Soit $(\mathbf{k_n})_{\mathbf{n \geqslant 1}}$ **une suite d'entiers croissante et tendant vers** $+\infty$ quand $n$ tend vers $+\infty$. Par la suite, nous supposerons qu'un arbre et/ou de taille $n$ ne peut contenir comme étiquettes plus de $k_n$ variables deux à deux distinctes, c'est à dire que, pour tout arbre $t$ de taille $n$, il existe un ensemble $\Gamma$ de variables de cardinal inférieur ou égal à $k_n$ tel que toute feuille de $t$ est étiquetée par une variable de $\Gamma$ ou par sa négation. Bien entendu, nous pouvons supposer $1 \leqslant k_n \leqslant n$, et ce pour tout entier $n \geqslant 1$.

**Définition 6.2.2**

*On note $T_n$ le nombre de classes d'équivalence d'arbres de taille $n$ (et tels qu'au plus $k_n$ variables différentes apparaissent comme étiquettes des feuilles). Soit $T(z)$ la série génératrice $T(z) = \sum_n T_n z^n$.*

Pour tout entier $n$, on note $\mathtt{Cat}_n$ le $n^{\text{ième}}$ nombre de Catalan (cf. Équation (1.1)), et pour tout couple d'entiers $(p, n)$, on note $\left\{ {n \atop p} \right\}$ le nombre de Stirling de seconde espèce associé à $n$ et $p$ (i.e. le nombre de partitions d'un ensemble à $n$ éléments en $p$ parties non-vides, voir par exemple [FS09, pages 735–737]).

**Proposition 6.2.3**

*Le nombre de classes d'équivalence d'arbres de taille $n$ vérifie :*

$$T_n = \mathtt{Cat}_{n-1} \cdot \sum_{p=1}^{k_n} \left\{ {n \atop p} \right\} 2^{2n-1-p}.$$

**Démonstration :** Le facteur $2^{n-1}\mathtt{Cat}_{n-1}$ compte le nombre de squelettes différents à $n$ feuilles : le terme $\mathtt{Cat}_{n-1}$ compte le nombre d'arbres non-étiquetés et le facteur $2^{n-1}$ compte le nombre d'étiquetages des nœuds internes. Il ne reste plus qu'à compter le nombre de façons d'étiqueter $n$ feuilles. Pour ce faire, choisissons un entier $p \in \{1, \ldots, k_n\}$ représentant le nombre de variables différentes apparaissant dans l'étiquetage des feuilles, puis partitionnons les $n$ feuilles en $p$ parties, signifiant que deux feuilles sont étiquetées par la même variable si, et seulement si, elles sont dans la même partie de la partition :

cela nous donne le facteur $\left\{ {n \atop p} \right\}$. Il ne reste plus qu'à choisir le signe de chaque étiquette de feuille : est-ce la variable elle-même ou sa négation ? Cela rajoute un terme $2^n$ que nous devons corriger par $2^{-p}$ car l'arbre obtenu en changeant tous les signes des feuilles d'une partie de la partition est équivalent à l'arbre de départ. ■

Étant donné un ensemble $\mathcal{S}$ de classes d'équivalence d'arbres et/ou, et en notant $S_n$ le nombre de ces classes dont les éléments sont de taille $n$, on appelle **proportion** ou **ratio** de $\mathcal{S}$ le réel suivant

$$\mu_n(\mathcal{S}) = \frac{S_n}{T_n}. \tag{6.1}$$

Étant donnée une fonction booléenne $f$, on note $T_n\langle f\rangle$ le nombre de classes d'équivalences d'arbres dont les éléments calculent une fonction de $\langle f\rangle$, et on appelle **probabilité** de $\langle f\rangle$ le ratio de cette famille de classes d'équivalence :

$$\mathbb{P}_n\langle f\rangle = \frac{T_n\langle f\rangle}{T_n}.$$

L'objectif de ce chapitre est d'étudier la distribution $\mathbb{P}_n$ [1] sur l'ensemble des classes d'équivalence de fonctions booléennes, notamment asymptotiquement quand $n$ tend vers $+\infty$.

### 6.2.3 Résultats

Nous décrivons dans ce chapitre le comportement de $\mathbb{P}_n\langle f\rangle$ pour toute fonction booléenne $f$ *fixée*, asymptotiquement quand $n$ tend vers $+\infty$.

**Définition 6.2.4**

*Soit $\langle f\rangle$ une classe d'équivalence de fonctions booléennes. On note $L\langle f\rangle$ (resp. $E\langle f\rangle$) la complexité commune (resp. le nombre de variables essentielles commun) des fonctions de $\langle f\rangle$. La* **multiplicité** *de $\langle f\rangle$, notée $R\langle f\rangle$, est le nombre entier $L\langle f\rangle - E\langle f\rangle \geqslant 0$ : c'est le nombre de répétitions dans un arbre minimal d'une fonction de $\langle f\rangle$.*

**Théorème 6.2.5**

*Soit $(k_n)_{n\geqslant 1}$ une suite d'entiers croissante et tendant vers $+\infty$ quand $n$ tend vers $+\infty$. Il existe une suite $(M_n)_{n\geqslant 1}$ telle que $M_n \sim \frac{n}{\ln n}$ (quand $n$ tend vers $+\infty$) et telle que, pour toute classe d'équivalence de fonctions booléennes $\langle f\rangle$ fixée, il existe une constante strictement positive $\lambda_{\langle f\rangle}$ telle que*

*(i) si, pour tout $n$ assez grand, $k_n \leqslant M_n$, alors, asymptotiquement quand $n$ tend vers $+\infty$,*

$$\mathbb{P}_n\langle f\rangle \sim \lambda_{\langle f\rangle} \cdot \left(\frac{1}{k_{n+1}}\right)^{R\langle f\rangle+1};$$

*(ii) si, pour tout $n$ assez grand, $k_n \geqslant M_n$, alors, asymptotiquement quand $n$ tend vers $+\infty$,*

$$\mathbb{P}_n\langle f\rangle \sim \lambda_{\langle f\rangle} \cdot \left(\frac{\ln n}{n}\right)^{R\langle f\rangle+1}.$$

1. Nous réutilisons ici une notation utilisée dans le Chapitre 5 pour un objet totalement différent : aucune confusion n'est cependant possible.

Remarquons tout d'abord que la constante $\lambda_{\langle f\rangle}$ est indépendante de $k_n$ (et de $n$, bien entendu), résultat qui a déjà été observé dans Genitrini et al. [GKZ07, GK12] dans le cas particulier de la fonction `Vrai`. Par ailleurs, ce théorème ne couvre pas tous les cas possibles : la suite $(k_n)_{n\geqslant 1}$ peut *osciller* entre valeurs plus petites et valeurs plus grandes que celles de la suite $(M_n)_{n\geqslant 1}$. Ceci dit, ce théorème couvre tous les cas *naturels* de suites $(k_n)_{n\geqslant 1}$ : l'idée est de prendre pour cette suite une suite très régulière du type $\ln n$, $\sqrt{n}$, $n^{3/4}$, etc.

Afin de pouvoir comparer nos résultats avec ceux concernant la distribution des arbres de Catalan (cf. Section 1.3.1), il nous faut traduire les résultats obtenus par Kozik (cf. Théorème 1.3.1) en terme de classes d'équivalence. Il suffit pour cela de remarquer que, si, pour tout $n \geqslant 0$, $k = k_n$, il y a $\binom{k}{E(f)}2^{E(f)}$ fonctions booléennes dans la classe d'équivalence de $f \in \mathcal{F}_k$. Dès lors, d'après le Théorème 1.3.1 (avec les notations de ce théorème), asymptotiquement quand $k$ tend vers $+\infty$,

$$\lim_{n\to+\infty} \mu_{n,k}\langle f\rangle = \sum_{g\in\langle f\rangle} \mu_{n,k}(g) = \Theta\left(\frac{1}{k^{L(f)-E(f)+1}}\right) = \Theta\left(\frac{1}{k^{R(f)+1}}\right).$$

Bien entendu, il serait abusif de considérer que la suite constante $(k_n = k)_{n\geqslant 0}$ tend vers $+\infty$ quand $n$ tend vers $+\infty$. Mais le résultat de Kozik étant vrai quand $k$ tend vers $+\infty$, il y a une certaine cohérence à supposer que le cas $k$ *fini* entre bien dans le cas $(i)$ de notre Théorème 6.2.5 (même si notre preuve ne s'appliquera pas à ce cas). Nous retrouvons bien le $\frac{1}{k}$ élevé à la puissance $R(f)+1$. Notre nouveau résultat est donc cohérent avec le modèle classique des arbres de Catalan.

Par ailleurs, le modèle étudié par Genitrini et al. [GKZ07, GK12] correspond a priori à la suite $(k_n)_{n\geqslant 0}$ constante égale à $+\infty$ (même si cette suite n'est pas autorisée dans ce chapitre). Remarquons cependant qu'un arbre à $n$ feuilles ne peut être étiqueté par plus de $n$ variables différentes : le cas où $(k_n)_{n\geqslant 0}$ est constante égale à $+\infty$ évoque donc le cas $k_n = n$ (pour tout entier $n$). Appliquer le Théorème 6.2.5 à la fonction `Vrai` permet de retrouver les résultats établis dans Genitrini et al [GKZ07, GK12].

Notre résultat est donc cohérent avec les deux cas *extrêmes* étudiés dans la littérature.

La preuve du Théorème 6.2.5 est l'objet de la suite du Chapitre. Elle se décompose en quatre parties : (1) une partie technique qui correspond à l'étude combinatoire du nombre $T_n$ de classes d'équivalence d'arbres de taille $n$ (cf. Section 6.3), (2) la généralisation de la théorie des motifs à notre nouveau modèle de classes d'équivalences (cf. Section 6.4), (3) l'étude de la probabilité de la classe d'équivalence de la fonction `Vrai` (cf. Sous-section 6.5.1), et (4) l'étude d'une classe d'équivalence de fonctions générales (cf. Sous-section 6.5.2).

## 6.3 Nombre de classes d'équivalence d'arbres

Cette partie est consacrée à une étude technique du comportement de $T_n$, asymptotiquement quand $n$ tend vers $+\infty$. Elle paraît un peu technique, mais les résultats obtenus seront fondamentaux pour la suite : c'est lors de cette étude que nous voyons apparaître le seuil d'ordre $\frac{n}{\ln n}$ observé dans le Théorème 6.2.5.

Rappelons que $T_n = 2^{2n-1}\mathtt{Cat}_{n-1}\sum_{p=1}^{k_n}\left\{{n \atop p}\right\}$ (cf. Proposition 6.2.3). Le facteur $2^{2n-1}\mathtt{Cat}_{n-1}$ est très bien connu : nous connaissons entre autres son comportement asymptotique : la difficulté de l'étude de $T_n$ vient donc du facteur somme. La proposition suivante, due à Comtet, et dont une preuve simple est établie par Sibuya [Sib88], nous permettra de borner ce facteur somme. Elle peut être vue comme un cas particulier des égalités de Bonferroni (cf. Comtet [Com74, Section 4.7]).

**Proposition 6.3.1 (Comtet [Com74], [Sib88])**

*Pour tout entier $n \geqslant 1$, pour tout $p \in \{1, \dots, n\}$,*

$$\frac{p^n}{p!} - \frac{(p-1)^n}{(p-1)!} \leqslant \left\{ {n \atop p} \right\} \leqslant \frac{p^n}{p!}.$$

Nous nous intéressons donc à la suite à double indice

$$a_p^{(n)} = \frac{p^n}{p!} 2^{-p}. \tag{6.2}$$

Lorsqu'aucune confusion ne sera possible, nous omettrons l'exposant $n$ et noterons $a_p$ au lieu de $a_p^{(n)}$. Montrons le lemme suivant

**Lemme 6.3.2**

*(i) La suite finie $(a_p)_{p\in\{1,\dots,n\}} = \left(\frac{p^n}{p!}2^{-p}\right)$ est unimodale. Plus précisément, pour tout $n \geqslant 1$, il existe un entier $M_n$ tel que $(a_p)_{p\in\{1,\dots,n\}}$ croît strictement sur $\{1, 2, \dots, M_n\}$ et décroît strictement sur $\{M_n + 1, \dots, n\}$.*

*(ii) La suite $(M_n)_{n\geqslant 0}$ est asymptotiquement croissante, et, asymptotiquement quand $n$ tend vers $+\infty$,*

$$M_n \sim \frac{n}{\ln n}.$$

**Démonstration :** $(i)$ Montrons que la suite $(a_p)_{1\leqslant p\leqslant n}$ est log-concave, i.e. que le suite $\left(\frac{a_{p+1}}{a_p}\right)_{1\leqslant p\leqslant n-1}$ est décroissante. Soit $p$ un entier de $\{1, \dots, n-1\}$. Par définition,

$$\frac{a_{p+1}}{a_p} = \left(\frac{p+1}{p}\right)^n \frac{1}{2(p+1)},$$

ce qui implique, pour tout $n \geqslant 0$,

$$\frac{a_{p+1}}{a_p} > 1 \Leftrightarrow n \ln\left(\frac{p+1}{p}\right) - \ln(2(p+1)) > 0.$$

Remarquons que la fonction $(\phi_n : p \mapsto n \ln\left(\frac{p+1}{p}\right) - \ln(2(p+1)))$ est strictement décroissante. Comme $\phi(1)$ tend vers $+\infty$ et $\phi(n-1)$ tend vers $-\infty$ quand $n$ tend vers $+\infty$, il existe un unique entier $M_n$ tel que $(a_p)_{p\in\{1,\dots,n\}}$ soit strictement croissante sur $\{1, \dots, M_n\}$ et strictement décroissante sur $\{M_n + 1, \dots, n\}$.

$(ii)$ Soit $x_n$ l'unique solution de l'équation

$$\left(\frac{x+1}{x}\right)^n \frac{1}{2(x+1)} = 1, \tag{6.3}$$

alors, $M_n = \lfloor x_n \rfloor$. Remarquons tout d'abord que la suite $(x_n)_{n\geqslant 1}$ est croissante. En effet, nous savons que $\phi_n(x_n) = 0$ et $\phi_{n+1}(x_{n+1}) = 0$. Dès lors, $\phi_n(x_{n+1}) = -\ln\left(1 + \frac{1}{x_{n+1}}\right) < 0$, ce qui implique, comme $\phi_n$ est décroissante, que $x_{n+1} \geqslant x_n$, et ce pour tout entier $n$ suffisamment grand. Dès lors, la suite $(M_n)_{n\geqslant 1}$ est asymptotiquement croissante.

De plus, comme, asymptotiquement quand $n$ tend vers $+\infty$,

$$\left(\frac{\frac{n}{\ln n}+1}{\frac{n}{\ln n}}\right)^n \frac{1}{2(\frac{n}{\ln n}+1)} \sim \frac{\ln n}{2} \geqslant 0,$$

alors $\frac{n}{\ln n} \leqslant x_n$, ce qui implique que $x_n$ tend vers $+\infty$ quand $n$ tend vers $+\infty$. Par ailleurs, l'Équation (6.3) évaluée en $x_n$ est équivalente à

$$n \ln\left(1 + \frac{1}{x_n}\right) = \ln 2 + \ln(x_n + 1), \tag{6.4}$$

ce qui implique $x_n \ln x_n \sim n$ quand $n$ tend vers $+\infty$. Nous en déduisons que $\ln x_n \sim \ln n$ puis que $x_n \sim \frac{n}{\ln n}$ quand $n$ tend vers $+\infty$. Comme $M_n = \lfloor x_n \rfloor$, nous en concluons que $M_n \sim n/\ln n$ asymptotiquement quand $n$ tend vers $+\infty$. ■

Extrayons de l'expression de $T_n$ le facteur dont l'analyse asymptotique semble être centrale :

**Définition 6.3.3**

*Pour toute suite $(u_n)_{n\geqslant 1}$, on pose, pour tout entier $n \geqslant 1$,*

$$B_{n,u_n} = \sum_{p=1}^{u_n} \left\{ {n \atop p} \right\} 2^{-p}.$$

*Dès lors, pour tout entier $n \geqslant 0$, $T_n = 2^{2n-1}\mathtt{Cat}_{n-1}B_{n,k_n}$ (cf. Proposition 6.2.3).*

Le lemme suivant explicite le comportement asymptotique de $B_{n,u_n}$ quand $n$ et $u_n$ tendent vers $+\infty$ : informellement, *avant le seuil, i.e. si $u_n \leqslant M_n$, $B_{n,u_n}$ est équivalent à la somme de quelques-uns de ses derniers termes, et après le seuil $M_n$, $B_{n,u_n}$ est équivalent à la somme de quelques termes autour du terme d'indice $M_n$*. Il faut retenir que les termes de plus grand poids de la suite $(a_p^{(n)})_{1\leqslant p\leqslant n}$ et donc les termes de plus grand poids de la suite $\left\{ {n \atop p} \right\} 2^{-p}$ sont les termes d'indices proches de $M_n$.

**Lemme 6.3.4**

*Soit $(u_n)_{n\geqslant 1}$ une suite croissante telle que $u_n \geqslant n$ pour tout entier $n \geqslant 1$ et $u_n$ tend vers $+\infty$ quand $n$ tend vers $+\infty$.*

*(i) si, pour tout $n$ assez grand, $u_n \leqslant M_n$, alors, pour toute suite $(\delta_n)_{n\geqslant 1}$ telle que $\delta_n = o(u_n)$ et $\frac{u_n\sqrt{\ln u_n}}{n} = o(\delta_n)$, nous avons, asymptotiquement quand $n$ tend vers $+\infty$,*

$$B_{n,u_n} = \Theta\left( \sum_{p=u_n-\delta_n}^{u_n} \frac{p^n}{p!} 2^{-p} \right) \tag{6.5}$$

*(ii) si, pour tout $n$ assez grand, $u_n \geqslant M_n$, alors, pour toute suite $(\delta_n)_{n\geqslant 1}$ telle que $\delta_n = o(u_n)$ et $\frac{u_n\sqrt{\ln u_n}}{n} = o(\delta_n)$, pour toute suite $(\eta_n)_{n\geqslant 1}$ telle que $\eta_n = o(M_n)$, $\lim_{n\to+\infty} \frac{\eta_n^2}{M_n} = +\infty$ et $\sqrt{M_n \ln(u_n - M_n)} = o(\eta_n)$, nous avons, asymptotiquement quand $n$ tend vers $+\infty$,*

$$B_{n,u_n} = \Theta\left( \sum_{p=M_n-\delta_n}^{\min\{M_n+\eta_n,u_n\}} \frac{p^n}{p!} 2^{-p} \right). \tag{6.6}$$

**Démonstration du Lemme 6.3.4, assertion** (6.5) **:** Au vu de la Proposition 6.3.1, nous pouvons borner $B_{n,u_n}$ : pour tout $n \geqslant 1$,

$$\frac{1}{2} \cdot \sum_{p=1}^{u_n-1} \frac{p^n}{p!\ 2^p} + \frac{u_n^n}{u_n!\ 2^{u_n}} \leqslant B_{n,u_n} \leqslant \sum_{p=1}^{u_n} \frac{p^n}{p!\ 2^p}. \tag{6.7}$$

Supposons $u_n \leqslant M_n$ pour tout $n$ assez grand et montrons que les deux bornes de l'Équation (6.7) sont du même ordre quand $n$ tend vers $+\infty$.

Notons, pour tout entier $N \geqslant 1$, $S_N = \sum_{p=1}^{N} a_p^{(n)}$ : l'Équation (6.7) implique

$$\frac{1}{2} S_{u_n} \leqslant B_{n,u_n} \leqslant S_{u_n}.$$

On sépare la somme $S_{u_n}$ en deux : les derniers $\delta_n$ termes, et les autres.

$$S_{u_n} = S_{u_n-\delta_n-1} + \sum_{p=u_n-\delta_n}^{u_n} a_p.$$

Rappelons que par hypothèse, $\delta_n = o(u_n)$ : nous pouvons donc choisir $n$ suffisamment grand de telle sorte que $u_n > \delta_n$. Montrons que $S_{u_n-\delta_n-1}$ est négligeable devant $a_{u_n}$, et donc devant $\sum_{p=u_n-\delta_n}^{u_n} a_p$. Rappelons que $(a_p)_{p\geqslant 1}$ est croissante sur $\{1,\dots,M_n\}$. Cela implique que

$$S_{u_n-\delta_n-1} \leqslant u_n a_{u_n-\delta_n}.$$

Pour tout entier $n$ suffisamment grand, via la formule de Stirling,

$$\begin{aligned}
\frac{a_{u_n-\delta_n}}{a_{u_n}} &= 2^{\delta_n} \left(\frac{u_n-\delta_n}{u_n}\right)^n \frac{u_n!}{(u_n-\delta_n)!} \\
&= \left(\frac{2u_n}{\mathsf{e}}\right)^{\delta_n} \left(\frac{u_n-\delta_n}{u_n}\right)^{n-u_n+\delta_n-1/2} (1+o(1)) \\
&= \exp\left[\delta_n \ln 2 - \delta_n + \delta_n u_n + (n-u_n+\delta_n-1/2)\ln\left(1-\frac{\delta_n}{u_n}\right) + o(1)\right].
\end{aligned}$$

Comme $\delta_n = o(u_n)$, nous avons $\ln\left(1-\frac{\delta_n}{u_n}\right) = -\frac{\delta_n}{u_n} - \frac{\delta_n^2}{2u_n^2}$, et

$$\begin{aligned}
\frac{a_{u_n-\delta_n}}{a_{u_n}} &= \exp\left[\delta_n \ln 2 - \delta_n + \delta_n u_n - \frac{n\delta_n}{u_n} + \delta_n - \frac{n\delta_n^2}{2u_n^2} + \mathcal{O}\left(\frac{n\delta_n^2}{u_n^2}\right)\right] \\
&= \exp\left[\delta_n \ln 2 + \delta_n u_n - \frac{n\delta_n}{u_n} - \frac{n\delta_n^2}{2u_n^2} + o\left(\frac{n\delta_n^2}{u_n^2}\right)\right],
\end{aligned}$$

car, par hypothèse, $\ln u_n = o\left(\frac{\delta_n^2}{u_n}\right)$, ce qui implique $\frac{n\delta_n^2}{u_n^2} = \Omega(\ln u_n)$. Rappelons que $u_n \leqslant M_n$, et, d'après (6.4), $\frac{n}{M_n} \geqslant \ln 2 + \ln M_n$. Dès lors,

$$\begin{aligned}
\frac{a_{u_n-\delta_n}}{a_{u_n}} &\leqslant \exp\left[\delta_n \ln 2 + \delta_n M_n - \frac{n\delta_n}{M_n} - \frac{n\delta_n^2}{2u_n^2} + o\left(\frac{n\delta_n^2}{u_n^2}\right)\right] \\
&\leqslant \exp\left[-\frac{n\delta_n^2}{2u_n^2} + o\left(\frac{n\delta_n^2}{u_n^2}\right)\right].
\end{aligned}$$

Comme $\frac{n\delta_n^2}{u_n^2} = \Omega(\ln u_n)$, nous en déduisons que

$$\frac{S_{u_n-\delta_n-1}}{a_{u_n}} \leqslant u_n \frac{a_{u_n-\delta_n}}{a_{u_n}} \leqslant \exp\left[\ln u_n - \frac{n\delta_n^2}{2u_n^2} + o\left(\frac{n\delta_n^2}{u_n^2}\right)\right] = o(1).$$

Dès lors, $S_{u_n} \sim \sum_{p=u_n-\delta_n}^{u_n} a_p$, ce qui conclut la preuve. ■

**Démonstration du Lemme 6.3.4, assertion** (6.6) **:** Supposons $u_n \geqslant M_n$ pour tout $n$ assez grand. Séparons les sommes des deux bornes de l'Équation (6.7) en trois parties : la première partie de l'indice 1 à l'indice $M_n - \delta_n - 1$ ; la seconde partie de l'indice $M_n - \delta_n$ à l'indice $M_n + \eta_n$ ; et la

troisième partie de $M_n + \eta_n + 1$ à $u_n$. Notons que, si $u_n \leqslant M_n + \eta_n$, alors, la troisième somme est vide et la deuxième somme n'est pas complète :

$$S_{u_n} = S_{M_n - \delta_n - 1} + \sum_{p=M_n-\delta_n}^{M_n+\eta_n} a_p + \sum_{p=M_n+\eta_n+1}^{u_n} a_p.$$

Par des arguments tout à fait similaires à ceux développés dans la preuve du $(i)$, nous pouvons montrer que $S_{M_n-\delta_n-1}$ est négligeable devant $a_{M_n}$, et donc devant $\sum_{p=M_n-\delta_n}^{M_n+\eta_n} a_p$. Dès lors, si $u_n \geqslant M_n + \eta_n$, l'assertion $(ii)$ est prouvée. Nous supposons maintenant que $u_n \geqslant M_n + \eta_n + 1$ : il nous reste à montrer que $\sum_{p=M_N+\eta_n+1}^{u_n} a_p$ est négligeable devant $a_{M_n}$, et donc devant $\sum_{p=M_n-\delta_n}^{M_n+\eta_n} a_p$ pour conclure la preuve.

Au vu du Lemme 6.3.2, nous avons

$$\sum_{p=M_n+\eta_n+1}^{u_n} a_p \leqslant (u_n - M_n - \eta_n) a_{M_n+\eta_n}.$$

Via la formule de Stirling,

$$\begin{aligned}
\frac{a_{M_n+\eta_n}}{a_{M_n}} &= 2^{-\eta_n} \left(\frac{M_n+\eta_n}{M_n}\right)^n \frac{M_n!}{(M_n+\eta_n)!} \\
&= \left(\frac{2(M_n+\eta_n)}{\mathsf{e}}\right)^{-\eta_n} \left(\frac{M_n+\eta_n}{M_n}\right)^{n-M_n-1/2} (1+o(1)) \\
&= \exp\left[-\eta_n \ln 2 + \eta_n - \eta_n \ln(M_n+\eta_n) + (n - M_n - 1/2)\ln\left(1+\frac{\eta_n}{M_n}\right) + o(1)\right].
\end{aligned}$$

Comme $\ln\left(1+\frac{\eta_n}{M_n}\right) \leqslant \frac{\eta_n}{M_n}$ et $\frac{\eta_n}{M_n} = o(1)$ par hypothèse, nous avons

$$\begin{aligned}
\frac{a_{M_n+\eta_n}}{a_{M_n}} &\leqslant \exp\left[-\eta_n \ln 2 + \eta_n - \eta_n \ln(M_n+\eta_n) + \frac{\eta_n}{M_n}(n - M_n - 1/2) + o(1)\right] \\
&= \exp\left[-\eta_n \ln 2 + \eta_n - \eta_n \ln(M_n+\eta_n) + \frac{n\eta_n}{M_n} - \eta_n + o(1)\right] \\
&= \exp\left[-\eta_n \ln 2 - \eta_n \ln(M_n+\eta_n) + \frac{n\eta_n}{M_n} + o(1)\right] \\
&= \exp\left[-\eta_n \ln 2 - \eta_n \ln M_n - \eta_n \ln\left(1+\frac{\eta_n}{M_n}\right) + \frac{n\eta_n}{M_n} + o(1)\right] \\
&= \exp\left[-\eta_n \ln 2 - \eta_n \ln M_n - \frac{\eta_n^2}{M_n} + \frac{n\eta_n}{M_n} + \mathcal{O}\left(\frac{\eta_n^3}{M_n^2}\right)\right]
\end{aligned}$$

Comme $M_n = \lfloor x_n \rfloor$, nous avons

$$n \ln\left(1+\frac{1}{x_n}\right) = n\left(\frac{1}{M_n} - \frac{1}{2M_n^2} + \mathcal{O}\left(\frac{1}{M_n^3}\right)\right),$$

et

$$\ln 2 + \ln(x_n + 1) = \ln 2 + \ln M_n + \mathcal{O}\left(\frac{1}{M_n}\right).$$

Au vu de l'Équation (6.4), cela implique

$$\frac{n}{M_n} = \ln 2 + \ln M_n + \frac{n}{2M_n^2} + \mathcal{O}\left(\frac{n}{M_n^3}\right) + \mathcal{O}\left(\frac{1}{M_n}\right) = \ln 2 + \ln M_n + \frac{n}{2M_n^2} + \mathcal{O}\left(\frac{n}{M_n^3}\right)$$

car $\frac{1}{M_n} = o(\frac{n}{M_n^3})$.

$$\begin{aligned}\frac{a_{M_n+\eta_n}}{a_{M_n}} &\leqslant \exp\left[-\frac{\eta_n^2}{M_n} + \mathcal{O}\left(\frac{\eta_n^3}{M_n^2}\right) + \mathcal{O}\left(\frac{n\eta_n}{M_n^3}\right)\right] \\ &= \exp\left[-\frac{\eta_n^2}{M_n} + o\left(\frac{\eta_n^2}{M_n}\right)\right],\end{aligned}$$

car, par hypothèse, $\frac{\eta_n^2}{M_n}$ tend vers $+\infty$ quand $n$ tend vers $+\infty$. Nous obtenons donc

$$\frac{\sum_{p=M_n+\eta_n+1}^{u_n} a_p}{a_{M_n}} \leqslant (u_n - M_n - \eta_n)\frac{a_{M_n+\eta_n}}{a_{M_n}} \leqslant \exp\left[\ln(u_n - M_n) - \frac{\eta_n^2}{M_n} + o\left(\frac{\eta_n^2}{M_n}\right)\right] = o(1)$$

car, par hypothèse, $\ln(u_n - M_n) = o\left(\frac{\eta_n^2}{M_n}\right)$. Dès lors, quand $n$ tend vers $+\infty$

$$S_{u_n} \sim \sum_{p=M_n-\delta_n}^{M_n+\eta_n} a_p,$$

ce qui conclut la preuve. ■

Nous pouvons désormais en déduire le comportement de $B_{n,k_n}$ :

**Lemme 6.3.5**

*Soit $(k_n)_{n\geqslant 1}$ une suite d'entiers croissante et tendant vers $+\infty$ quand $n$ tend vers $+\infty$. Supposons que $k_n \leqslant M_n$ pour tout $n$ assez grand. Alors, asymptotiquement quand $n$ tend vers $+\infty$,*

$$\frac{B_{n,k_{n+1}}}{B_{n+1,k_{n+1}}} = \Theta\left(\frac{1}{k_{n+1}}\right).$$

**Démonstration : Supposons tout d'abord que $k_{n+1} \leqslant M_n$.** Soit $(\delta_n)_{n\geqslant 1}$ une suite d'entiers vérifiant $\delta_n = o(k_{n+1})$ et $\frac{k_{n+1}\sqrt{\ln k_{n+1}}}{n} = o(\delta_n)$ quand $n$ tend vers $+\infty$. Le Lemme 6.3.4 appliqué à $u_n = k_{n+1}$ nous donne, asymptotiquement quand $n$ tend vers $+\infty$,

$$B_{n,k_{n+1}} = \Theta\left(\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n)}\right).$$

De même, comme $k_{n+1} \leqslant M_{n+1}$, et comme la suite $(\delta_{n-1})_{n\geqslant 1}$ vérifie $\delta_{n-1} = o(k_n)$, et $\frac{k_n\sqrt{\ln k_n}}{n} = o(\delta_{n-1})$, en appliquant le Lemme 6.3.4 à la suite $u_n = k_n$, nous obtenons, asymptotiquement quand $n$ tend vers $+\infty$,

$$B_{n,k_n} = \Theta\left(\sum_{p=k_n-\delta_{n-1}}^{k_n} a_p^{(n)}\right),$$

donc

$$B_{n+1,k_{n+1}} = \Theta\left(\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n+1)}\right).$$

Dès lors,

$$\mathsf{rat}_{n+1} := \frac{B_{n,k_{n+1}}}{B_{n+1,k_{n+1}}} = \Theta\left(\frac{\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n)}}{\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n+1)}}\right).$$

Nous avons

$$(k_{n+1}-\delta_n)\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n)} \leqslant \sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} p a_p^{(n)} = \sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n+1)} = \sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} p a_p^{(n)} \leqslant k_{n+1} \sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n)}.$$

Dès lors,

$$\mathsf{rat}_{n+1} = \frac{B_{n,k_{n+1}}}{B_{n+1,k_{n+1}}} = \Theta\left(\frac{1}{k_{n+1}}\right).$$

**Supposons, au contraire $M_{n+1} \geqslant k_{n+1} > M_n$.** Soit $(\delta_n)_{n\geqslant 1}$ une suite d'entiers telle que $\delta_n = o(k_{n+1})$ et $\frac{k_{n+1}\sqrt{\ln k_{n+1}}}{n+1} = o(\delta_n)$. Soit $(\eta_n)_{n\geqslant 1}$ une suite d'entiers telle que $\eta_n = o(M_n)$, $\lim_{n\to+\infty} \frac{\eta_n^2}{M_n} = +\infty$ et $\sqrt{M_n \ln(u_n - M_n)} = o(\eta_n)$. En appliquant le Lemme 6.3.4, assertion (6.6), à la suite $u_n = k_n$, nous obtenons

$$B_{n,k_{n+1}} = \Theta\left(\sum_{p=M_n-\delta_n}^{\min\{M_n+\eta_n,k_{n+1}\}} a_p^{(n)}\right).$$

De plus, comme $\delta_{n-1} = o(k_n)$ et $\frac{k_n\sqrt{\ln k_n}}{n} = o(\delta_{n-1})$, via le Lemme 6.3.4, assertion (6.5) appliqué à la suite $u_n = k_n$,

$$B_{n+1,k_{n+1}} = \Theta\left(\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n+1)}\right).$$

Remarquons comme ci-dessus que

$$(k_{n+1}-\delta_n)\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n)} \leqslant B_{n+1,k_{n+1}} \leqslant k_{n+1} \sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n+1)}.$$

De plus, comme $k_{n+1} \geqslant M_n$, via des arguments similaires à ceux utilisés dans la preuve du Lemme 6.3.4, assertion (6.6),

$$\sum_{p=k_{n+1}-\delta_n}^{k_{n+1}} a_p^{(n)} \sim \sum_{p=k_{n+1}-\delta_n}^{\min\{k_{n+1},M_n+\eta_n\}} a_p^{(n)} \sim B_{n,k_{n+1}}.$$

Dès lors, comme $\delta_n = o(k_{n+1})$,

$$\mathsf{rat}_{n+1} := \frac{B_{n,k_{n+1}}}{B_{n+1,k_{n+1}}} = \mathcal{O}\left(\frac{1}{k_{n+1}-\delta_n}\right) = \mathcal{O}\left(\frac{1}{k_{n+1}}\right).$$

De même,

$$\mathsf{rat}_{n+1} = \Omega\left(\frac{1}{k_{n+1}}\right),$$

ce qui conclut la preuve. ■

**Lemme 6.3.6**

*Supposons que $k_n \geqslant M_n$ pour tout $n$ assez grand. Alors, asymptotiquement quand $n$ tend vers $+\infty$,*

$$\frac{B_{n,k_{n+1}}}{B_{n+1,k_{n+1}}} = \Theta\left(\frac{\ln n}{n}\right).$$

**Démonstration :** Nous avons, par hypothèse, $k_{n+1} \geqslant M_{n+1}$, ce qui implique $k_{n+1} \geqslant M_n$. Soit $(\delta_n)_{n\geqslant 1}$ une suite d'entiers telle que $\delta_n = o(M_n)$ et $\frac{M_n\sqrt{\ln M_n}}{n} = o(\delta_n)$. Soit $(\eta_n)_{n\geqslant 1}$ une suite d'entiers telle que $\eta_n = o(M_n)$, $\lim_{n\to+\infty} \frac{\eta_n^2}{M_n} = +\infty$ et $\sqrt{M_n \ln(k_{n+1} - M_n)} = o(\eta_n)$. Nous pouvons donc appliquer le Lemme 6.3.4, assertion (6.6), à $u_n = k_{n+1}$ : asymptotiquement quand $n$ tend vers $+\infty$,

$$B_{n,k_{n+1}} = \Theta\left(\sum_{p=M_n-\delta_n}^{\min\{M_n+\eta_n,k_{n+1}\}} a_p^{(n)}\right).$$

De plus, comme la suite $(\delta_n)_{n\geqslant 1}$ vérifie $\delta_n = o(M_n)$ et $\frac{M_n\sqrt{\ln M_n}}{n} = o(\delta_n)$, et comme la suite $(\eta_n)_{n\geqslant 1}$ vérifie $\eta_n = o(M_n)$, $\lim_{n\to+\infty} \frac{\eta_n^2}{M_n} = +\infty$ et $\sqrt{M_n \ln(k_n - M_n)} = o(\eta_n)$, nous avons,

$$B_{n,k_n} = \Theta\left(\sum_{p=M_n-\delta_n}^{\min\{M_n+\eta_n,k_n\}} a_p^{(n)}\right),$$

et donc

$$B_{n+1,k_{n+1}} = \Theta\left(\sum_{p=M_{n+1}-\delta_{n+1}}^{\min\{M_{n+1}+\eta_{n+1},k_{n+1}\}} a_p^{(n+1)}\right).$$

Remarquons que

$$(M_{n+1} - \delta_n) \sum_{p=M_{n+1}-\delta_{n+1}}^{\min\{M_{n+1}+\eta_{n+1},k_{n+1}\}} a_p^{(n)} \leqslant B_{n+1,k_{n+1}} \leqslant (M_{n+1} + \eta_{n+1}) \sum_{p=M_{n+1}-\delta_{n+1}}^{\min\{M_{n+1}+\eta_{n+1},k_{n+1}\}} a_p^{(n)}.$$

De plus, comme $k_{n+1} \geqslant M_{n+1} \geqslant M_n$, via des arguments similaires à ceux utilisés dans la preuve du Lemme 6.3.4, assertion (6.6),

$$\sum_{p=M_{n+1}-\delta_{n+1}}^{\min\{M_{n+1}+\eta_{n+1},k_{n+1}\}} a_p^{(n)} \sim \sum_{p=M_{n+1}-\delta_{n+1}}^{\min\{M_n+\eta_n,k_{n+1}\}} a_p^{(n)}.$$

Nous devons donc comparer

$$S_n = \sum_{p=M_{n+1}-\delta_{n+1}}^{\min\{M_n+\eta_n,k_{n+1}\}} a_p^{(n)}$$

et

$$T_n = \sum_{p=M_n-\delta_n}^{\min\{M_n+\eta_n,k_{n+1}\}} a_p^{(n)}$$

et montrer que ces deux sommes sont équivalentes. Décomposons $S_n$ comme suit :

$$S_n = T_n + \sum_{p=\min\{M_n+\eta_n,k_{n+1}\}}^{\min\{M_{n+1}+\eta_{n+1},k_{n+1}\}} a_p^{(n)} - \sum_{p=M_n-\delta_n}^{M_{n+1}-\delta_{n+1}} a_p^{(n)}.$$

Le second terme est négligeable devant $T_n$, via la preuve du Lemme 6.3.4, assertion (6.6). Supposons que le troisième terme n'est pas vide : $M_{n+1} - \delta_{n+1} > M_n - \delta_n$ (si ce terme est nul, alors nous avons déjà prouvé $S_n \sim T_n$). Via le Lemme 6.3.2, comme $\frac{M_{n+1}}{M_n} = 1 + o(\frac{1}{M_n})$,

$$\begin{aligned}\sum_{p=M_n-\delta_n}^{M_{n+1}-\delta_{n+1}} a_p^{(n)} &\leqslant (M_{n+1} - \delta_{n+1} - M_n + \delta_n) a_{M_n-\delta_n}^{(n)} \\ &= (\delta_n - \delta_{n+1} + o(1)) a_{M_n-\delta_n}^{(n)} \\ &\leqslant (\delta_n + o(1)) a_{M_n-\delta_n}^{(n)} = o(a_{M_n}^{(n)})\end{aligned}$$

au vu de la preuve du Lemme 6.3.4, assertion (6.5). Dès lors, comme $a_{M_n}^{(n)} \leqslant T_n$, nous avons $S_n \sim T_n$ quand $n$ tend vers $+\infty$, ce qui implique

$$\frac{1}{M_{n+1}+\eta_{n+1}}\Theta(1) \leqslant \mathsf{rat}_{n+1} := \frac{B_{n,k_{n+1}}}{B_{n+1,k_{n+1}}} \leqslant \frac{1}{M_{n+1}-\delta_{n+1}}\Theta(1),$$

et comme $\eta_n = o(M_n)$ et $\delta_n = o(M_n)$, nous obtenons

$$\mathsf{rat}_{n+1} = \Theta\left(\frac{1}{M_{n+1}}\right) = \Theta\left(\frac{\ln n}{n}\right).$$

■

**Définition 6.3.7**

*On notera*

$$\mathsf{rat}_n = \frac{B_{n-1,k_n}}{B_{n,k_n}}$$

*Le comportement de ce ratio est déterminé par les Lemmes 6.3.5 et 6.3.6 : notons en particulier que* $\mathsf{rat}_n \to 0$ *quand* $n$ *tend vers* $+\infty$.

## 6.4 Généralisation de la théorie des motifs

Rappelons que l'on note $\mathcal{I}$ la famille des squelettes, $I_n$ le nombre de squelettes de taille $n$, et $I(z) = \sum_{n\geqslant 1} I_n z^n$ leur série génératrice. Par la méthode symbolique,

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

ce qui implique

$$I(z) = \frac{1-\sqrt{1-8z}}{4}.$$

La singularité dominante de $I(z)$ est donc $\frac{1}{8}$.

L'objet de cette Section est de généraliser la théorie des motifs de Kozik (cf. Section 1.3.1) à notre nouveau modèle. Le lemme suivant est une généralisation du lemme de Kozik 1.3.5 : il ne prend en compte que les répétitions, et non l'ensemble des restrictions car les restrictions ne sont plus pertinentes lorsque l'on parle de classes d'équivalence d'arbres.

**Lemme 6.4.1**

*Soit* $L$ *un langage de motifs non-ambigu et sous-critique pour la famille* $\mathcal{I}$ *des squelettes. Soit* $T_n^{[r]}$ *(resp.* $T_n^{[\geqslant r]}$*) le nombre de classes d'équivalence d'arbres ayant exactement (resp. au moins)* $r$ $L$*-répétitions. Dès lors, asymptotiquement quand* $n$ *tend vers* $+\infty$,

$$\frac{T_n^{[r]}}{T_n} = \mathcal{O}\left(\mathsf{rat}_n^r\right) \qquad \text{et} \qquad \frac{T_n^{[\geqslant r]}}{T_n} = \mathcal{O}\left(\mathsf{rat}_n^r\right).$$

**Démonstration :** Le nombre de classes d'équivalence d'arbres de taille $n$ ayant au moins $r$ $L$-répétitions est donné par

$$T_n^{[\geqslant r]} = \sum_{d=r+1}^{n} I_n(d) Lab(n, k_n, d, r),$$

où $I_n(d)$ est le nombre de classes de squelettes ayant $d$ $L$-feuilles de motif, et où $Lab(n, k_n, d, r)$ est le nombre de façons (à équivalence près) d'étiqueter les $n$ feuilles de ces squelettes de façon à obtenir au moins $r$ $L$-répétitions. Nous avons l'inégalité suivante :

$$Lab(n, k_n, d, r) \leqslant 2^n \cdot \sum_{j=1}^{r} \binom{d}{r+j} \left\{ {r+j \atop j} \right\} B_{n-r-j+1,k_n}.$$

En effet, le facteur $2^n$ correspond au choix du signe de chaque littéral sur chaque feuille ; l'indice $j$ représente le nombre de variables distinctes qui réalisent les $r$ répétitions ; le facteur binomial représente le nombre de façons de choisir les feuilles de motifs qui réalisent les $r$ répétitions ; le nombre de Stirling représente le nombre de façons de partitionner ces $r + j$ feuilles en $j$ parties ; et, pour finir, le facteur $B_{n-r-j+1,k_n}$ étiquette les $n - r - j + 1$ feuilles restantes. Cependant, nous comptons plusieurs fois certains étiquetages, d'où l'inégalité suivante, car la suite $(B_{m,k_n})_{m\geqslant 1}$ est croissante en $m$ :

$$T_n^{[\geqslant r]} \leqslant 2^n \cdot B_{n-r,k_n} \sum_{j=1}^{r} \left\{ {r+j \atop j} \right\} \sum_{d=r+j}^{n} I_n(d) \binom{d}{r+j}.$$

Soit $\ell(x, y)$ la fonction génératrice du langage de motifs $L$. Alors, pour tout $p \geqslant 0$,

$$\frac{z^p}{p!} \frac{\partial^p \ell}{\partial x^p}(z, I(z)) = \sum_{n=1}^{\infty} \sum_{d=1}^{\infty} I_n(d) \binom{d}{p} z^n,$$

ce qui implique

$$\frac{T_n^{[\geqslant r]}}{T_n} \leqslant \frac{B_{n-r,k_n}}{B_{n,k_n}} \sum_{j=1}^{r} \left\{ {r+j \atop j} \right\} \frac{[z^n] z^{r+j} \frac{\partial^{r+j} \ell}{\partial x^{r+j}}(z, I(z))}{[z^n] I(z)}.$$

Comme $z^{r+j} \frac{\partial^{r+j} \ell}{\partial x^{r+j}}(z, I(z))$ et $I(z)$ ont la même singularité car $L$ est sous-critique pour la famille $\mathcal{I}$, chaque terme de la somme converge vers une constante quand $n$ tend vers $+\infty$, ce qui implique que

$$\frac{T_n^{[r]}}{T_n} \leqslant \frac{T_n^{[\geqslant r]}}{T_n} = \mathcal{O}\left( \frac{B_{n-r,k_n}}{B_{n,k_n}} \right) = \mathcal{O}\left( \mathsf{rat}_n^r \right).$$

■

## 6.5 Comportement de la distribution de probabilité

Maintenant que nous avons généralisé le lemme de Kozik, nous pouvons commencer l'étude de la distribution de probabilité $\mathbb{P}_n$. Notre première étape est, comme pour tous les modèles d'arbres booléens aléatoires, l'étude de la fonction constante `Vrai` (et donc, par symétrie celle de la fonction `Faux`). Nous montrons ensuite le Théorème 6.2.5.

### 6.5.1 Tautologies

Pour étudier les tautologies, la stratégie est la même que dans le cadre de la distribution des arbres de Catalan : nous calculons l'équivalent de la fraction des tautologies simples (cf. Définition 1.3.6) quand $n$ tend vers $+\infty$, puis, nous montrons que, asymptotiquement quand $n$ tend vers $+\infty$, presque toute tautologie est simple. Rappelons que l'on note $\mathcal{ST}$ l'ensemble des classes d'équivalence de tautologies simples, et $ST_n$ le nombre de classes d'équivalence de tautologies simples de taille $n$.

**Lemme 6.5.1**

*La fraction (cf. Équation* (6.1)*) des tautologies simples vérifie, asymptotiquement quand $n$ tend*

*vers* $+\infty$,

$$\mu_n(\mathcal{ST}) = \frac{ST_n}{T_n} \sim \frac{3}{4}\mathsf{rat}_n.$$

*De plus, asymptotiquement quand* $n$ *tend vers* $+\infty$*, presque toute tautologie est simple.*

**Démonstration :** Calculons tout d'abord la fraction des tautologies simples. Tout comme dans le cas de l'étude de la distribution des arbres de Catalan, ou tout comme dans le chapitre 2 où nous utilisons la même approche pour les tautologies associatives et commutatives, considérons le langage de motifs non-ambigu $S = \bullet|S \vee S| \square \wedge \square$. Un arbre dont deux $S$-feuilles de motif sont étiquetées par une variable et sa négation est une tautologie simple.

La fonction génératrice de $S$ est

$$s(x, y) = \frac{1}{2}(1 - \sqrt{1 - 4(x + y^2)}).$$

La singularité $\rho = \frac{1}{8}$ de $I(z)$ vérifie $I(\rho) = \frac{1}{4}$. Soit $\varepsilon > 0$. Supposons $|x| \leqslant \frac{1}{8} + \varepsilon$ et $|y| \leqslant \frac{1}{4} + \varepsilon$. Dès lors, $|4(x + y^2)| \leqslant \frac{3}{4} + 6\varepsilon + 4\varepsilon^2$. Il est donc possible de choisir $\varepsilon$ de façon à ce que $S(x, y)$ soit analytique sur

$$\{(x, y) \in \mathbb{C}^2 \mid |x| \leqslant \frac{1}{8} + \varepsilon, |y| \leqslant \frac{1}{4} + \varepsilon\}.$$

Le langage $S$ est donc sous-critique pour la famille $\mathcal{I}$.

La fonction génératrice $\tilde{I}(z) = \frac{1}{2}\partial^2/\partial x^2(s(xz, I(z))_{|x=1}$ compte le nombre de squelettes dans lequels on a pointé deux $S$-feuilles de motif. Dès lors, $DC_n = 2^{n-1}\tilde{I}_n B_{n-1,k_n}$ est le nombre de tautologies simples, comptées plusieurs fois, une fois par paire de feuilles qui réalise la tautologie simple. Comme nous faisons du double-comptage, nous notons cette famille $\mathcal{DC}$, pour signifier que certaines tautologies simples sont comptées plus d'une fois. Comme $T_n = 2^{n+1}I_n B_{n,k_n}$,

$$\mu_n(\mathcal{DC}) = \frac{DC_n}{T_n} = \frac{2^{n-1}\tilde{I}_n B_{n-1,k_n}}{2^{n+1}I_n B_{n,k_n}},$$

ce qui implique, via le Lemme 1.5.1,

$$\lim_{n\to\infty} \frac{\tilde{I}_n}{I_n} = \lim_{z\to\frac{1}{8}} \frac{\tilde{I}'(z)}{I'(z)} = 3.$$

Nous obtenons donc

$$\mu_n(\mathcal{ST}) \leqslant \mu_n(\mathcal{DC}) \sim \frac{3}{4}\mathsf{rat}_n.$$

Pour obtenir une borne inférieure, il faut se concentrer sur le double-comptage que nous avons effectué. Dans le famille $\mathcal{DC}$ les tautologies réalisées par une unique paire de feuilles sont comptées exactement une fois, celles qui sont réalisées par deux paires de feuilles sont comptées deux fois, et ainsi de suite. Notons $\mathcal{ST}^i$ la famille des tautologies simples comptées au moins $i$ fois dans $\mathcal{DC}$. Dès lors, $DC_n = \sum_{i\geqslant 1} ST_n^i$.

Notre objectif est maintenant de soustraire à $DC_n$ les tautologies simples que nous avons surcomptées. Pour ce faire, comptons les tautologies simples réalisées par trois $S$-feuilles de motif étiquetées par $\alpha/\underline{\alpha}/\bar{\alpha}$ où $\alpha$ est un littéral, et les tautologies réalisées par quatre feuilles de motif étiquetées par $\alpha/\bar{\alpha}/\beta/\bar{\beta}$ où $\alpha$ et $\beta$ sont deux littéraux. Pour cela, notons

$$I_3(z) = \frac{1}{3!}\frac{\partial^3}{\partial x^3}s(xz, I(z))_{|x=1}$$

la série génératrice des squelettes dans lesquels trois $S$-feuilles de motif sont pointées, et

$$I_4(z) = \frac{1}{4!}\frac{\partial^4}{\partial x^4}s(xz, I(z))_{|x=1}$$

la série génératrice des squelettes dans lesquels quatre $S$-feuilles de motif sont pointées. Dès lors, notons

$$DC_n^{(3)} = 3 \cdot 2^{n-2} B_{n-2,k_n}[z^n] I_3(z),$$

et

$$DC_n^{(4)} = 6 \cdot 2^{n-2} B_{n-2,k_n}[z^n] I_4(z).$$

L'entier $DC_n^{(3)}$ compte (éventuellement plusieurs fois) les arbres dans lesquels trois $S$-feuilles de motif ont été pointées, deux d'entre elles étiquetées par un littéral et l'autre par sa négation. Le facteur 3 vient du choix des deux feuilles ayant la même étiquette. Nous savons que ces trois feuilles sont étiquetées par la même variable. L'étiquetage des feuilles de l'arbre est donc compté par $B_{n-2,k_n} 2^{n-2}$. L'entier $DC_n^{(4)}$ compte (éventuellement plusieurs fois) les arbres dans lesquels quatre $S$-feuilles de motif ont été pointées, deux d'entre elles étiquetées par deux littéral (associés à deux variables distinctes) et les deux autres par leurs négations. Notons qu'un arbre ayant six feuilles de motifs étiquetées respectivement par $\alpha/\alpha/\bar{\alpha}/\beta/\beta/\bar{\beta}$ est compté 2 fois par $DC_n^{(3)}$ et une fois par $DC_n^{(4)}$.

Notons que, pour tout entier $i$ une tautologie simple comptée au moins $i$ fois par $DC_n$ est comptée au moins $(i-1)$ fois par $DC_n^{(3)} + DC_n^{(4)}$. Dès lors,

$$ST_n \geqslant DC_n - (DC_n^{(3)} + DC_n^{(4)}).$$

De plus, comme le langage de motifs $S$ est sous-critique pour la famille des squelettes $\mathcal{I}$, nous avons

$$\frac{DC_n^{(3)}}{T_n} \leqslant c_3 \cdot \frac{B_{n-2,k_n}}{B_{n,k_n}} = c_3 \cdot \frac{B_{n-2,k_n}}{B_{n,k_n}} = \mathcal{O}(\mathsf{rat}_n^2)$$

et

$$\frac{DC_n^{(4)}}{T_n} \leqslant c_4 \cdot \frac{B_{n-2,k_n}}{B_{n,k_n}} = c_4 \cdot \frac{B_{n-2,k_n}}{B_{n,k_n}} = \mathcal{O}(\mathsf{rat}_n^2),$$

où $c_3$ et $c_4$ sont des constantes strictement positives.

Ainsi, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mu_n(\mathcal{ST}) = \mu_n(DC) + o\,(\mathsf{rat}_n) \sim 3/4 \cdot \mathsf{rat}_n.$$

Montrons désormais que, asymptotiquement quand $n$ tend vers $+\infty$, presque toute tautologie est simple : pour cela considérons le langage de motifs $N = \bullet | N \vee N | N \wedge \square$. Ce langage de motifs est non-ambigu et sa série génératrice est donnée par $\frac{1}{2}(1 - y - \sqrt{(1-y)^2 - 4x})$. Il est aisé de voir que $N$ est donc sous-critique pour la famille $\mathcal{I}$ des squelettes.

Nous faisons le même raisonnement que dans l'étude de la distribution de Catalan (cf. [Koz08] et Section 1.3.1). Une tautologie a au moins une $N$-répétition, car sinon, toutes ses $N$-feuilles de motif peuvent être assignées à `Faux` et l'arbre total calcule `Faux` pour cette affectation partielle des variables, ce qui est impossible pour une tautologie.

Nous pouvons montrer, comme dans le Chapitre 2, Section 2.3.2 (cf. preuve du Lemme 2.3.21) que toute tautologie admet au moins une $N[N]$-répétition, que toute tautologie ayant exactement une $N[N]$-répétition est une tautologie simple, et que la famille des arbres ayant au moins 2 $N[N]$-répétitions est négligeable devant celle des tautologies simples. ■

Nous avons donc montré le Théorème 6.2.5 pour les fonctions de complexité 0 : asymptotiquement quand $n$ tend vers $+\infty$,

$$\mathbb{P}_n(\texttt{Vrai}) = \mathbb{P}_n(\texttt{Faux}) \sim \frac{3}{4}\mathsf{rat}_n,$$

où le comportement asymptotique de $\mathsf{rat}_n$ est déterminé par les Lemmes 6.3.5 et 6.3.6 selon les propriétés de la suite $(k_n)_{n\geqslant 1}$ et sa position par rapport à la suite $(M_n)_{n\geqslant 1}$.

**Remarque :** Notons que la répétitivité des deux fonctions constantes `Vrai` est *Faux* est égale à 0 car leur complexité est 0 et qu'elle n'ont pas de variables essentielles.

### 6.5.2 Cas général

Tout comme dans l'étude de la distribution des arbres de Catalan, nous allons montrer que, asymptotiquement quand $n$ tend vers $+\infty$, presque tout arbre calculant $\langle f \rangle$ est une expansion d'un arbre minimal de $\langle f \rangle$. Les expansions que nous considérons dans ce chapitre sont les T-expansions et les X-expansions (cf. Proposition 2.5.3). Les preuves seront similaires à celles développées dans le Chapitre 2 : à partir du moment où la théorie des motifs s'applique, toutes les preuves s'adaptent sans difficulté.

Dans toute la suite, $\langle f \rangle$ sera une classe d'équivalence fixée (et $f$ un représentant de cette classe) de complexité $r$. Rappelons les définitions des langages de motifs $N$ et $P$ :

$$N = \bullet | N \vee N | N \vee \square$$
$$P = \bullet | P \wedge P | P \vee \square.$$

Si toutes les $N$-feuilles de motifs d'un arbre sont affectées à `Faux`, l'arbre complet restreint à cette affectation partielle des variables calcule la fonction `Faux`, et si toutes les $P$-feuilles de motif d'un arbre sont affectées à `Vrai`, alors l'arbre complet restreint à cette affectation partielle des variables calcule la fonction `Vrai`. Nous considérons les langages de motifs $L = N^{(r+1)}[N \oplus P]$ and $\bar{L} = N^{(r+1)}[(N \oplus P)^2]$, qui sont tous deux non-ambigus et sous-critiques pour la famille de squelettes $\mathcal{I}$, car $N$ et $P$ le sont.

La première étape de la preuve du Théorème 6.2.5 est la proposition suivante :

**Proposition 6.5.2**

*Soit $f$ une fonction booléenne fixée et soit $\Gamma_f$ l'ensemble de ses variables essentielles. Un arbre $t$ calculant $f$ et ayant au moins une $L$-feuille de motif de niveau $(r+2)$, a au moins $R(f)+1$ $(L, \Gamma_F)$-restrictions (on rappelle que $r$ est défini par $r = L(f)$).*

La preuve de cette proposition est identique à celle développée dans [Koz08] :

**Démonstration :** Supposons que $t$ calcule $f$, a au moins une $L$-feuille de motif de niveau $(r+2)$ et au plus $R(f)$ $L$-répétitions. Soit $i$ le plus petit entier tel que le nombre de $(N^{(i)}, \Gamma_f)$-restrictions est égal au nombre de $(N^{(i-1)}, \Gamma_f)$-restrictions. Si un tel $i$ n'existe pas dans $\{1, \dots, r+2\}$, on posera $i = +\infty$.

Il y a au moins une restriction parmi les $L$-feuilles de motifs : s'il n'y en a pas, nous pouvons affecter toutes les $N$-feuilles de motifs de $t$ à `Faux` sans changer la fonction calculée par l'arbre, ce qui implique que $f \equiv$ `Faux`, ce qui est absurde. Dès lors, $i \leqslant r+1$.

**Premier cas :** Supposons que $t$ contient au plus $(r-1)$ $(N^{(i)}, \Gamma_f)$-restrictions. Nous savons qu'il n'y a ni répétition, ni variable essentielle parmi les $L$-feuilles de motif de niveau $i$. Nous pouvons donc affecter ces feuilles à `Faux` et ainsi, tous les arbres greffés dans les emplacements de niveau $(i-1)$ calculent `Faux`, et l'arbre total calcule toujours $f$. Affectons de plus la valeur `Faux` à toutes les variables non-essentielles qui ne sont pas déjà affectées. On simplifie l'arbre de façon à faire disparaître les feuilles étiquetées par des constantes `Vrai` ou `Faux`. Nous obtenons un arbre noté $t^\star$, qui calcule toujours $f$, dont les feuilles sont les anciennes $N^{(i-1)}$-feuilles de motif qui étaient étiquetées par des variables essentielles dans $t$. Dès lors, $t^\star$ a au plus $(r-1)$ feuilles, ce qui est impossible car $f$ est de complexité $r$.

**Second cas :** Supposons que $t$ a exactement $r$ $(N^{(i)}, \Gamma_f)$-restrictions. Comme $i \leqslant r+1$, il n'y a pas de restriction dans les arbres greffés dans les emplacements de niveau $r$. Nous pouvons donc remplacer ces emplacements par des $\star$, signifiant ainsi que ces nouvelles feuilles étiquetées par $\star$ peuvent être assignées à `Vrai` ou `Faux`, indépendamment les unes des autres, et sans changer la fonction calculée par $t$. Nous remplaçons aussi par $\star$ les feuilles étiquetées par des variables non-essentielles et non-répétées.

Nous simplifions ces $\star$ (cf. règles de simplifications (2.14) et (2.16)) : une telle simplification supprime au moins une feuille non $\star$. Si cette feuille est étiquetée par une variable essentielle mais

non-répétée, alors $t^\star$ ne dépend plus de cette variable de $f$ mais calcule toujours $f$ : c'est absurde. Dès lors, la feuille supprimée est une répétition : $t^\star$ a donc au plus $R(f)$ répétitions, ce qui est impossible. ■

La généralisation du lemme de Kozik (cf. Lemme 6.4.1) ne concerne que les répétitions, et non les restrictions, qui ne sont pas pertinentes en terme de classes d'équivalence. Le lemme suivant est nécessaire pour montrer le Théorème 6.2.5 : il traite le cas des restrictions.

**Lemme 6.5.3**

*Soit $L$ un langage de motifs non ambigu, sous-critique pour $\mathcal{I}$. Soient $f$ une fonction booléenne, $\mathcal{M}_f$ l'ensemble de ses arbres minimaux et $\Gamma_f$ l'ensemble des ses variables essentielles. Soit $E_p(\mathcal{M}_f)$ la famille des expansions d'arbres minimaux de $f$ obtenus en greffant un arbre ayant exactement $p$ $(L,\Gamma_f)$-restrictions. Dès lors, il existe une constante $\alpha > 0$ telle que*

$$\mu_n\langle E_p(\mathcal{M}_f)\rangle \sim \alpha \cdot \mathsf{rat}_n^{R(f)+p}.$$

**Démonstration :** Soit $E_n$ le nombre d'arbres de taille $n$ de $E_p(\mathcal{M}_f)$. On notera $i$ le nombre de feuilles réalisant les $p$ $(L,\Gamma_f)$-restrictions de l'expansion : $p+1 \leqslant i \leqslant 2p$. L'égalité suivante n'est qu'un équivalent car certains arbres sont double-comptés, mais ce double comptage est négligeable (ce fait est admis ici) :

$$\mu_n(E_p(\mathcal{M}_f)) \sim \frac{E_n}{T_n} = \sum_{i=p+1}^{2p} [z^{n-L(f)}]\frac{\partial^i}{i!\partial x^i}\left(\ell(xz, I(z))\right)_{|x=1} \frac{2^n B_{n-p-R(f),k_n}}{2^n I_n B_{n,k_n}}.$$

Comme $L$ est sous-critique pour $\mathcal{I}$, il existe une constante $\alpha > 0$ telle que

$$\sum_{i=p+1}^{2p} \frac{[z^{n-L(f)}]\frac{\partial^i}{i!\partial x^i}\left(\ell(xz, I(z))\right)_{|x=1}}{I_n} \sim \alpha \cdot \frac{I_{n-L(f)}}{I_n} \sim \alpha \left(\frac{1}{8}\right)^{L(f)} > 0$$

asymptotiquement quand $n$ tend vers $+\infty$. Dès lors, via la Section 6.3,

$$\mu_n\langle E_p(\mathcal{M}_f)\rangle \sim \alpha \cdot \mathsf{rat}_n^{R(f)+p}.$$

■

Considérons la famille des $T$-expansions des arbres minimaux de $f$. Comme toute tautologie a au moins une $N$-répétition, d'après le Lemme 6.5.3,

$$\frac{E_n}{T_n} \sim \alpha \cdot \mathsf{rat}_n^{R(f)+1},$$

et donc,

$$\mathbb{P}_n\langle f\rangle = \Omega\left(\mathsf{rat}_n^{R(f)+1}\right).$$

Au vu du Lemme 6.4.1, nous savons que la famille des classes d'équivalence d'arbres calculant $\langle f\rangle$ ayant au moins $R(f)+2$ $L$-répétitions est négligeable devant $\mathsf{rat}_n^{R(f)+1}$. Nous savons donc que

$$\mathbb{P}_n\langle f\rangle = \Theta\left(\mathsf{rat}_n^{R(f)+1}\right),$$

ce qui n'est pas tout à fait suffisant pour conclure la preuve du Théorème 6.2.5.

Montrons que presque tout arbre calculant $f$ est une expansion d'un arbre minimal de $f$.

**Lemme 6.5.4**

*Le ratio des classes d'équivalence d'expansions d'arbres minimaux d'une fonction de $\langle f \rangle$ vérifie, asymptotiquement quand $n$ tend vers $+\infty$,*

$$\mu_n\langle E[\mathcal{M}_f]\rangle = cst \cdot \mathsf{rat}_n^{R(f)+1} + o\left(\mathsf{rat}_n^{R(f)+1}\right),$$

*où $cst$ est une constante strictement positive.*

Ce lemme est une conséquence directe du Lemme 6.5.3 car les arbres greffés lors d'une T-expansion ou lors d'une X-expansion admettent au moins une $(N, \Gamma_f)$-restriction.

**Lemme 6.5.5**

*Soit $\langle f \rangle$ une classe d'équivalence de fonctions booléennes. Asymptotiquement quand $n$ tend vers $+\infty$,*

$$\mathbb{P}_n\langle f\rangle \sim \mu_n(E[\mathcal{M}_f]).$$

**Démonstration :** Soit $t$ un arbre calculant $f$. Un tel arbre doit avoir au moins $R(f)+1$ $\bar{L}$-répétitions. De plus, au vu du Lemme 6.4.1, la famille des classes d'équivalence d'arbres ayant au moins $R(f)+2$ $\bar{L}$-répétitions est négligeable. Montrons que les arbres ayant exactement une $R(f)+1$ $\bar{L}$-répétitions sont exactement les expansions d'arbres minimaux de $f$. Cette preuve est très similaire à celle développée pour les arbres associatifs plans dans le Chapitre 2, Section 2.5.1.

Supposons que la taille $n$ de $t$ est assez grande : dès lors, $t$ a au moins $R(f)+1$ $L$-répétitions (cf. Proposition 6.5.2). Dès lors, les $\bar{L}$-feuilles de niveau $(r+3)$ n'ajoutent pas de nouvelle répétition.

Soit $i$ le plus petit entier tel qu'il y ait autant de $(N^{(i)}, \Gamma_f)$-restrictions que de $(N^{(i-1)}, \Gamma_f)$-restrictions. Comme il doit y avoir une restriction au premier niveau, et qu'il n'y a que $r+1$ restrictions, $i \leqslant r+1$.

**Premier cas :** Supposons qu'une variable essentielle $\alpha$ apparaisse parmi les feuilles de motif de niveau $(r+3)$. Dès lors, $t$ a au plus $L(f)$ $(N^{(i)}, \Gamma_f)$-restrictions. Remplaçons les emplacements du niveau $(i-1)$ par `Faux` et affectons toutes les variables non-essentielles non encore affectées à `Faux` (par exemple). Simplifions l'arbre et notons $t^\star$ l'arbre obtenu après simplification. Les feuilles de ce nouvel arbre sont les $N^{(i-1)}$-feuilles de motif de $t$ qui étaient étiquetées par des variables essentielles de $f$, et $t^\star$ calcule toujours $f$. Durant la simplification, nous avons dû supprimer au moins une $N^{(i)}$-feuille de motif étiquetée par une variable essentielle $\beta$, mais comme $t^\star$ calcule $f$, cette variable essentielle apparaît toujours parmi les feuilles de $t^\star$, et était donc répétée dans $t$. De plus, la variable $\alpha$ est essentielle pour $f$ : elle apparaît donc toujours parmi les feuilles de $t^\star$. En supprimant son occurrence parmi les feuilles de motif de niveau $(r+3)$, nous avons supprimé deux répétitions : $t^\star$ a donc au plus $R(f)-1$ répétitions (au vu de la Proposition 6.5.2), ce qui est impossible (notons que ce raisonnement reste valide si $\alpha = \beta$).

**Second cas :** Supposons qu'il n'y ait pas de variable essentielle parmi les feuilles de motif de niveau $(r+3)$. Comme il n'y a pas non plus de répétition à ce niveau, nous pouvons remplacer les emplacements de niveau $(r+2)$ par des $\star$, signifiant ainsi qu'ils peuvent être affectés à `Vrai` ou `Faux`, indépendamment les uns des autres, et sans changer la fonction calculée par l'arbre $t$. Remplaçons les variables non-essentielles et non-répétées restantes par $\star$. Nous pouvons ensuite simplifier les $\star$ et obtenir un arbre simplifié $t^\star$. L'arbre $t^\star$ est un arbre et/ou dont les feuilles sont les anciennes $R$-feuilles de motif de $t$, essentielles ou répétées. Pendant l'étape de simplification, nous avons simplifié aussi au moins une de ces $L$-feuilles de motif : $t^\star$ a au plus $r = L(f)$ feuilles, c'est donc un arbre minimal de $f$.

Nous pouvons montrer que le dernier ancêtre commun des $\star$ a été simplifié pendant le processus de simplification : supposons qu'il ne l'ait pas été, alors deux $\star$ ont été simplifiées indépendamment,

et, au vu des règles de simplifications (2.14) et (2.16), au moins deux $N^{(i)}$-feuilles de motif essentielles ou répétées ont été simplifiées, ce qui implique que $t^\star$ est de taille $L(f)-1$, ce qui est absurde.

Notons $t_e$ le sous-arbre enraciné en $\nu$, le dernier ancêtre commun des $\star$. Nous avons montré que, asymptotiquement quand $n$ tend vers $+\infty$, presque tout arbre calculant $f$ est une expansion d'un arbre minimal de $f$.

Il ne nous reste plus qu'à montrer que, asymptotiquement quand $n$ tend vers $+\infty$, presque tout arbre calculant $f$ est une T-expansion ou une X-expansion d'un arbre minimal de $f$. Supposons tout d'abord que $t_e$ (défini ci-dessus) n'a pas de $((N \oplus P), \Gamma_f)$-restriction. Dès lors, nous pouvons remplcer $t_e$ par une $\star$, puis simplifier cette $\star$. Cette simplification donne un arbre de taille strictement inférieure à $L(f)$ calculant $f$. C'est absurde. La famille des classes d'équivalence d'expansions d'arbres minimaux de $f$ telles que l'arbre greffé $t_e$ a au moins 2 $((N \oplus P), \Gamma_f)$-restrictions est négligeable au vu du Lemme 6.5.3. Nous pouvons donc supposer que $t_e$ a exactement une $((N \oplus P), \Gamma_f)$-restriction. Si cette restriction est une répétition, nous pouvons montrer que $t_e$ est alors une tautologie ou une contradiction et si c'est une variable essentielle de $f$, alors, nous pouvons montrer que $t$ est une X-expansion d'un arbre minimal de $f$. ■

## 6.6 Conclusion

Nous avons défini et étudié dans ce chapitre un nouveau modèle d'arbres booléens dans lequel le nombre de variables utilisées pour l'étiquetage de l'arbre dépend de la taille de cet arbre : nous avons défini une suite $k_n$, croissante et qui tend vers $+\infty$ quand $n$ tend vers $+\infty$, puis avons considéré la famille des arbres de taille $n$ et tels qu'au plus $k_n$ variables différentes apparaissent comme étiquettes des feuilles d'un même arbre.

Ces arbres, modulo une relation d'équivalence, induisent une distribution de probabilité sur $\mathcal{F}_\infty$. Nous avons étudié le comportement de cette distribution de probabilité quand $n$ tend vers $+\infty$, et avons montré que $\mathbb{P}_n\langle f\rangle$ se comporte en $\lambda_{\langle f\rangle} \cdot \mathsf{rat}_n^{R(f)+1}$, où le comportement de $\mathsf{rat}_n$ quand $n$ tend vers $+\infty$ dépend de la suite $k_n$. En résumé, l'idée est que, si $k_n \leqslant {}^{n}\!/_{\ln n}$, alors $\mathsf{rat}_n = \Theta\left({}^{1}\!/_{k_n}\right)$, et si $k_n \geqslant \frac{n}{\ln n}$, alors $\mathsf{rat}_n = \Theta\left({}^{\ln n}\!/_{n}\right)$.

Notre résultat est très général : nous avons restreint notre étude à des suites $(k_n)_{n\geqslant 1}$ *raisonnables* au sens où ces suites sont croissantes et tendent vers $+\infty$ quand $n$ tend vers $+\infty$. Nous montrons un théorème global qui met en évidence un phénomène de saturation en $k_n \sim {}^{n}\!/_{\ln n}$ : la distribution sur $\mathcal{F}_\infty$ obtenue pour $k_n = {}^{\ln n}\!/_{n}$ et celle obtenue pour $k_n = n$ ont des comportements asymptotiques similaires. De plus, quelle que soit la suite $(k_n)_{n\geqslant 1}$ d'entiers, tendant vers $+\infty$, presque toute tautologie est simple, asymptotiquement quand $n$ tend vers $+\infty$, et ce comme dans le modèle classique des arbres de Catalan.

# Conclusion et perspectives

Nous avons étudié dans cette partie différents modèles d'arbres booléens aléatoires : les arbres non binaires, non plans, l'arbre bourgeonnant (issu de l'arbre binaire de recherche aléatoire), et les arbres et/ou binaires plans de taille $n$ étiquetés sur $k(n)$ variables. Dans tous ces modèles, ainsi que dans la littérature, nous pouvons observer une certaine universalité de la distribution induite sur l'ensemble des fonctions booléennes. Deux comportements sont observés, manifestement caractérisés par le niveau de saturation de l'arbre aléatoire considéré.

Nous avons montré (cf. Théorème 5.5.2) que si le niveau de saturation de l'arbre sous-jacent tend vers $+\infty$ en probabilité quand la taille de l'arbre $n$ tend vers $+\infty$, alors, la distribution induite sur l'ensemble des fonctions booléennes est dégénérée, au sens ou elle ne charge que les fonctions constantes expressibles dans le système logique choisi. La réciproque de cette assertion est vraie, elle-aussi. Plus précisément, nous n'avons montré ce théorème que dans le cadre des arbres et/ou binaires plans, et dans le cadre du modèle suivant : soit $(T_n)_{n\geqslant 0}$ une suite d'arbres binaires aléatoires (non étiquetés), soit $(\hat{T}_n)_{n\geqslant 0}$ la suite de ces arbres après étiquetage aléatoire uniforme dans le système et/ou à $k$ variables, soit $f_{\hat{T}_n}$ la fonction booléenne représentée par l'arbre booléen aléatoire $\hat{T}_n$, si le niveau de saturation de $T_n$ tend vers $+\infty$ en probabilité quand $n$ tend vers $+\infty$, alors la distribution de cette fonction converge vers la distribution qui donne probabilité $1/2$ à la fonction constante `Vrai` et à la fonction constante `Faux`.

Pouvons-nous étendre ce résultat à d'autres systèmes logiques ? à des familles d'arbres plus larges (non binaires, non plans) ? à un étiquetage aléatoire non uniforme ? Ces généralisations semblent en effet assez raisonnables et leur preuve semble accessible. Il serait intéressant de réussir à obtenir un résultat le plus général possible permettant de caractériser les modèles qui induisent une distribution de probabilité dégénérée sur l'ensemble des fonctions booléennes. Par ailleurs, nous pouvons nous inspirer de l'étude du Chapitre 6 pour essayer de montrer cette dégénérescence, si elle a lieu, dans un modèle où l'étiquetage de l'arbre aléatoire $T_n$ est uniforme sur $k(n)$ variables. Ce dernier modèle, à la croisée des Chapitres 5 et 6 est inexploré à ce jour : comment se comporte, asymptotiquement quand $n$ tend vers $+\infty$, la distribution induite sur l'ensemble des fonctions booléennes par l'arbre bourgeonnant de taille $n$ étiqueté uniformément au hasard avec $k(n)$ variables ? Cette distribution est-elle dégénérée ?

Par ailleurs, le Théorème 5.5.2 nous assure que si le niveau de saturation de l'arbre $T_n$ ne tend pas vers $+\infty$ en probabilité quand $n$ tend vers $+\infty$, alors la distribution induite sur les fonctions booléennes par $T_n$ quand $n$ tend vers $+\infty$, si cette limite existe, n'est pas dégénérée. Au vu des différents modèles étudiés dans la littérature et dans ce mémoire, ces distributions non-dégénérées ont toutes le même comportement : elles donnent plus de poids aux fonctions de petite complexité. Plus précisément, notons $\mu_{k,n}$ la distribution induite par l'arbre $T_n$ étiqueté uniformément au hasard sur $k$ variables ($k$ indépendant de $n$), supposons que $\mu_k = \lim_{n\to+\infty} \mu_{n,k}$ existe. Alors, il semble raisonnable de conjecturer qu'il existe une constante $c \in \mathbb{N}$ telle que, asymptotiquement quand $k$

tend vers $+\infty$,

$$\mu_k(f) = \Theta\left(\frac{1}{k^{L(f)+c}}\right),$$

où $L(f)$ désigne la complexité de la fonction $f$ dans le modèle étudié. Quelles sont les bonnes hypothèses sur la famille d'arbres $(T_n)_{n\geqslant 0}$ nous assurant de la convergence vers une distribution limite $\mu_k$ ? sous quelles hypothèses cette conjecture est-elle vraie ? et comment prouver une telle conjecture ?

Finir l'étude du modèle d'arbres associatifs dont la taille est comptée en termes de nœuds permettrait de confirmer cette conjecture, et le fait que ce modèle semble se résoudre via des méthodes distinctes de celles utilisées dans les autres modèles peut nous donner des indices sur la marche à suivre en toute généralité. Par exemple, la théorie des motifs développée par Kozik pour les arbres de Catalan et que nous avons généralisée tout au long de cette partie à différents modèles d'arbres (non binaires, non plans, $k$ dépendant de $n$) ne semble pas être une approche universelle.

Par ailleurs, de nouveaux modèles d'arbres pourraient être étudiés pour confirmer cette conjecture. En effet, si l'on se restreint au cas binaire plan, seuls deux modèles d'arbres permettent de confirmer notre conjecture dans le cas non dégénéré : le modèle des arbres de Catalan et le modèle de Galton-Watson. L'arbre de Ford, introduit par Ford en 2005 [For05] en vue de modéliser des arbres phylogénétiques, est un processus d'arbres aléatoires à paramètre $\alpha \in [0,1]$ qui peut être vu comme une généralisation de l'algorithme de Rémy. Dans l'algorithme de Rémy, à chaque étape, on fait bourgeonner une arête tirée uniformément au hasard dans l'arbre. Dans le modèle de Ford, le tirage n'est plus fait uniformément au hasard, mais en pondérant par $\alpha$ les arêtes internes et par $1-\alpha$ les arêtes externes. Pour $\alpha = 0$, ce processus est l'arbre bourgeonnant, pour $\alpha = \frac{1}{2}$ c'est l'arbre de Catalan, et pour $\alpha = 1$, c'est le processus déterministe du peigne. La hauteur de cet arbre est d'ordre $n^\alpha$ [HMPW08], et il peut être montré que son niveau de saturation est d'ordre constant dès que $\alpha > 0$. Autrement dit, la distribution induite sur les fonctions booléennes, si sa limite existe quand la taille des arbres tend vers $+\infty$, est non-dégénérée. Parvenir à étudier ce modèle permettrait de confirmer ou infirmer notre conjecture pour toute une famille, indexée par $\alpha$, d'arbres aléatoires.

Enfin, la question de l'effet Shannon reste ouverte pour le plupart des modèles présentés dans ce mémoire (cf. Chapitres 2 et 4) : il serait intéressant de progresser dans cette étude, et de développer une approche générale pour cette question.

Au delà de ces nombreuses perspectives en matière de logique quantitative, un domaine voisin est celui des expressions arithmétiques aléatoires. Au lieu d'étiqueter les nœuds internes des arbres considérés par des connecteurs logiques, pourquoi ne pas les étiqueter par des opérations arithmétiques, $+$, $\times$, $-$, etc. Par exemple, le modèle de l'arbre bourgeonnant avec un étiquetage $+$, $-$ est abordé dans la thèse de Nguyen The [Ngu04]. L'idée la plus proche de la logique quantitative serait de considérer des arbres dont les nœuds internes sont étiquetés par min ou max, et dont les feuilles sont étiquetées par des variables $\{x_1, \ldots, x_k\} \in [0,1]^k$ ou par leurs compléments $\{\bar{x}_1, \ldots, \bar{x}_k\}$ où, pour tout $i \in \{1, \ldots, k\}$, $\bar{x}_i = 1 - x_i$. Un tel arbre représente une fonction de $[0,1]^k$ dans $[0,1]$ : quelle est la distribution de cette fonction selon le modèle d'arbres aléatoires choisi ?

# Deuxième partie

# Urnes de Pólya

# Chapitre 7

# Introduction

## 7.1 Contexte

Une urne de Pólya est un processus aléatoire décrit comme suit : une urne contient des boules noires et des boules rouges ; à chaque étape, nous piochons au hasard une de ces boules, regardons sa couleur, la remettons dans l'urne et rajoutons un certain nombre de boules noires et rouges déterminé par une règle pré-établie et par la couleur de la boule piochée. Quelle est la composition de l'urne après $n$ étapes ? à l'infini ?

L'urne de Pólya originelle a été introduite par Pólya et Eggenberger pour modéliser des phénomènes de contagion : à chaque étape, on rajoute $S$ boules de la couleur de la boule piochée. Cette urne originelle a été largement étudiée (cf. Eggenberger et Pólya [EP23], Blackwell et Kendall [BK64]), notamment dans son extension naturelle à $d$ couleurs. Asymptotiquement, le vecteur composition de l'urne à l'étqpae $n$ (dont la $i^{\text{ième}}$ coordonnée est le nombre de boules de couleur $i$ dans l'urne) converge en loi vers un vecteur de loi de Dirichlet dont les paramètres sont connus, quand $n$ tend vers $+\infty$.

Friedman [Fri49] généralise le modèle de Pólya-Eggenberger à deux couleurs de la façon suivante : à chaque étape, nous ajoutons dans l'urne $a$ boules de la couleur de la boule piochée et $c$ boules de l'autre couleur. Au delà des approches par combinatoire de Friedman, Freedman [Fre65] développe des résultats asymptotiques concernant ce modèle : si $\sigma = \frac{a-c}{a+c} < 1/2$ (autrement dit si $a < 3c$), nous dirons dans ce cas que l'urne est *petite*, le comportement asymptotique du vecteur composition est gaussien, et la limite est indépendante de la composition initiale de l'urne.

Motivés par les applications en informatique fondamentale, notamment aux structures de données, Bagchi et Pal [BP85] montrent, via la méthode des moments, l'universalité d'un comportement gaussien pour les *petites urnes* générales. Leur modèle est le suivant : lorsque l'on pioche une boule rouge, on rajoute $a$ boules rouges et $b$ boules noires à l'urne, si l'on tire une boule noire, on ajoute $c$ boules rouges et $d$ boules noires à l'urne. Sous l'hypothèse de balance ($a + b = c + d$, le nombre de boule dans l'urne est donc déterministe), le comportement de ces urnes semble universel : dès que $\sigma = \frac{a-c}{a+b} < 1/2$, le vecteur composition converge en loi vers un vecteur gaussien indépendant de la composition initiale de l'urne. Ces urnes plus générales donnent lieu à une littérature variée : Gouet [Gou97] généralise à des urnes à $d$ couleurs, Smythe [Smy96] à des règles de remplacement non déterministes. Une contribution d'importance à l'étude des urnes de Pólya est celle d'Athreya et Karlin [AK68] qui plongent les urnes de Pólya en temps continu, obtenant ainsi des processus de branchement multitypes. Ce plongement en temps continu leur permet d'obtenir des théorèmes limites pour *petites urnes* ($\sigma \leqslant 1/2$) et *grandes urnes* ($\sigma > 1/2$) de Friedman. L'article de Janson [Jan04], basé sur le plongement en temps continu, est une des contributions essentielles

à l'étude du comportement asymptotique des urnes à $d$ couleurs : ce comportement est décrit en détail aussi bien en temps continu qu'en temps discret, pour petites et grandes urnes.

Plus récemment, Flajolet et al. [FGP05] développent des méthodes alternatives de combinatoire analytique pour l'étude des ces urnes : cette étude permet de décrire l'évolution de l'urne par un système différentiel vérifié par des séries génératrices. Ce système ne peut être résolu en toute généralité, mais sa résolution dans certains cas particuliers précis permet d'obtenir des résultats fins comme des théorèmes de limite locale. Une dernière approche, développée par Pouyanne [Pou08] exploite la description algébrique de l'urne à $d$ couleurs pour en déduire des théorèmes asymptotiques.

Comme évoqué précédemment, les urnes de Pólya sont des modèles pertinents en informatique fondamentale, notamment en ce qui concerne les structures de données arborescentes telles les arbres 2-3 de recherche [FGP05], les arbres $m$-aires de recherche [CP04, FK05], les AVL (arbres binaires de recherche rééquilibrés par rotation) [Mah98].

Dans ce mémoire, nous nous intéressons aux *grandes urnes* dont l'asymptotique fait intervenir une variable aléatoire $W$ assez méconnue pour l'instant. Il a déjà été montré par Chauvin et al. [CPS11], dans le cadre des urnes à deux couleurs en temps continu, que la variable $W$ admet une densité sur $\mathbb{R}$, et que sa transformée de Laplace a un rayon de convergence nul. Cette étude utilise de l'analyse de Fourier très précise et aboutit au calcul explicite de la transformée de Fourier de $W$. De nombreuses zones d'ombres persistent cependant : quel est l'ordre des moments de cette variable aléatoire ? est-elle déterminée par ses moments ? comment se comporte sa transformée de Fourier au voisinage de zéro ? en l'infini ? Par ailleurs, l'étude de Chauvin et al. se restreint au modèle d'urne plongé en temps continu : que pouvons-nous dire de la variable $W$ issue du processus en temps discret ?

Contrairement aux *petites urnes*, la limite du processus d'une grande urne dépend de la composition initiale de l'urne. Ainsi, nous devons étudier toute une famille $W_{(\alpha,\beta)}$ de variables aléatoires indicées par $(\alpha, \beta)$, signifiant que l'urne contient initialement $\alpha$ boules rouges et $\beta$ boules noires. Nous montrerons comment la structure arborescente de l'urne nous permet de réduire l'étude à deux variables aléatoires : $W_{(1,0)}$ et $W_{(0,1)}$, et nous montrerons que ces deux variables aléatoires sont solutions d'un système d'équations en loi. Comme cette étude peut être faite aussi bien en temps discret qu'en temps continu, nous obtenons deux systèmes en loi pour deux couples $(W_{(1,0)}, W_{(0,1)})$ différents.

Notre but est d'appliquer à ce système d'équations en loi des méthodes développées dans la littérature pour l'étude des équations de point fixe, ou *smoothing equations* en anglais. Les équations de point fixe sont l'objet d'une vaste littérature. Une revue de littérature est disponible dans l'article d'Aldous et Bandyopadhyay [AB05]. Des équations de points fixes apparaissent notamment dans l'étude des processus de branchement (cf. Liu [Liu99], Biggins et Kiprianou [BK05] et Alsmeyer et al. [ABM12]). Il se trouve qu'une urne de Pólya en temps continu est un processus de branchement multitype : il n'est pas étonnant que des systèmes de point fixe d'équations en loi apparaissent dans leur étude. Les équations de point fixe apparaissent aussi dans l'étude de cascades de Mandelbrot [Man74, BM10]. Enfin, ces équations sont aussi largement utilisées dans l'étude d'algorithmes récursifs (comme l'étude de Quicksort par Rösler [Rös92]) et de structures de données : une revue de littérature sur le sujet est disponible par Rösler et Rüschendorf [RR01], ou Neininger et Rüschendorf [NR06].

Dans ce mémoire, nous nous inspirons plus particulièrement des méthodes de Liu [Liu99], reposant principalement sur l'analyse de transformées de Fourier. Ces méthodes ont déjà été adaptées par Chauvin et al. [CLP12b, CLP12a] à une urne très particulière issue de l'analyse de l'arbre $m$-aire de recherche. Dans ce mémoire, nous nous attacherons à décrire un cas plus général à $d$ couleurs. Il est intéressant de noter qu'indépendamment, Knape et Neininger [KN13] utilisent les équations de point fixe dans l'étude des urnes de Pólya. Leur centre d'intérêt n'est pas la variable $W$ et leurs

conclusions sont donc différentes de celles de ce mémoire : leur étude donne une idée de preuve alternative de certains théorèmes limites de Janson [Jan04] aussi bien dans le cas des *petites urnes* que des *grandes urnes*.

Le Chapitre 8 concerne les urnes à deux couleurs : les résultats de ce chapitre, obtenus en collaboration avec Brigitte Chauvin et Nicolas Pouyanne (UVSQ, France), sont à paraître Journal of Theoritical Probability. Nous montrons dans ce chapitre que les variables lois des variables $W$ sont déterminées par leurs moments, aussi bien en temps discret qu'en temps continu. Nous montrons en outre que la série de Laplace de la variable $W$ issue du processus d'urne en temps discret admet un rayon de convergence infini. Nous montrons que la variable $W$ issue du processus en temps discret admet une densité, proposant au passage une preuve alternative de ce même résultat en temps continu (déjà prouvé par Chauvin et al. [CPS11]). Les méthodes de point fixe et d'analyse de transformées de Fourier que nous détaillons dans ce chapitre, en plus de montrer la détermination par les moments, se généralisent à des urnes à $d$ couleurs comme nous le verrons dans le Chapitre 9.

Nous présentons dans la suite de cette introduction quelques résultats évoqués ci-dessus qui seront utiles dans la suite de cette partie consacrée aux grandes urnes de Pólya : nous résumons tout d'abord les résultats asymptotiques obtenus dans la littérature concernant les urnes à deux couleurs, introduisons les variables aléatoires $W$, sujets de cette partie, puis détaillons l'étude de l'urne originelle de Pólya qui aura un rôle par la suite.

## 7.2 Préliminaires

Une urne de Pólya est décrite par un vecteur composition initiale $U(0)$, et par une matrice de remplacement $R$

$$U(0) = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{et} \quad R = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

où $\alpha, \beta \geqslant 0$ sont tels que $\alpha + \beta \neq 0$ et $a, b, c, d \in \mathbb{Z}$. Cela signifie que l'urne contient initialement $\alpha$ boules noires et $\beta$ boules rouges ; et qu'à chaque étape, on tire uniformément au hasard une boule dans l'urne, on regarde sa couleur, on la remet dans l'urne, et on ajoute $a$ boules rouges et $b$ boules noires si elle était rouge, ou $c$ boules rouges et $d$ boules noires si elle était noire. Notons que si, $a, b, c$ ou $d$ sont négatifs, alors on retire $-a, -b, -c$ ou $-d$ boules au lieu d'en ajouter.

Nous ferons trois hypothèses classiques dans ce mémoire :

- Les urnes considérées seront équilibrées, i.e. $a + b = c + d = S$ où l'entier $S$ sera appelée **balance** de l'urne. Cela signifie qu'à l'étape $n$, il y a $\alpha + \beta + nS$ boules dans l'urne.
- Nous supposerons la **non-extinction** de l'urne, i.e. la probabilité que l'urne finisse vide est égale à zéro. La plupart des résultats de la littérature (cf. Janson [Jan04]) restent vrais conditionnellement à la non-extinction. Pour plus de simplicité, nous nous restreindrons à considérer des urnes pour lesquelles la probabilité d'extinction est nulle (une discussion sur les urnes dont la probabilité d'extinction est nulle peut être lue dans le livre de Mahmoud [Mah08]).
- Nous supposerons que les urnes sont **irréductibles**, i.e. $bc \neq 0$. Les urnes triangulaires, aussi appelées réductibles peuvent être étudiées (cf. Gouet [Gou93] et [Jan06]), mais nécessitent un traitement particulier que nous ne détaillons pas ici.

On note $U^{DT}_{(\alpha,\beta)}(n) =^t (R_n, B_n)$ (où $^t$ représente la transposée du vecteur et où l'exposant $DT$ nous rappelle que le processus d'urne considéré est en temps discret) la composition de l'urne à l'étape $n$, c'est à dire le nombre de boules noires $B_n$ et le nombre de boules rouges $R_n$ qu'elle contient. Le but de l'étude d'une urne est de décrire ce vecteur composition aléatoire au temps $n$.

La balance $S$ est valeur propre de la matrice $R$ et donc de sa transposée $^tR$. Notons $m \leqslant S$ la seconde valeur propre de $^tR$. Soit $v_1$ un vecteur propre de $^tR$ associé à $S$ et $v_2$ un vecteur propre

associé à $m$. Nous noterons $(u_1, u_2)$ la base duale de $(v_1, v_2)$ : $u_1$ est une projection sur l'espace propre associé à $S$, et $u_2$ est une projection sur l'espace propre associé à $m$. Plusieurs choix de vecteurs propres sont possibles : nous faisons le choix canonique suivant

$$v_1 = \frac{S}{b+c}\begin{pmatrix} c \\ b \end{pmatrix} \qquad v_2 = \frac{S}{b+c}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

ce qui implique, pour tout $(x, y) \in \mathbb{R}^2$

$$u_1(x, y) = \frac{x+y}{S} \qquad u_2(x, y) = \frac{bx - cy}{S}.$$

Comme le montre le théorème suivant, le comportement asymptotique d'une urne dépend du rapport entre ses deux valeurs propres $\sigma = \frac{m}{S}$.

**Théorème 7.2.1 (cf. Janson [Jan04] ou Pouyanne [Pou08] par exemple)**

– *Si $\sigma < \frac{1}{2}$, on dit que l'urne est* **petite**, *et on a le théorème limite suivant :*

$$\frac{U^{DT}_{(\alpha,\beta)}(n) - nv_1}{\sqrt{n}} \to G$$

*en loi, quand $n$ tend vers $+\infty$, avec $G$ un vecteur Gaussien centré, de matrice de covariance*

$$\Sigma^2 = \frac{1}{1-2\sigma}\frac{bcm^2}{(b+c)^2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

– *Si $\sigma = \frac{1}{2}$, on dit aussi que l'urne est petite, et on a le théorème limite suivant :*

$$\frac{U^{DT}_{(\alpha,\beta)}(n) - nv_1}{\sqrt{n \ln n}} \to G,$$

*en loi, quand $n$ tend vers $+\infty$, avec $G$ un vecteur Gaussien centré, de matrice de covariance*

$$\Sigma^2 = \frac{bc}{4}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

– *Si $\sigma > \frac{1}{2}$, on dit que l'urne est* **grande**, *et on a le théorème limite suivant :*

$$U^{DT}_{(\alpha,\beta)}(n) = nv_1 + n^\sigma W^{DT}_{(\alpha,\beta)} v_2 + o(n^\sigma) \tag{7.1}$$

*presque sûrement et dans tous les $L^p$, $p \geqslant 1$, quand $n$ tend vers l'infini.*

Il est intéressant de remarquer que la limite du processus ne dépend pas des conditions initiales dans le cas d'une petite urne, alors qu'elle en dépend, a priori, dans le cas des grandes urnes. Nous nous intéresserons dans ce mémoire à la variable aléatoire $W^{DT}_{(\alpha,\beta)}$ : quel est son support ? admet-elle une densité ? quels sont ses moments ?

Il est classique depuis les travaux d'Athreya et Karlin [AK68] de plonger le processus discret de l'urne de Pólya en temps continu. A l'instant initial, il y a $\alpha$ boules rouges et $\beta$ boules noires dans l'urne. Chacune de ces boules est équipée d'une horloge qui sonnera au bout d'un temps aléatoire de loi exponentielle de paramètre 1, et ce indépendamment des autres. Lorsque l'horloge d'une boule

sonne, cette boule se divise en $a+1$ boules rouges et $b$ boules noires si elle était rouge, ou en $c$ boules rouges et $d+1$ boules noires si elle était noire. On note $U^{CT}_{(\alpha,\beta)}(t)$ la composition de l'urne au temps $t$.

On notera $\tau_n$ la date de la $n^{\text{ème}}$ sonnerie. Comme, via les propriétés de la loi exponentielle, la première horloge qui sonne parmi $p$ (pour tout entier $p \geqslant 1$) horloges est tirée *uniformément* parmi les $p$ horloges, nous avons la relation suivante entre le processus en temps discret et le processus en temps continu :

$$(U^{DT}(n))_{n\geqslant 0} = (U^{CT}(\tau_n))_{n\geqslant 0}, \tag{7.2}$$

presque sûrement, et, de plus, la suite des temps d'arrêts $(\tau_n)_{n\geqslant 0}$ est indépendante de $(U^{CT}(\tau_n))_{n\geqslant 0}$. C'est cette relation qui nous permettra de traduire tout résultat obtenu en temps discret pour le processus en temps continu, et vice-versa. Notons que le plongement en temps continu a déjà été appliqué dans ce mémoire à l'arbre bourgeonnant, ou arbre binaire de recherche aléatoire (cf. Chapitre 5), dont le plongement en temps continu est appelé arbre de Yule. Nous avions alors la même connexion entre processus en temps discret et processus en temps continu. Ce plongement en temps continu nous avait autorisé à travailler sur le processus en temps continu, plus *simple*, avant de traduire nos résultats en temps discret (cf. Section 5.2.3).

Le processus d'urne de Pólya en temps continu a lui aussi longuement été étudié dans la littérature : nous rappelons le théorème limite suivant, concernant les grands urnes :

**Théorème 7.2.2 (cf. Janson [Jan04])**

*Si $\sigma > 1/2$, asymptotiquement quand $t$ tend vers $+\infty$,*

$$U^{CT}_{\alpha,\beta}(t) = \mathsf{e}^{St}\xi v_1(1+o(1)) + \mathsf{e}^{mt}W^{CT}_{(\alpha,\beta)}v_2(1+o(1))$$

*presque sûrement et dans tous les $L^p$, $p \geqslant 1$. La variable aléatoire $\xi$ est connue : elle suit une loi Gamma$\left(\frac{\alpha+\beta}{S}\right)$.*

Nous nous intéressons dans ce mémoire aux variables aléatoires $W^{DT}_{(\alpha,\beta)}$ et $W^{CT}_{(\alpha,\beta)}$. Ces variables, qui apparaissent dans les Théorèmes 7.2.1 et 7.2.2, sont définies comme limites de martingales : $W^{DT}_{(\alpha,\beta)}$, est, à une constante près, la limite de la martingale

$$u_2\left(\frac{U^{DT}_{(\alpha,\beta)}(n)}{\prod_{j=1}^{n}\left(1+\frac{\sigma}{\frac{\alpha+\beta}{S}+j-1}\right)}\right),$$

et vérifie

$$W^{DT}_{(\alpha,\beta)} = \lim_{n\to+\infty} u_2\left(\frac{U^{DT}_{(\alpha,\beta)}(n)}{n^{\sigma}}\right). \tag{7.3}$$

Par ailleurs,

$$W^{CT}_{(\alpha,\beta)} = \lim_{t\to+\infty} u_2\left(\frac{U^{CT}_{(\alpha,\beta)}(t)}{\mathsf{e}^{mt}}\right). \tag{7.4}$$

Via ces définitions comme limites de martingales, et via la relation (7.2), nous déduisons deux égalités en loi, que nous appellerons **connexions** :

$$W^{CT}_{\alpha,\beta} \overset{(loi)}{=} \xi^{\sigma}\cdot W^{DT}_{\alpha,\beta}, \tag{7.5}$$

avec $\xi$ de loi Gamma$\left(\frac{\alpha+\beta}{S}\right)$, et $\xi$ et $W_{\alpha,\beta}^{DT}$ indépendantes ; et

$$W_{\alpha,\beta}^{DT} \overset{(loi)}{=} \xi^{-\sigma} \cdot W_{\alpha,\beta}^{CT},$$

avec $\xi$ de loi Gamma$\left(\frac{\alpha+\beta}{S}\right)$, mais $\xi$ et $W_{\alpha,\beta}^{CT}$ *non* indépendantes.

En conséquence, si nous avons des informations sur $W^{CT}$, nous en déduirons des informations sur $W^{DT}$, et réciproquement. Comme évoqué précédemment, il est montré par Chauvin et al. [CPS11] que la variables aléatoires $W^{CT}$ admet une densité quelle que soit la composition initiale de l'urne $(\alpha, \beta)$ et que, de plus, la transformée de Laplace de $W^{CT}$ a un rayon de convergence nul, ce qui implique, que, pour toute constante $C > 0$, pour tout entier $p_0$ il existe $p \geqslant p_0$ tel que,

$$C^p \leqslant \frac{\mathbb{E}[(W^{CT})^p]}{p!}.$$

De tels résultats n'existent pas dans la littérature concernant la variable aléatoire $W^{DT}$.

## 7.3 L'urne "originelle"

L'urne originelle est une urne à $d \geqslant 2$ couleurs, représentée par la matrice de remplacement $SI_d$ (ou $I_d$ est la matrice identité en dimension $d$), et de composition initiale $(\alpha_1, \ldots, \alpha_d)$. Le théorème limite suivant est standard : un énoncé légèrement différent est prouvé par Athreya [Ath69], il est prouvé dans la cas particulier $S = 1$ et $\alpha_1 = \ldots = \alpha_d = 1$ par Blackwell et Kendall [BK64], et la preuve par méthode des moments que nous présentons ici pour sa simplicité et son autonomie est évoquée mais non développée dans le livre de Johnson et Kotz [JK97].

**Théorème 7.3.1**

*Si $U(n)$ est le vecteur composition de l'urne originelle au temps $n$, alors,*

$$\frac{U(n)}{nS} \to V$$

*presque sûrement et dans tous les $L^p$, $p \geqslant 1$, quand $n$ tend vers l'infini, où $V$ est un vecteur aléatoire de Dirichlet de paramètres $\left(\frac{\alpha_1}{S}, \ldots, \frac{\alpha_d}{S}\right)$.*

Il est utile de rappeler que les marginales d'un vecteur aléatoire de Dirichlet de paramètres $\left(\frac{\alpha_1}{S}, \ldots, \frac{\alpha_d}{S}\right)$ sont des lois Bêta de paramètres respectifs $\left(\frac{\alpha_k}{S}, \sum_{j\neq k} \frac{\alpha_j}{S}\right)$, pour tout $k \in \{1, \ldots, d\}$.

Mais avant tout, détaillons la définition et quelques propriétés utiles de la loi de Dirichlet. Soit $d \geqslant 2$ un entier. Soit $\Sigma$ le simplexe de dimension $(d-1)$ :

$$\Sigma = \left\{ (x_1, \ldots, x_d) \in [0,1]^d, \ \sum_{j=1}^{d} x_j = 1 \right\}.$$

Nous avons l'égalité suivante, généralisation de la définition de la fonction Bêta d'Euler : pour tous entiers non-nuls $\nu_1, \ldots, \nu_d$,

$$\int_\Sigma \prod_{j=1}^{d} x_j^{\nu_j - 1} d\Sigma(x_1, \ldots, x_d) = \frac{\Gamma(\nu_1) \ldots \Gamma(\nu_d)}{\Gamma(\nu_1 + \ldots + \nu_d)} \tag{7.6}$$

où $d\Sigma$ est la mesure positive sur le simplexe $\Sigma$, définie comme suit : pour toute fonction $f$ définie sur $\Sigma$,

$$f(x_1,\dots,x_d)d\Sigma(x_1,\dots,x_d) = f\left((x_1,\dots,x_{d-1},1-\sum_{j=1}^{d-1}x_j\right)\mathbb{1}_{\left\{x\in[0,1]^{d-1},\ \sum_{j=1}^{d-1}x_j\leqslant 1\right\}}dx_1\dots dx_{d-1}.$$

La distribution de Dirichlet de paramètres $(\nu_1,\dots,\nu_d)$ est la loi qui a pour densité sur $\Sigma$

$$\frac{\Gamma(\nu_1+\ldots+\nu_d)}{\Gamma(\nu_1)\dots\Gamma(\nu_d)}\prod_{j=1}^{d}x_j^{\nu_j-1}d\Sigma(x_1,\dots,x_d).$$

En particulier, si $D=(D_1,\dots,D_d)$ est un vecteur aléatoire de loi de Dirichlet de paramètres $(\nu_1,\dots,\nu_d)$, alors, pour tout $p=(p_1,\dots,p_d)\in\mathbb{N}^d$, le moment joint d'ordre $p$ de $D$ est donné par

$$\mathbb{E}(D^p)=\mathbb{E}(D_1^{p_1}\dots D_d^{p_d})=\frac{\Gamma(\nu)}{\Gamma(\nu+|p|)}\prod_{j=1}^{d}\frac{\Gamma(\nu_j+p_j)}{\Gamma(\nu_j)}$$

où $\nu=\sum_{j=1}^{d}\nu_j$ et $|p|=\sum_{j=1}^{d}p_j$.

De plus, chaque variable aléatoire $D_j$, à valeurs dans $[0,1]$, suit la loi Bêta de paramètres $(\nu_j,\nu-\nu_j)$ et est donc de densité

$$\frac{1}{B(\nu_j,\nu-\nu_j)}t^{\nu_j-1}(1-t)^{\nu-\nu_j-1}\mathbb{1}_{[0,1]}dt.$$

Une description alternative de la distribution de Dirichlet est la suivante :

**Proposition 7.3.2 (cf. Bertoin [Ber06, page 63])**

*Si $\xi_1,\dots,\xi_d$ sont $d$ variables indépendantes de lois Gamma de paramètres respectifs $(\nu_1,\nu),\dots,(\nu_d,\nu)$, si $\xi=\sum_{i=1}^{d}\xi_i$, alors $\xi$ est de loi Gamma$(\nu_1+\ldots+\nu_d,\nu)$, et le vecteur aléatoire $\left(\frac{\xi_1}{\xi},\dots,\frac{\xi_d}{\xi}\right)$ suit la loi de Dirichlet de paramètres $(\nu_1,\dots,\nu_d)$ et est indépendant de $\xi$.*

**Démonstration du Théorème 7.3.1 :** Soit $\alpha=\sum_{j=1}^{d}\alpha_j\geqslant 1$. On note $\mathcal{F}_n$ la filtration engendrée par le processus $(U(n))_{n\geqslant 0}$ jusqu'au temps $n$, nous avons donc

$$\mathbb{E}(U(n+1)|\mathcal{F}_n)=\frac{\alpha+(n+1)S}{\alpha+nS}U(n),$$

ce qui implique que $\left(\frac{U(n)}{\alpha+nS}\right)_{n\geqslant 0}$ est une martingale à valeur dans $[0,1]^d$, convergente, et de moyenne $\left(\frac{\alpha_1}{\alpha},\dots,\frac{\alpha_d}{\alpha}\right)$. On notera $V$ la limite de cette martingale. Pour toute fonction $f$ de $\mathbb{R}^d$ dans $\mathbb{R}$,

$$\mathbb{E}(f(U(n+1)|\mathcal{F}_n)=\left(I+\frac{\Phi}{\alpha+nS}\right)(f)(U(n)),$$

où, pour tout réel $v$, pour toute fonction $f$,

$$\Phi(f)(v)=\sum_{j=1}^{d}v_j(f(v+Se_j)-f(v)),$$

où $e_j$ est le $j^{\text{ème}}$ vecteur de la base canonique de $\mathbb{R}^d$ et où $v=\sum_{j=1}^{d}v_je_j$. On peut ainsi vérifier que, si $p=(p_1,\dots,p_d)\in\mathbb{N}^d$ et si l'on note $|p|=\sum_{j=1}^{d}p_j$, la fonction

$$\Gamma_p(v)=\prod_{j=1}^{d}\frac{\Gamma\left(\frac{v_j}{S}+p_j\right)}{\Gamma\left(\frac{v_j}{S}\right)},$$

définie sur $\mathbb{R}^d$, est une fonction propre de l'opérateur $\Phi$, associée à la valeur propre $|p|S$. Dès lors, par récurrence, pour tout $p \in \mathbb{N}^d$,

$$\mathbb{E}[\Gamma_p(U(n))] = \frac{\Gamma\left(\frac{\alpha}{S} + n + |p|\right)}{\Gamma\left(\frac{\alpha}{S} + n\right)} \cdot \frac{\Gamma\left(\frac{\alpha}{S}\right)}{\Gamma\left(\frac{\alpha}{S} + |p|\right)} \cdot \Gamma_p(U(0)).$$

Via la formule de Stirling, nous en déduisons que, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mathbb{E}[\Gamma_p(U(n))] = n^{|p|} \cdot \frac{\Gamma\left(\frac{\alpha}{S}\right)}{\Gamma\left(\frac{\alpha}{S} + |p|\right)} \cdot \Gamma_p(U(0)) \cdot \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Par ailleurs, si nous écrivons la décomposition des polynômes $X^p = X_1^{p_1} \dots X_d^{p_d}$ dans la base de fonction propres $(\Gamma_p)_{p \in \mathbb{N}^d}$, nous obtenons

$$X^p = S^{|p|}\Gamma_p + \sum_{\substack{j \in \mathbb{N}^d \\ |j| \leqslant |p|-1}} a_{p,j}\Gamma_j(X)$$

où les $a_{p,j}$ sont des rationnels. Dès lors, asymptotiquement quand $n$ tend vers $+\infty$,

$$\mathbb{E}\left(\frac{U(n)}{\alpha + nS}\right)^p = \frac{\Gamma\left(\frac{\alpha}{S}\right)}{\Gamma\left(\frac{\alpha}{S} + |p|\right)}\Gamma_p(U(0))\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

ce qui implique que, pour tout $p \in \mathbb{N}^d$,

$$\mathbb{E}(V^p) = \frac{\Gamma\left(\frac{\alpha}{S}\right)}{\Gamma\left(\frac{\alpha}{S} + |p|\right)} \prod_{j=1}^{d} \frac{\Gamma\left(\frac{\alpha_j}{S} + p_j\right)}{\Gamma\left(\frac{\alpha_j}{S}\right)}. \tag{7.7}$$

Nous avons donc montré que la martingale converge dans $L^t$, pour tout $t \geqslant 1$. Comme une loi de Dirichlet est bornée, donc déterminée par ses moments, nous en déduisons que la loi de $V$ est une loi de Dirichlet de paramètres $\left(\frac{\alpha_1}{S}, \dots, \frac{\alpha_d}{S}\right)$. ■

# Chapitre 8

# Grandes urnes à deux couleurs

## 8.1 Motivations

Dans ce chapitre, nous nous intéressons aux grandes urnes à deux couleurs, et plus précisément aux variables $W^{DT}$ et $W^{CT}$ décrivant le comportement asymptotique d'une urne équilibrée et non triangulaire.

Ces deux variables aléatoires sont déjà étudiées par Chauvin et al. [CPS11] qui montrent que $W^{CT}$ admet une densité sur $\mathbb{R}$. Leurs travaux se fondent sur l'analyse de la transformée de Fourier de cette variable aléatoire. Les auteurs montrent que les transformées de Fourier des variables $W^{CT}$ issues des deux configurations initiales ${}^t(1,0)$ et ${}^t(0,1)$ sont solutions d'un système différentiel non linéaire. À partir de ce système, une expression explicite des transformées de Fourier de ces deux variables aléatoires est déduite, permettant ainsi d'appliquer un théorème d'inversion de Fourier et de prouver ainsi l'existence de la densité des $W^{CT}$. Il est cependant à noter que ces transformées de Fourier ne sont ni $L^2$, ni $L^1$, ce qui contraint les auteurs à appliquer un théorème d'inversion de Fourier *ad hoc*. Cette expression permet à Chauvin et al. de montrer que la série de Laplace de $W^{CT}$ est de rayon de convergence nul. Par ailleurs, la forme de cette transformée de Fourier ne permet pas de très bien appréhender la distribution de $W^{CT}$ et beaucoup de questions restent ouvertes : notamment l'ordre exact des moments de $W^{CT}$ reste inconnu. De plus, aucune information ne peut être traduite pour le processus en temps discret : $W^{DT}$ admet-elle une densité ? quel est l'ordre de ses moments ?

L'objectif de ce chapitre est de reprendre *ab initio* l'étude de ces variables $W^{CT}$ et $W^{DT}$ via d'autres méthodes. Notre objectif est d'explorer de nouvelles méthodes, non seulement pour compléter notre connaissance des $W^{DT}$ et $W^{CT}$ qui est pour l'instant très partielle, mais aussi en vue de généraliser ces nouvelles méthodes à l'étude des urnes à $d$ couleurs. Notre nouvelle approche utilise la structure arborescente de l'urne de Pólya. La composition de l'urne de Pólya de composition initiale ${}^t(\alpha,\beta)$ *ressemble* à la somme des compositions de $\alpha$ urnes de compositions initiales ${}^t(1,0)$ et $\beta$ urnes de composition initiale ${}^t(0,1)$ : pouvons-nous rendre cette heuristique rigoureuse ? Par ailleurs, supposons que l'urne contienne initialement une unique boule, par exemple rouge. Alors, la première pioche est déterministe, et après cette première étape, la composition de l'urne est ${}^t(a+1,b)$, et une urne de composition initiale ${}^t(1,0)$ *ressemble* à la somme des compositions de $a+1$ urnes de composition initiale ${}^t(1,0)$ et $b$ urnes de composition initiale ${}^t(0,1)$.

C'est grâce à des raisonnements de ce type que nous montrons dans la Section 8.2 que l'on peut en effet se réduire à l'étude de deux compositions initiales, et donc à l'étude de deux variables aléatoires $W_{(1,0)}$ et $W_{(0,1)}$, aussi bien en temps discret qu'en temps continu, et que ces deux variables aléatoires sont solutions d'un système d'équations en loi, ou système de point fixe, en temps discret,

comme en temps continu. Nous montrons dans la Section 8.3 que ces deux systèmes de point fixe admettent une unique solution dans un espace de mesures muni de la distance de Wasserstein, et ce via le théorème de point fixe de Banach. La Section 8.4 est consacrée à l'étude des moments de $W^{CT}_{(1,0)}$ et $W^{CT}_{(0,1)}$ : nous montrons, par récurrence sur l'ordre des moments, que ces deux variables aléatoires vérifient le critère de Carleman et leurs lois sont donc déterminées par leurs moments. Les connexions entre temps discret et temps continu nous permettent d'étendre ce résultat à $W^{DT}_{(1,0)}$ et $W^{DT}_{(0,1)}$, puis à toute composition initiale au vu des résultats de la Section 8.2. Nous montrerons en outre que, au contraire de celle de $W^{CT}_{(\alpha,\beta)}$ (cf. [CPS11]), la série de Laplace de $W^{DT}_{(\alpha,\beta)}$ a un rayon de convergence infini. Enfin, dans la Section 8.5, nous montrons que la transformée de Fourier de $W^{DT}_{(\alpha,\beta)}$ est $L^1$ et donc que $W^{DT}_{(\alpha,\beta)}$ admet une densité dans $\mathbb{R}$, et ce pour toute composition initiale $(\alpha, \beta)$.

Nous reprenons dans ce Chapitre les notations introduites dans le Chapitre 7. **Dans tout ce chapitre, nous supposons que la matrice $R$ de valeurs propres $S$ et $m$ définit une grande urne à deux couleurs équilibrée et non triangulaire. Nous supposons de plus que la non-extinction de cette urne est presque sûre.**

## 8.2 Arborescence

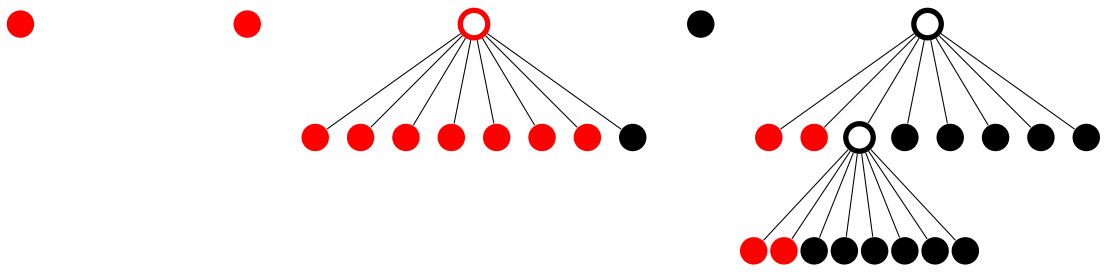
Dans cette section, nous explorons la structure arborescente de l'urne. Dans un premier temps, nous réduirons l'étude des variables $W_{(\alpha,\beta)}$ à l'étude des deux variables aléatoires $W_{(1,0)}$ et $W_{(0,1)}$, dans un second temps, nous montrerons que $W_{(1,0)}$ et $W_{(0,1)}$ sont solutions d'un système de point fixe en loi. Nous détaillons le raisonnement en temps discret mais pas en temps continu car l'indépendance entre les sous-arbres rend le raisonnement plus direct en temps continu. De plus, les résultats en temps continu sont mentionnés dans [Jan04].

### 8.2.1 Décomposition en temps discret

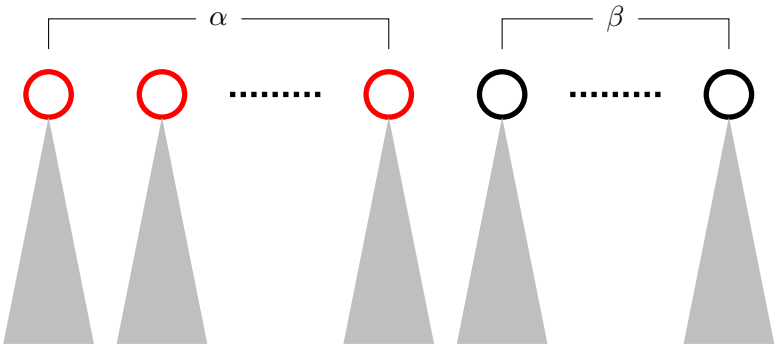
Cette section ne concerne que le processus en temps discret : nous prendrons donc la liberté d'omettre les exposants $DT$ lorsqu'aucune ambigüité n'est possible.

Soit $(\mathcal{T}_n)$ le processus de forêts décrit comme suit : à l'instant initial, la forêt est constituée de $\alpha$ racines rouges et de $\beta$ racines noires ($\alpha, \beta \in \mathbb{N}$ tels que $\alpha + \beta > 0$) : à chaque étape, nous piochons uniformément au hasard une feuille de la forêt, cette feuille devient un nœud interne ayant $S + 1$ enfants constitués de $a + 1$ feuilles rouges et $b$ feuilles noires si la feuille piochée était rouge, et de $c$ feuilles rouges et $d + 1$ feuilles noires si la feuille piochée était noire.

L'ensemble des feuilles de la forêt est une urne de Pólya de composition initiale ${}^t(\alpha, \beta)$ et de matrice de transition $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Ainsi, la figure suivante représente une réalisation possible de l'urne de composition initiale ${}^t(3, 2)$ et de matrice de remplacement $\begin{pmatrix} 6 & 1 \\ 2 & 5 \end{pmatrix}$ après trois tirages :



La forêt est donc composée de $\alpha+\beta$ arbres, $\alpha$ d'entre eux enracinés par une boule rouge et $\beta$ d'entre eux par une boule noire :




La composition de l'urne est donc la somme des compositions des $\alpha+\beta$ sous-arbres de la forêt décrite ci-dessus, et chacun de ces sous-arbres décrit une urne de Pólya de composition initiale $(1,0)$ ou $(0,1)$ :

$$U_{(\alpha,\beta)} \overset{(loi)}{=} \sum_{k=1}^{\alpha} U_{(1,0)}^{(k)}(T_k(n)) + \sum_{k=\alpha+1}^{\beta} U_{(0,1)}^{(k)}(T_k(n)),$$

où $T_k(n)$ représente le *temps à l'intérieur du $k^{\text{ème}}$ sous-arbre de la forêt*, i.e. le nombre de pioches effectuées parmi les feuilles du $k^{\text{ème}}$ sous-arbre, et où les processus d'urnes $(U^{(k)})_{k\in\{1,\dots,\alpha+\beta\}}$ sont des copies indépendantes de $U_{(1,0)}$ et $U_{(0,1)}$, respectivement. Autrement dit, comme à chaque fois que l'on pioche dans le $k^{\text{ème}}$ sous-arbre, on ajoute $S$ feuilles dans ce sous-arbre, si l'on note $D_n(k)$ le nombre de feuilles du $k^{\text{ème}}$ sous-arbre, alors

$$T_n(k) = \frac{D_k(n)-1}{S}.$$

Dès lors,

$$U_{(\alpha,\beta)} \overset{(loi)}{=} \sum_{k=1}^{\alpha} U_{(1,0)}^{(k)}\left(\frac{D_k(n)-1}{S}\right) + \sum_{k=\alpha+1}^{\beta} U_{(0,1)}^{(k)}\left(\frac{D_k(n)-1}{S}\right). \tag{8.1}$$

Remarquons que $(D_1(0),\dots,D_{\alpha+\beta}(0)) = (1,\dots,1)$, que, à chaque étape, une feuille est choisie uniformément au hasard parmi les feuilles de la forêt, et que, si cette feuille appartient au $k^{\text{ème}}$ sous-arbre, on ajoute $S$ feuilles à ce sous-arbre. Cette description est la description d'une urne de Pólya-Eggenberger à $(\alpha+\beta)$ couleurs (on oublie les couleurs des feuille rouges et noires, et deux feuilles sont de la même couleur si et seulement si elle sont dans le même sous-arbre de la forêt), de composition initiale ${}^t(1,\dots,1)$, et de matrice de remplacement $SI_{\alpha+\beta}$ (où $I_{\alpha+\beta}$ est la matrice identité de dimension $\alpha+\beta$). Dès lors, d'après le Théorème 7.3.1, le vecteur $(D_1(n),\dots,D_{\alpha+\beta}(n))$ vérifie le théorème limite suivant : asymptotiquement quand $n$ tend vers $+\infty$,

$$\frac{1}{nS}(D_1(n),\dots,D_{\alpha+\beta}(n)) \to (V_1,\dots,V_{\alpha+\beta}),$$

presque sûrement, où $(V_1,\dots,V_{\alpha+\beta})$ suit la loi de Dirichlet de paramètres $\left(\frac{1}{S},\dots,\frac{1}{S}\right)$.

Dès lors, via l'Équation 8.1, en utilisant le fait que, par définition de $W$ (cf. Équation (7.3)), presque sûrement,

$$W_{(\alpha,\beta)} = \lim_{n\to+\infty} u_2\left(\frac{U_{(\alpha,\beta)}(n)}{n^\sigma}\right),$$

nous obtenons le théorème suivant après renormalisation, projection selon $u_2$de l'Équation (8.1) et limite quand $n$ tend vers $+\infty$ :

**Théorème 8.2.1**

*Pour tout* $(\alpha,\beta) \in \mathbb{N}\backslash\{(0,0)\}$,

$$W_{(\alpha,\beta)} \overset{(loi)}{=} \sum_{k=1}^{\alpha} V_k^{\sigma} W_{(1,0)}^{(k)} + \sum_{k=\alpha+1}^{\alpha+\beta} V_k^{\sigma} W_{(0,1)}^{(k)},$$

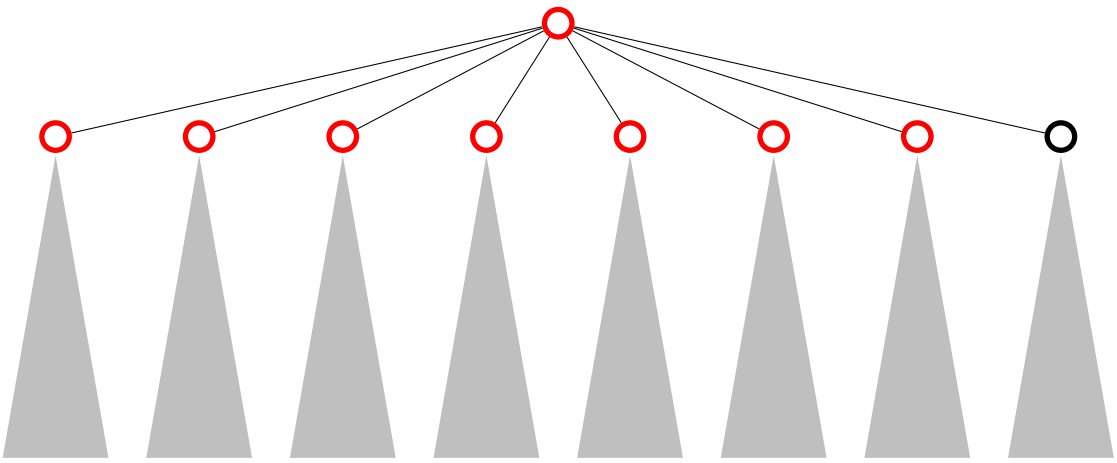
*où* $V = (V_1,\dots,V_{\alpha+\beta})$ *est un vecteur aléatoire de loi de Dirichlet de paramètres* $\left(\frac{1}{S},\dots,\frac{1}{S}\right)$, *et où les* $W_{(1,0)}^{(k)}$ *et* $W_{(1,0)}^{(k)}$ *sont des copies indépendantes de* $W_{(1,0)}$ *et* $W_{(0,1)}$, *respectivement, indépendantes du vecteur* $V$.

Grâce à ce théorème, nous pouvons nous concentrer sur les deux variables aléatoires $W_{(1,0)}$ et $W_{(0,1)}$ : les informations que nous obtiendrons sur ces deux variables pourront, via le théorème ci-dessus, se traduire pour toute la famille des $W_{(\alpha,\beta)}$.

### 8.2.2 Dislocation en temps discret

En utilisant de nouveau la structure arborescente de l'urne de Pólya, nous allons montrer que les deux variables $W_{(1,0)}$ et $W_{(0,1)}$ sont solutions d'un système d'équations en loi, point de départ des travaux réalisés dans ce chapitre. Avant tout, afin d'alléger les notations, notons $X = W_{(1,0)}$ et $Y = W_{(0,1)}$.

Commençons par étudier la variable $X$, issue l'urne de composition initiale ${}^t(1,0)$. Le premier tirage est déterministe : nous piochons la boule rouge de l'urne et le vecteur de composition à l'étape 1 est donné par ${}^t(a+1,b)$. Si l'on reprend l'analogie décrite précédemment avec une forêt dont les feuilles représentent les boules de l'urne, la forêt associée à l'urne de composition initiale ${}^t(1,0)$ est composée d'un unique arbre enraciné par un nœud interne (l'ancienne boule rouge initiale) ayant $S+1$ enfants : $a+1$ feuilles rouges et $b$ feuilles noires. Ces enfants sont racines de $S+1$ sous-arbres et la composition de l'urne est la somme des compositions de ces $S+1$ sous-arbres. Ces sous-arbres sont eux mêmes associés à des processus d'urnes de composition initiale une boule unique (rouge ou noire), et de matrice de remplacement $R$.



Nous avons donc

$$U_{(1,0)}(n) = \sum_{k=1}^{a+1} U_{(1,0)}^{(k)}(T_k(n)) + \sum_{k=a+2}^{S+1} U_{(0,1)}^{(k)}(T_k(n)),$$

où $T_k(n)$ est le *temps à l'intérieur du $k^{ième}$ sous-arbre de la racine*, i.e. le nombre de feuilles piochées dans ce sous-arbre avant l'étape $n$, et où les processus $(U^{(k)})_{k\in\{1,\dots,S+1\}}$ sont des copies indépendantes de $U_{(1,0)}$ et $U_{(0,1)}$, respectivement, indépendantes de $(T_1(n),\dots,T_{S+1}(n))$. Tout comme précédemment, remarquons que

$$T_k(n) = \frac{D_k(n)-1}{S},$$

où $D_k(n)$ est le nombre de feuilles du $k^{\text{ième}}$ sous-arbre à l'étape $n$. De plus, le vecteur $(D_1(n),\dots,D_{S+1}(n))$ est le vecteur composition au temps $(n-1)$ d'une urne de Pólya à $S+1$ couleurs, de composition initiale ${}^t(1,\dots,1)$ et de matrice de transition $SI_{S+1}$. Dès lors, au vu du Théorème 7.3.1, nous avons, asymptotiquement quand $n$ tend vers $+\infty$,

$$\frac{1}{nS}(D_1(n),\dots,D_{S+1}(n)) \to (V_1,\dots,V_{S+1}),$$

presque sûrement, où le vecteur $(V_1,\dots,V_{S+1})$ est un vecteur aléatoire de loi de Dirichlet de paramètre $\left(\frac{1}{S},\dots,\frac{1}{S}\right)$. Nous obtenons donc, via l'Équation (7.3), le théorème suivant :

**Théorème 8.2.2**

*Les variables aléatoires $X = W_{(1,0)}$ et $Y = W_{(0,1)}$ vérifient le système suivant :*

$$\begin{cases} X \overset{(loi)}{=} \displaystyle\sum_{k=1}^{a+1} V_k^{\sigma} X^{(k)} + \sum_{k=a+2}^{S+1} V_k^{\sigma} Y^{(k)} \\ Y \overset{(loi)}{=} \displaystyle\sum_{k=1}^{c} V_k^{\sigma} X^{(k)} + \sum_{k=c+1}^{S+1} V_k^{\sigma} Y^{(k)} \end{cases}$$

*où le vecteur $V = (V_1,\dots,V_{S+1})$ est un vecteur aléatoire de loi Dirichlet de paramètres $\left(\frac{1}{S},\dots\frac{1}{S}\right)$, et où les $X^{(k)}$ et $Y^{(k)}$ sont des copies indépendantes de $X$ et $Y$, respectivement, indépendantes de $V$.*

L'objectif de la suite de ce chapitre est d'appliquer à ce système de point fixe des méthodes utilisées dans l'étude d'équations de point fixe (ou *smoothing equations*) dans la littérature de façon à en déduire un maximum d'information sur la loi de $W_{(1,0)}$ et $W_{(0,1)}$.

### 8.2.3 Résultats analogues en temps continu et connexion

Bien entendu, le processus d'urne plongé en temps continu admet lui aussi une structure arborescente sous-jacente. Les équations sont d'ailleurs plus directes qu'en temps discret grâce à l'indépendance entre les sous-arbres. Ainsi,

$$U^{CT}_{(\alpha,\beta)}(t) = \langle\alpha\rangle U^{CT}_{(1,0)}(t) + \langle\beta\rangle U^{CT}_{(0,1)}(t),$$

où la notation $\langle n\rangle X$, pour tout entier $n$ et toute variable aléatoire $X$ représente la somme de $n$ copies indépendantes de $X$. Dès lors, via la définition de $W^{CT}$ comme limite de martingale (cf. Équation (7.4)), nous obtenons

$$W^{CT}_{(\alpha,\beta)} = \langle\alpha\rangle W^{CT}_{(1,0)} + \langle\beta\rangle W^{CT}_{(0,1)}. \tag{8.2}$$

Ce résultat est l'analogue du Théorème 8.2.1 en temps continu : il permet de se ramener à l'étude de deux variables aléatoires, $W^{CT}_{(1,0)}$ et $W^{CT}_{(0,1)}$, au lieu d'étudier toute une famille.

De même qu'en temps discret, nous pouvons montrer le théorème suivant :

**Théorème 8.2.3 (cf. Janson [Jan04] ou [CPS11])**

*Si l'on pose $X = W^{CT}_{(1,0)}$ et $Y = W^{CT}_{(0,1)}$, alors ces deux variables aléatoires vérifient*

$$\begin{cases} X \overset{(loi)}{=} U^m \left( \sum_{k=1}^{a+1} X^{(k)} + \sum_{k=a+2}^{S+1} Y^{(k)} \right) \\ Y \overset{(loi)}{=} U^m \left( \sum_{k=1}^{c} X^{(k)} + \sum_{k=c+1}^{S+1} Y^{(k)} \right), \end{cases}$$

*où $U$ est une variable aléatoire uniforme sur $[0,1]$ et où les $X^{(k)}$ et $Y^{(k)}$ sont des copies indépendantes de $X$ et $Y$, respectivement, et indépendantes de $U$.*

Il est intéressant de noter que l'on peut déduire le système vérifié par $W^{CT}_{(1,0)}$ et $W^{CT}_{(0,1)}$ de celui vérifié par $W^{DT}_{(1,0)}$ et $W^{DT}_{(1,0)}$, et ce via la connexion (7.5) :

**Proposition 8.2.4**

*Soient $X$ et $Y$ solutions du système du Théorème 8.2.2, soit $\xi$ une variable aléatoire de loi Gamma$\left(\frac{1}{S}\right)$. Alors les variables aléatoires $\xi^\sigma X$ et $\xi^\sigma Y$ sont solution du système du Théorème 8.2.3.*

Cette proposition, via la connexion (7.5), nous assure que le Théorème 8.2.3 peut être vu comme une conséquence du Théorème 8.2.2. L'implication réciproque, si elle existe, n'est pas encore connue. Cette proposition est un corollaire du lemme suivant, il suffit pour cela de remarquer que si $U$ est une variable aléatoire uniforme sur $[0,1]$, $U^S$ est de loi Bêta$\left(\frac{1}{S}, 1\right)$ :

**Lemme 8.2.5**

*Considérons les deux équations en loi suivantes d'inconnues $X, X_1, \dots, X_{S+1}$ :*

$$X \overset{(loi)}{=} \sum_{k=1}^{S+1} V_k^\sigma X_k \tag{8.3}$$

*où $V = (V_1, \dots, V_{S+1})$ est un vecteur aléatoire de loi de Dirichlet de paramètre $\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$ indépendant des $X_1, \dots, X_{S+1}$ ; et*

$$X \overset{(loi)}{=} V^\sigma \sum_{k=1}^{S+1} X_k \tag{8.4}$$

*où $V$ est une variable aléatoire de loi Bêta$\left(\frac{1}{S}, 1\right)$ indépendante des $X_1, \dots, X_{S+1}$.*

*Soient $V, \xi_1, \dots, \xi_{S+1}$ des variables aléatoires indépendantes telles que $\xi_1, \dots, \xi_{S+1}$ sont de loi Gamma$\left(\frac{1}{S}\right)$ et $V$ est de loi Bêta$\left(\frac{1}{S}, 1\right)$. On pose*

$$\xi = V \sum_{i=1}^{S+1} \xi_i,$$

*et pour tout $k \in \{1, \dots, S+1\}$,*

$$V_k = \frac{\xi_k}{\sum_{i=1}^{S+1} \xi_i}.$$

*Dès lors,*

*(i) la variable aléatoire $\xi$ est de loi Gamma$\left(\frac{1}{S}\right)$,*

*(ii) le vecteur aléatoire* $(V_1, \dots, V_{S+1})$ *est indépendant de* $\xi$ *et de loi Dirichlet*$\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$,

*(iii) si* $X, X_1, \dots, X_{S+1}$ *satisfont l'Équation* (8.3), *si* $X_1, \dots, X_{S+1}$ *sont indépendants de* $V, \xi_1, \dots, \xi_{S+1}$ *et si* $X$ *est indépendant de* $\xi$, *alors* $\xi^\sigma X$, $\xi_1^\sigma X_1$, …, $\xi_{S+1}^\sigma X_{S+1}$ *satisfont l'Équation* (8.4).

**Démonstration :** *(i)* Cette affirmation se démontre par exemple par calcul des moments : les lois Bêta et Gamma sont déterminées par leurs moments, le $p^{\text{ième}}$ moment d'une loi Bêta de paramètres $(\alpha, \beta)$ est donné par

$$\frac{\Gamma(\alpha+p)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+p)\Gamma(\alpha)}$$

où $\Gamma$ est la fonction Gamma d'Euler, et le $p^{\text{ème}}$ moment d'une loi Gamma de paramètre $\alpha$ est donné par

$$\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}.$$

Il faut aussi remarquer que la somme de $d$ variables aléatoires indépendantes de lois Gamma de paramètres respectifs $\alpha_1, \dots, \alpha_d$ est de loi Gamma$(\alpha_1 + \dots + \alpha_d)$.

*(ii)* Cette affirmation est une conséquence directe de la Proposition 7.3.2.

*(iii)* Supposons que $X, X_1, \dots, X_{S+1}$ soient solution de l'Équation (8.3) pour ces $V_1, \dots, V_{S+1}$. Multiplions cette équation par $\xi^\sigma$ de chaque côté : cette multiplication dans une équation en loi est licite car $\xi$ est indépendant des deux membres de l'Équation (8.3), et remplaçons les $V_k$ par leur valeur :

$$\xi^\sigma X \overset{(loi)}{=} \xi^\sigma \sum_{k=1}^{S+1} \frac{\xi_k^\sigma}{\left(\sum_{i=1}^{S+1} \xi_i\right)^\sigma} X_k = V^\sigma \sum_{k=1}^{S+1} \xi_k^\sigma X_k,$$

ce qui conclut la preuve. ■

La suite du présent chapitre est consacrée à l'étude des systèmes des Théorèmes 8.2.2 et 8.2.3.

## 8.3 Unicité des solutions

Nous montrons dans cette section l'unicité des solutions des systèmes de point fixe des théorèmes 8.2.2 et 8.2.3. Notons que l'unicité de la solution du système en temps discret peut être déduite du théorème plus général de Neininger et Rüshendorf [NR06]. Nous en donnons néanmoins une courte preuve ici dans le cas particulier des urnes de Pólya : cette preuve est inspirée des travaux de Fill et Kapur [FK05].

Soit $A$ un réel, et soit $\mathcal{M}_2(A)$ l'ensemble des mesures de probabilité de moyenne $A$ et de carré intégrable. Équipons cet espace de la distance de Wasserstein définie ci-après : cet espace est un espace métrique complet dans lequel nous pourrons donc appliquer le théorème de point fixe de Banach. Tout d'abord, remarquons que si $X$ et $Y$ sont solution du système en temps discret ou du système en temps continu, si $\mathbb{E}X = B$ et $\mathbb{E}Y = C$, alors $cB + bC = 0$. Nous montrerons que si $B$ et $C$ sont deux réels tels que $cB + bC = 0$, alors les systèmes issus des Théorèmes 8.2.2 et 8.2.3 ont chacun une unique solution dans l'espace métrique produit $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$.

Comme le vecteur aléatoire $(W_{(1,0)}, W_{(0,1)})$ est d'espérance proportionnelle à $(b, -c)$, en temps discret aussi bien qu'en temps continu, ce résultat d'unicité nous assure que les systèmes des Théorèmes 8.2.2 et 8.2.3 caractérisent les distributions de $W_{(1,0)}^{DT}$ et $W_{(0,1)}^{DT}$ en temps discret, et $W_{(1,0)}^{CT}$ et $W_{(0,1)}^{CT}$ en temps continu.

### 8.3.1 Distance de Wasserstein

Soit $A \in \mathbb{R}$. La distance de Wasserstein sur $\mathcal{M}_2(A)$ est définie comme suit :

$$d_W(\mu_1, \mu_2) = \min_{(X_1, X_2)} \left(\mathbb{E}(X_1 - X_2)^2\right)^{1/2},$$

où le minimum est pris sur les vecteurs aléatoires $(X_1, X_2)$ de $\mathbb{R}^2$ dont les marginales sont $\mu_1$ et $\mu_2$. Le théorème de Kantorovich-Rubinstein nous assure que ce minimum est atteint. Par ailleurs, l'espace $\mathcal{M}_2(A)$ équipé de la distance de Wasserstein est un espace métrique complet (voir par exemple Dudley [Dud02]).

Soit $(B, C) \in \mathbb{R}^2$. Équipons l'espace produit $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$ de la distance produit, définie par exemple comme suit :

$$d\left((\mu_1, \nu_1), (\mu_2, \nu_2)\right) = \max\{d_W(\mu_1, \mu_2), d_W(\nu_1, \nu_2)\}.$$

L'espace produit $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$ muni de cette distance est un espace métrique complet.

### 8.3.2 Méthode de contraction en temps discret

Rappelons ci-dessous le système de point fixe vérifié par $(X^{DT}, Y^{DT})$ (cf. Théorème 8.2.2) :

$$\begin{cases} X \overset{(loi)}{=} \displaystyle\sum_{k=1}^{a+1} V_k^\sigma X^{(k)} + \sum_{k=a+2}^{S+1} V_k^\sigma Y^{(k)} \\ \\ Y \overset{(loi)}{=} \displaystyle\sum_{k=1}^{c} V_k^\sigma X^{(k)} + \sum_{k=c+1}^{S+1} V_k^\sigma Y^{(k)}. \end{cases} \tag{8.5}$$

Soit $\mathcal{M}_2$ l'espace des mesures de probabilité de carré intégrable sur $\mathbb{R}$. Pour tout $(B, C) \in \mathbb{R}^2$, soit $K_1$ la fonction définie sur $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$ par

$$\begin{array}{rccl} K_1 : & \mathcal{M}_2(B) \times \mathcal{M}_2(C) & \longrightarrow & \mathcal{M}_2 \\ & (\mu, \nu) & \longmapsto & \mathcal{L}\left(\displaystyle\sum_{k=1}^{a+1} V_k^\sigma X^{(k)} + \sum_{k=a+2}^{S+1} V_k^\sigma Y^{(k)}\right) \end{array}$$

où $X^{(1)}, \ldots, X^{(a+1)}$ sont des variables indépendantes de loi $\mu$, $Y^{(a+2)}, \ldots, Y^{(S+1)}$ sont des variables aléatoires indépendantes de loi $\nu$ et où $V = (V_1, \ldots, V_{S+1})$ est un vecteur aléatoire de loi de Dirichlet de paramètres $\left(\frac{1}{S}, \ldots, \frac{1}{S}\right)$, les $X^{(k)}$, $Y^{(k)}$ et $V$ étant indépendants. De même, soit $K_2$ la fonction définie par

$$\begin{array}{rccl} K_2 : & \mathcal{M}_2(B) \times \mathcal{M}_2(C) & \longrightarrow & \mathcal{M}_2 \\ & (\mu, \nu) & \longmapsto & \mathcal{L}\left(\displaystyle\sum_{k=1}^{c} V_k^\sigma X^{(k)} + \sum_{k=c+1}^{S+1} V_k^\sigma Y^{(k)}\right). \end{array}$$

Un calcul direct montre que si $(\mu, \nu) \in \mathcal{M}_2(B) \times \mathcal{M}_2(C)$, alors

$$\mathbb{E}K_1(\mu, \nu) = \frac{(a+1)B + bC}{m+1},$$

et

$$\mathbb{E}K_2(\mu, \nu) = \frac{cB + (d+1)C}{m+1},$$

de telle sorte que, comme $m = a - c = d - b$, l'équation $cB + bC = 0$ est une condition nécessaire et suffisante pour que la fonction produit $(K_1, K_2)$ soit une fonction de $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$ dans lui-même.

**Lemme 8.3.1**

*Soient $B$ et $C$ deux réels tels que $cB + bC = 0$. Alors, la fonction*

$$\begin{aligned} K: \quad \mathcal{M}_2(B) \times \mathcal{M}_2(C) &\longrightarrow \mathcal{M}_2(B) \times \mathcal{M}_2(C) \\ (\mu, \nu) &\longmapsto (K_1(\mu,\nu), K_2(\mu,\nu)) \end{aligned}$$

*est $\sqrt{\frac{S+1}{2m+1}}$-Lipschitz. C'est donc une contraction dès que $\sigma = \frac{m}{S} > \frac{1}{2}$, ce qui est le cas car l'urne considérée est grande.*

**Théorème 8.3.2**

*(i) Si $B$ et $C$ sont deux réels tels que $cB + bC = 0$, alors le Système (8.5) admet une unique solution dans $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$.*

*(ii) Le couple $(X^{DT}, Y^{DT})$ est l'unique solution du Système (8.5) de moyenne*

$$\left( \frac{\Gamma\left(\frac{1}{S}\right)}{\Gamma\left(\frac{m+1}{S}\right)} \frac{b}{S}, -\frac{\Gamma\left(\frac{1}{S}\right)}{\Gamma\left(\frac{m+1}{S}\right)} \frac{c}{S} \right)$$

*et de carré intégrable.*

Le Théorème 8.3.2 est une conséquence directe du Lemme 8.3.1 et du théorème de point fixe de Banach.

**Démonstration du Lemme 8.3.1 :** Soient $(\mu_1, \nu_1)$ et $(\mu_2, \nu_2)$ deux éléments de $\mathcal{M}_2(B) \times \mathcal{M}_2(C)$. Soit $V = (V_1, \ldots, V_{S+1})$ un vecteur aléatoire de loi de Dirichlet de paramètre $\left(\frac{1}{S}, \ldots, \frac{1}{S}\right)$. Soit $X_1^{(1)}, \ldots, X_1^{(a+1)}$ des variables aléatoires de loi $\mu_1$, $Y_1^{(a+2)}, \ldots, Y_1^{(S+1)}$ des variables aléatoires de loi $\nu_1$, $X_2^{(1)}, \ldots, X_1^{(c)}$ des variables aléatoires de loi $\mu_2$, et $Y_2^{(c+1)}, \ldots, Y_2^{(S+1)}$ des variables aléatoires de loi $\nu_2$, indépendantes entre elles et indépendantes de $V$. Dès lors,

$$\begin{aligned} d_W(K_1(\mu_1,\nu_1), K_1(\mu_2,\nu_2))^2 &\leqslant \left\| \sum_{k=1}^{a+1} V_k^\sigma \left(X_1^{(k)} - X_2^{(k)}\right) + \sum_{k=a+2}^{S+1} V_k^\sigma \left(Y_1^{(k)} - Y_2^{(k)}\right) \right\|_2^2 \\ &= \mathtt{Var}\left[ \sum_{k=1}^{a+1} V_k^\sigma \left(X_1^{(k)} - X_2^{(k)}\right) + \sum_{k=a+2}^{S+1} V_k^\sigma \left(Y_1^{(k)} - Y_2^{(k)}\right) \right] \\ &= \mathbb{E}\mathtt{Var}\left( \sum_{k=1}^{a+1} V_k^\sigma \left(X_1^{(k)} - X_2^{(k)}\right) + \sum_{k=a+2}^{S+1} V_k^\sigma \left(Y_1^{(k)} - Y_2^{(k)}\right) | V \right) \\ &+ \mathtt{Var}\mathbb{E}\left( \sum_{k=1}^{a+1} V_k^\sigma \left(X_1^{(k)} - X_2^{(k)}\right) + \sum_{k=a+2}^{S+1} V_k^\sigma \left(Y_1^{(k)} - Y_2^{(k)}\right) | V \right) \end{aligned}$$

via la loi de la variance totale [1]. Comme $V = (V_1, \ldots, V_{S+1})$ est indépendant des $X_j^{(k)}$ et des $Y_j^{(k)}$,

1. La loi de la variance totale nous assure que, pour toutes variables aléatoires $X$ et $Y$, si $Y$ est de carré intégrable, alors, $\mathtt{Var}Y = \mathbb{E}\mathtt{Var}(Y|X) + \mathtt{Var}\mathbb{E}(Y|X)$.

nous obtenons

$$\begin{aligned} d_W(K_1(\mu_1,\nu_1),K_1(\mu_2,\nu_2))^2 &\leqslant \sum_{k=1}^{a+1}\mathbb{E}V_k^{2\sigma}\mathtt{Var}\left(X_1^{(k)}-X_2^{(k)}\right)+\sum_{k=a+2}^{S+1}\mathbb{E}V_k^{2\sigma}\mathtt{Var}\left(Y_1^{(k)}-Y_2^{(k)}\right)\\ &\leqslant \mathtt{Var}\left(X_1^{(1)}-X_2^{(1)}\right)\sum_{k=1}^{a+1}\mathbb{E}V_k^{2\sigma}+\mathtt{Var}\left(Y_1^{(1)}-Y_2^{(1)}\right)\sum_{k=a+2}^{S+1}\mathbb{E}V_k^{2\sigma}\\ &= \frac{a+1}{2m+1}\left\|X_1^{(1)}-X_2^{(1)}\right\|_2^2+\frac{b}{2m+1}\left\|Y_1^{(1)}-Y_2^{(1)}\right\|_2^2. \end{aligned}$$

Comme cette inégalité est vraie pour toutes variables aléatoires $X_1^{(1)}$, $X_2^{(1)}$, $Y_1^{(1)}$ et $Y_2^{(1)}$ de lois respectives $\mu_1$, $\mu_2$, $\nu_1$ et $\nu_2$, nous en déduisons

$$\begin{aligned} d_W(K_1(\mu_1,\nu_1),K_1(\mu_2,\nu_2))^2 &\leqslant \frac{a+1}{2m+1}d_W(\mu_1,\mu_2)^2+\frac{b}{2m+1}d_W(\nu_1,\nu_2)^2\\ &\leqslant \frac{S+1}{2m+1}d\left((\mu_1,\nu_1),(\mu_2,\nu_2)\right)^2. \end{aligned}$$

De même, nous pouvons montrer que

$$d_W\left(K_2(\mu_1,\nu_1),K_2(\mu_2,\nu_2)\right)^2 \leqslant \frac{S+1}{2m+1}d\left((\mu_1,\nu_1),(\mu_2,\nu_2)\right)^2,$$

ce qui implique finalement

$$d\left(K(\mu_1,\nu_1),K(\mu_2,\nu_2)\right)^2 \leqslant \frac{S+1}{2m+1}d\left((\mu_1,\nu_1),(\mu_2,\nu_2)\right)^2,$$

ce qui conclut la preuve. Notons que c'est l'hypothèse $\sigma=\frac{m}{S}>\frac{1}{2}$ qui garantit que la constante de Lipschitz est strictement plus petite que 1. Autrement dit, l'unicité de la solution du Système (8.5) n'est montrée que dans le cas d'une grande urne. ■

### 8.3.3 Méthode de contraction en temps continu

En temps continu, les lois de $X^{CT}$ et $Y^{CT}$ sont solutions du système suivant (cf. Théorème 8.2.3) :

$$\begin{cases} X \overset{(loi)}{=} U^m\left(\displaystyle\sum_{k=1}^{a+1}X^{(k)}+\sum_{k=a+2}^{S+1}Y^{(k)}\right)\\ \\ Y \overset{(loi)}{=} U^m\left(\displaystyle\sum_{k=1}^{c}X^{(k)}+\sum_{k=c+1}^{S+1}Y^{(k)}\right), \end{cases} \tag{8.6}$$

Le théorème suivant, qui est la version en temps continu du Théorème 8.3.2, peut être démontré de deux façons différentes. Nous pouvons traduire le Théorème 8.3.2, valide en temps discret via la connexion établie en Proposition 8.2.4, ou reprendre les arguments utilisés lors de la preuve du Théorème 8.3.2 pour refaire une preuve directe. Nous omettons les détails.

**Théorème 8.3.3**

*(i) Soient $B$ et $C$ deux réels tels que $cB+bC=0$, alors, le Système (8.6) admet une unique solution dans $\mathcal{M}_2(B)\times\mathcal{M}_2(C)$.*

*(ii) Le couple $(X^{CT},Y^{CT})$ est l'unique solution du Système d'équations en loi (8.6) d'espérance*

$\left(\frac{b}{S}, -\frac{c}{S}\right)$ *et de carré intégrable.*

## 8.4 Moments

Cette section est consacrée à l'étude des moments des variables aléatoires $W^{DT}$ et $W^{CT}$, en temps discret tout comme en temps continu. Nous savons déjà, via la convergence dans tous les $L^p$ ($p \geqslant 1$) dans les Théorèmes 7.2.1 et 7.2.2, que ces variables admettent des moments de tous ordres. Par ailleurs, il est montré par Chauvin et al. [CPS11] que ces moments sont "grands" au sens où la série de Laplace des $W^{CT}$ a un rayon de convergence nul. Nous montrons dans cette section que les lois des $W^{CT}$ vérifient le critère de Carleman et sont donc déterminées par leurs moments. Nous étudions dans cette section le système en temps continu :

$$\begin{cases} X \overset{(loi)}{=} U^m \left( \sum_{k=1}^{a+1} X^{(k)} + \sum_{k=a+2}^{S+1} Y^{(k)} \right) \\ Y \overset{(loi)}{=} U^m \left( \sum_{k=1}^{c} X^{(k)} + \sum_{k=c+1}^{S+1} Y^{(k)} \right), \end{cases}$$

où $U$ est une variable aléatoire uniforme sur $[0,1]$, et où $X$, $X^{(k)}$ et $Y$, $Y^{(k)}$ sont des copies respectives de $X^{CT}$ et $Y^{CT}$, indépendantes les unes des autres et indépendantes de $U$.

Nous savons déjà que les moments de $W^{CT}$ sont *grands* :

**Théorème 8.4.1 ([CPS11])**

*Les séries de Laplace de $X = W^{CT}_{(1,0)}$ et $Y = W^{CT}_{(0,1)}$ sont de rayon de convergence nulle, ce qui implique que pour toute constante $C > 0$, pour tout $p_0 \geqslant 1$, il existe $p \geqslant p_0$ tel que,*

$$C^p \leqslant \frac{\mathbb{E}|X|^p}{p!} \quad \text{et} \quad C^p \leqslant \frac{\mathbb{E}|Y|^p}{p!}.$$

Le lemme suivant donne une borne supérieure pour $\frac{\mathbb{E}|X|^p}{p!}$ et $\frac{\mathbb{E}|Y|^p}{p!}$ : ce lemme est l'argument clef du Théorème 8.4.4 qui affirme que les lois de $X$ et $Y$ sont déterminées par leurs moments.

**Lemme 8.4.2**

*Si $X$ et $Y$ sont des solutions intégrables du Système* (8.6)*, alors elles admettent des moments de tous ordres et les suites $\left(\frac{\mathbb{E}|X|^p}{p!\ln^p p}\right)^{\frac{1}{p}}$ et $\left(\frac{\mathbb{E}|Y|^p}{p!\ln^p p}\right)^{\frac{1}{p}}$ sont bornées.*

**Démonstration :** Soit $\varphi(p) := \ln^p(p+2)$ et soit

$$u_p := \frac{\mathbb{E}|X|^p}{p!\varphi(p)} \quad \text{et} \quad v_p := \frac{\mathbb{E}|Y|^p}{p!\varphi(p)}.$$

Montrons par récurrence que, pour tout entier $p \geqslant 1$, $\left(\frac{\mathbb{E}|X|^p}{p!\varphi(p)}\right)^{\frac{1}{p}}$ and $\left(\frac{\mathbb{E}|Y|^p}{p!\varphi(p)}\right)^{\frac{1}{p}}$ sont finis et définissent deux suites bornées. Une stratégie similaire est développée par Kahane et Peyrière [KP76]. Élevons

la première équation du Système (8.6) à la puissance $p$. Comme $\mathbb{E}U^{mp} = \frac{1}{mp+1}$, et $S+1 = a+b+1$,

$$\mathbb{E}|X|^p \leqslant \frac{1}{mp+1}((a+1)\mathbb{E}|X|^p + b\mathbb{E}|Y|^p + \sum_{\substack{p_1+\cdots+p_{S+1}=p \\ p_j \leqslant p-1, \forall j \in \{1,\ldots,S+1\}}} \frac{p!}{p_1!\ldots p_{S+1}!}\mathbb{E}|X|^{p_1}\ldots\mathbb{E}|X|^{p_{a+1}}\mathbb{E}|Y|^{p_{a+2}}\ldots\mathbb{E}|Y|^{p_{S+1}}),$$

ou encore,

$$(mp-a)\mathbb{E}|X|^p \leqslant b\mathbb{E}|Y|^p + \sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j \leqslant p-1}} \frac{p!}{p_1!\ldots p_{S+1}!}\mathbb{E}|X|^{p_1}\ldots\mathbb{E}|X|^{p_{a+1}}\mathbb{E}|Y|^{p_{a+2}}\ldots\mathbb{E}|Y|^{p_{S+1}}.$$

Nous pouvons bien entendu faire le même calcul pour la deuxième équation du Système (8.6), et nous obtenons :

$$\begin{cases} (mp-a)u_p \leqslant bv_p + \displaystyle\sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j \leqslant p-1}} u_{p_1}\ldots u_{p_{a+1}}v_{p_{a+2}}\ldots v_{p_{S+1}}\frac{\varphi(p_1)\ldots\varphi(p_{S+1})}{\varphi(p)} \\ (mp-d)v_p \leqslant cu_p + \displaystyle\sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j \leqslant p-1}} u_{p_1}\ldots u_{p_c}v_{p_{c+1}}\ldots v_{p_{S+1}}\frac{\varphi(p_1)\ldots\varphi(p_{S+1})}{\varphi(p)}. \end{cases} \tag{8.7}$$

Comme les valeurs propres de $R$ sont $S$ et $m$, et comme $\frac{m}{S} > \frac{1}{2}$ (l'urne considérée est une grande urne), pour tout $p \geqslant 2$, la matrice $(mpI_2 - R)$ est inversible. De plus, l'inverse de $(mpI_2 - R)$ est donné par :

$$(mpI_2 - R)^{-1} = \frac{1}{(mp-d)(mp-a)-bc}\begin{pmatrix} mp-d & b \\ c & mp-a \end{pmatrix}.$$

Pour tout $p \geqslant 2$, $mp-a > S-a = b \geqslant 0$, $mp-d > c \geqslant 0$ et $(mp-d)(mp-a)-bc > 0$, et l'inverse de $(mpI_2 - R)$ est donc à coefficients positifs. Nous pouvons réécrire le Système (8.7) comme

$$(mpI_2 - R)\begin{pmatrix} u_p \\ v_p \end{pmatrix} \leqslant t_{p-1},$$

où l'inégalité se lit coefficient par coefficient, et où le vecteur $t_{p-1}$ a pour coordonnées les deux sommes intervenant le Système (8.7). Dès lors, comme l'inverse de $(mpI_2 - R)$ est à coefficients positifs, nous avons

$$\begin{pmatrix} u_p \\ v_p \end{pmatrix} \leqslant (mpI_2 - R)^{-1}t_{p-1}.$$

Comme le vecteur $t_{p-1}$ ne dépend que des moments de $X$ et $Y$ d'ordre inférieurs ou égaux à $p-1$, par récurrence sur $p$, les solutions $X$ et $Y$ du Système (8.6) admettent des moments de tous ordres dès lors qu'elles sont intégrables.

Soit $p_0$ le plus petit entier tel que, pour tout $p \geqslant p_0$,

$$\frac{m(p-1)}{(mp-a)(mp-d)-bc}\left(1+8\ln(p+2)\right)^{S+1} \leqslant 1.$$

Un tel $p_0$ existe car le terme de gauche tend vers 0 quand $p$ tend vers $+\infty$. Soit

$$A := \max_{1 \leqslant q \leqslant p_0}\{(u_q)^{\frac{1}{q}}, (v_q)^{\frac{1}{q}}\}.$$

Montrons par récurrence sur $p \geqslant p_0+1$ que, pour tout $q \leqslant p-1, (u_q)^{\frac{1}{q}} \leqslant A$ et $(v_q)^{\frac{1}{q}} \leqslant A$. Supposons que cette affirmation soit vraie pour $p \geqslant p_0+1$. Dès lors,

$$\left\{ \begin{array}{l} (mp-a)u_p \leqslant bv_p + A^p \displaystyle\sum_{\substack{p_1+\cdots+p_{S+1}=p \\ p_j \leqslant p-1}} \frac{\varphi(p_1)\ldots\varphi(p_{S+1})}{\varphi(p)} \\ \\ (mp-d)v_p \leqslant cu_p + A^p \displaystyle\sum_{\substack{p_1+\cdots+p_{S+1}=p \\ p_j \leqslant p-1}} \frac{\varphi(p_1)\ldots\varphi(p_{S+1})}{\varphi(p)}. \end{array} \right.$$

Soit

$$\Phi(p) := \sum_{\substack{p_1+\cdots+p_{S+1}=p \\ p_j \leqslant p-1}} \frac{\varphi(p_1)\ldots\varphi(p_{S+1})}{\varphi(p)}. \tag{8.8}$$

Alors,

$$\left\{ \begin{array}{l} (mp-a)u_p \leqslant bv_p + A^p\Phi(p) \\ \\ (mp-d)v_p \leqslant cu_p + A^p\Phi(p), \end{array} \right.$$

ce qui implique

$$\left\{ \begin{array}{l} u_p \leqslant \dfrac{m(p-1)}{(mp-a)(mp-d)-bc} A^p\Phi(p) \\ v_p \leqslant \dfrac{m(p-1)}{(mp-a)(mp-d)-bc} A^p\Phi(p), \end{array} \right.$$

Nous montrerons ultérieurement (cf. Lemme 8.4.3) que, pour tout $p \geqslant 2$, $\Phi(p) \leqslant (1+8\ln(p+2))^{S+1}$. Ce résultat implique que

$$u_p \leqslant \frac{m(p-1)}{(mp-a)(mp-d)-bc} A^p \left(1+8\ln(p+2)\right)^{S+1}.$$

Par définition de $p_0$, cela implique que $(u_p)^{\frac{1}{p}} \leqslant A$ : la preuve sera donc terminée dès que nous aurons démontré le Lemme 8.4.3. ■

**Lemme 8.4.3**

*Pour tout $p \geqslant 2$, $\Phi(p) \leqslant (1+8\ln(p+2))^{S+1}$.*

**Démonstration :** Les définitions de $\varphi$ et $\Phi$ nous assurent que

$$\begin{aligned} \Phi(p) &= \sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j \leqslant p-1}} \frac{\ln^{p_1}(p_1+2)\ldots\ln^{p_{S+1}}(p_{S+1}+2)}{\ln^p(p+2)} \\ &= \sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j \leqslant p-1}} \left(1+\frac{\ln\left(1-\frac{p-p_1}{p+2}\right)}{\ln(p+2)}\right)^{p_1} \ldots \left(1+\frac{\ln\left(1-\frac{p-p_{S+1}}{p+2}\right)}{\ln(p+2)}\right)^{p_{S+1}}. \end{aligned}$$

Comme $\log(1-u) \leqslant -u$ pour tout $u<1$,

$$\Phi(p) \leqslant \sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j \leqslant p-1}} \left(1-\frac{p-p_1}{(p+2)\ln(p+2)}\right)^{p_1} \ldots \left(1-\frac{p-p_{S+1}}{(p+2)\ln(p+2)}\right)^{p_{S+1}},$$

ce qui implique

$$\Phi(p) \leqslant \sum_{\substack{p_1+\cdots+p_{S+1}=p \\ p_j \leqslant p-1}} \exp\left\{-\frac{p^2}{(p+2)\ln(p+2)} \sum_{j=1}^{S+1} \frac{p_j}{p}\left(1-\frac{p_j}{p}\right)\right\}.$$

Soit $\psi_p(x) := \exp\left(-\frac{p^2}{(p+2)\ln(p+2)}x(1-x)\right)$. Nous avons donc

$$\begin{aligned}\Phi(p) &\leqslant \sum_{\substack{p_1+\ldots+p_{S+1}=p \\ p_j\leqslant p-1}} \psi_p\left(\frac{p_1}{p}\right)\ldots\psi_p\left(\frac{p_{S+1}}{p}\right) \\ &\leqslant \sum_{0\leqslant p_1,\ldots,p_{S+1}\leqslant p-1} \psi_p\left(\frac{p_1}{p}\right)\ldots\psi_p\left(\frac{p_{S+1}}{p}\right) \\ &= \left(\sum_{k=0}^{p-1}\psi_p\left(\frac{k}{p}\right)\right)^{S+1}.\end{aligned}$$

Via un calcul élémentaire, nous obtenons

$$\sum_{k=0}^{p-1}\psi_p\left(\frac{k}{p}\right) \leqslant 1 + p\int_0^1 \psi_p(t)dt;$$

comme, pour tout $\alpha > 0$

$$\int_0^1 \exp\left(-\alpha x(1-x)\right)dt \leqslant \frac{4}{\alpha},$$

cela implique

$$\int_0^1 \psi_p(t)dt \leqslant 4\frac{(p+2)\ln(p+2)}{p^2}.$$

Dès lors,

$$\Phi(p) \leqslant \left(1 + 4\frac{(p+2)\ln(p+2)}{p}\right)^{S+1},$$

ce qui, comme $\frac{p+2}{p} \leqslant 2$ (car $p \geqslant 2$) conclut la preuve du lemme. ■

La borne supérieure des moments obtenue dans le Lemme 8.4.2 mène au théorème suivant :

**Théorème 8.4.4**

*(i) Soient $X$ et $Y$ solutions intégrables du Système d'équations en loi* (8.6)*. Dès lors, $X$ et $Y$ admettent des moments de tous ordres et les distributions de probabilité de $|X|$, $|Y|$, $X$ et $Y$ sont déterminées par leurs moments.*

*(ii) Soient $X$ et $Y$ solutions intégrables du Système* (8.5)*. Dès lors, $X$ et $Y$ admettent des séries de Laplace de rayons de convergence infinis (et sont donc déterminées par leurs moments).*

**Démonstration :** (i) Le Lemme 8.4.2 nous assure que, si $X$ et $Y$ sont solutions intégrables du Système (8.6), alors elles admettent des moments de tous ordres, et il existe une constante $C > 0$ telle que, pour tout $p$ assez grand,

$$(\mathbb{E}|X|^p)^{-\frac{1}{p}} \geqslant C\frac{(p!)^{-\frac{1}{p}}}{\ln p}. \tag{8.9}$$

Via la formule de Stirling, quand $p$ tend vers $+\infty$,

$$\frac{(p!)^{-\frac{1}{p}}}{\ln p} \sim \frac{e}{p\ln p}$$

qui est le terme général d'une série de Bertrand divergente. Le critère de Carleman [2] s'applique donc et permet de conclure que $X$ et $Y$ sont déterminées par leurs moments.

2. Le critère de Carleman est le suivant : toute variable aléatoire dont les moments $(m_p)_{p\geqslant 1}$ vérifient $\sum_{p\geqslant 1} m_{2p}^{-2p} = +\infty$ est déterminée par ses moments.

(ii) Si $X$ et $Y$ sont solutions intégrables de (8.5) et si $\xi$ est une variable aléatoire indépendante de $X$ et $Y$, de loi Gamma$\left(\frac{1}{S}\right)$, alors, au vu de la Proposition 8.2.4, $\xi X$ et $\xi Y$ sont solutions intégrables de (8.6) et satisfont donc l'Équation (8.9), ce qui implique, pour tout entier $p$,

$$\frac{\mathbb{E}|\xi^\sigma X|^p}{p!} \leqslant C^p \ln^p p.$$

Dès lors, par indépendance de $X$ et $\xi$,

$$\frac{\mathbb{E}|X|^p}{p!} \leqslant C^p \frac{\ln^p p}{\mathbb{E}|\xi|^{\sigma p}}.$$

Comme

$$\mathbb{E}|\xi|^{\sigma p} = \frac{\Gamma\left(\frac{1}{S} + \sigma p\right)}{\Gamma\left(\frac{1}{S}\right)},$$

il existe une constante $D$ positive telle que

$$\frac{\mathbb{E}|X|^p}{p!} \leqslant D^p \frac{\ln^p p}{\Gamma\left(\sigma p + \frac{1}{S}\right)}.$$

Nous pouvons en déduire que les variables $X$ et $Y$, solutions intégrables du système (8.5), ont une série de Laplace dont le rayon de convergence est infini. ■

**Corollaire 8.4.5**

*(i) Pour toute composition initiale $(\alpha, \beta)$, la loi de $W^{CT}_{(\alpha,\beta)}$ est déterminée par ses moments.*

*(ii) Pour toute composition initiale $(\alpha, \beta)$, la série de Laplace de $W^{DT}_{(\alpha,\beta)}$ est de rayon de convergence infini.*

**Démonstration :** (i) Grâce à l'Équation (8.2), nous avons

$$||W^{CT}_{(\alpha,\beta)}||_p \leqslant \alpha||W^{CT}_{(1,0)}||_p + \beta||W^{CT}_{(0,1)}||_p.$$

Comme $W^{CT}_{(1,0)}$ et $W^{CT}_{(0,1)}$ satisfont (8.9), $W^{CT}_{(\alpha,\beta)}$ vérifie le critère de Carleman et est donc déterminée par ses moments.

(ii) est une conséquence directe des Théorèmes 8.4.4 et 8.2.1. ■

## 8.5 Densité

Cette section est consacrée à l'étude des densité de $W^{DT}$ et $W^{CT}$. Si l'existence de cette densité est démontrée dans [CPS11] en temps continu, c'est un problème ouvert en ce qui concerne $W^{DT}$ : il nous faudra donc, avant toute chose, prouver son existence. Nous nous attachons ici à l'étude des variables $W^{DT}_{(1,0)}$ et $W^{DT}_{(0,1)}$ via le Système (8.5), et montrons que ces deux variables admettent chacune une densité, en généralisant des méthodes développées par Liu [Liu01] pour les équations de point fixe. Dans les travaux de Liu, il s'agit d'étudier la solution d'une équation de point fixe, solution dont le support est inclus dans $\mathbb{R}^+$. Notre problème diffère donc en deux points de celui de Liu : nous nous intéressons à un système d'équations au lieu d'une équation, et nos variables aléatoires ne sont pas, a priori, positives.

Pour montrer l'existence d'une densité pour les variables aléatoires $X = W^{DT}_{(1,0)}$ et $Y = W^{DT}_{(0,1)}$, nous allons montrer que leurs transformées de Fourier sont $L^1$, donc inversibles : leurs inverses seront alors les densités respectives de $X$ et $Y$. On notera

$$\varphi_X(t) = \mathbb{E}\mathsf{e}^{itX} \quad \text{et} \quad \varphi_Y(t) = \mathbb{E}\mathsf{e}^{itY}.$$

Nous montrons le résultat suivant :

**Théorème 8.5.1**

*Rappelons que l'on pose* $X = W^{DT}_{(1,0)}$ *et* $Y = W^{DT}_{(0,1)}$.

*(i) Le support des variables aléatoires* $X$ *et* $Y$ *est* $\mathbb{R}$.

*(ii) Pour tout* $\rho \in \left]0, \frac{a+1}{m}\right[$, *il existe une constante* $C > 0$ *telle que, pour tout* $t \in \mathbb{R}\backslash\{0\}$,

$$|\varphi_X(t)| \leqslant \frac{C}{|t|^\rho}.$$

*(iii) Pour tout* $\rho \in \left]0, \frac{d+1}{m}\right[$, *il existe une constante* $C > 0$ *telle que, pour tout* $t \in \mathbb{R}\backslash\{0\}$,

$$|\varphi_Y(t)| \leqslant \frac{C}{|t|^\rho}.$$

*(iv) Les variables aléatoires* $X$ *et* $Y$ *admettent chacune une densité bornée et continue sur* $\mathbb{R}$.

Via le Théorème 8.2.1, nous pouvons étendre ce résultat à n'importe quelle composition initiale :

**Théorème 8.5.2**

*Pour tout* $(\alpha, \beta) \in \mathbb{N}\backslash\{(0,0)\}$, *la variable aléatoire* $W^{DT}_{(\alpha,\beta)}$ *admet une densité sur* $\mathbb{R}$.

Remarquons par ailleurs, que ce résultat, couplé à la connexion (7.5), permet de redémontrer l'existence d'une densité pour $W^{CT}_{(\alpha,\beta)}$ : nous proposons donc une preuve alternative à celle développée dans [CPS11].

La preuve du Théorème 8.5.1 se décompose en plusieurs étapes : (1) nous montrons tout d'abord que le support des variables $X$ et $Y$ est $\mathbb{R}$, nous montrons ensuite que (2) les transformées de Fourier de $X$ et $Y$ ne valent 1 qu'en $t = 0$, (3) qu'elles tendent vers 0 lorsque $|t|$ tend vers $+\infty$, (4) puis qu'elles sont bornées au voisinage de 0 par une puissance de $|t|$ suffisamment négative pour pouvoir appliquer l'inversion de Fourier.

Le Lemme suivant est la première étape de cette preuve : il correspond au $(i)$ du Théorème 8.5.1 :

**Lemme 8.5.3**

*Le support des variables* $X$ *et* $Y$ *est égal à* $\mathbb{R}$.

**Démonstration :** On notera $\texttt{Supp}(X)$ et $\texttt{Supp}(Y)$ les supports respectifs de $X$ et $Y$. Comme $\mathbb{E}X > 0$ et $\mathbb{E}Y < 0$ (cf. Théorème 8.3.2), il existe $x \in \texttt{Supp}(X) \cap \mathbb{R}^+$ et il existe $y \in \texttt{Supp}(Y) \cap \mathbb{R}^-$. Au vu du Système (8.5) dont $X$ et $Y$ sont solutions, pour tout $v = (v_1, \dots, v_{S+1}), w = (w_1, \dots, w_{S+1})$ de $[0,1]^{S+1}$ tels que $\sum_{1\leqslant k\leqslant S+1} v_k = \sum_{1\leqslant k\leqslant S+1} w_k = 1$, i.e. pour tout $v, w$ dans le support d'une loi de Dirichlet de paramètres $\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$,

$$\left( x \sum_{k=1}^{a+1} v_k^\sigma + y \sum_{k=a+2}^{S+1} v_k^\sigma, x \sum_{k=1}^{c} w_k^\sigma + y \sum_{k=c+1}^{S+1} w_k^\sigma \right) \in \texttt{Supp}(X) \times \texttt{Supp}(Y). \tag{8.10}$$

Montrons tout d'abord qu'il existe $\varepsilon > 0$ tel que $[-\varepsilon, \varepsilon] \subseteq \texttt{Supp}(X) \cap \texttt{Supp}(Y)$. Pour cela, fixons $t \in [0,1]$, et appliquons l'Équation (8.10) à $v = w = (t, 0, \dots, 0, 1-t)$ : cela implique que le segment $[y, x]$ est inclus dans $\texttt{Supp}(X) \cap \texttt{Supp}(Y)$. Nous pouvons donc poser $\varepsilon = \min\{x, -y\}$ : dès lors, $[-\varepsilon, \varepsilon] \subseteq \texttt{Supp}(X) \cap \texttt{Supp}(Y)$, comme annoncé.

Montrons ensuite qu'il existe $\eta > 0$ tel que, pour tout $z \in \texttt{Supp}(X) \cap \texttt{Supp}(Y)$, $(1+\eta)z \in \texttt{Supp}(X) \cap \texttt{Supp}(Y)$. En effet, appliquons l'Équation (8.10) à $v = \left(\frac{1}{a+1}, \dots, \frac{1}{a+1}, 0, \dots, 0\right)$ et $w = \left(0, \dots, 0, \frac{1}{d+1}, \dots, \frac{1}{d+1}\right)$ (où le nombre de coordonnées nulles de $v$ et $w$ est choisi de telle sorte que ces

deux éléments soient dans $[0,1]^{S+1}$). Dès lors, si $z \in \mathtt{Supp}(X) \cap \mathtt{Supp}(Y)$, alors $(1+a)^{1-\sigma} z \in \mathtt{Supp}(X)$ et $(1+d)^{1-\sigma} z \in \mathtt{Supp}(Y)$. Il nous suffit donc de prendre $\eta = \min\{(1+a)^{1-\sigma} - 1, (1+d)^{1-\sigma} - 1\}$.

Enfin, il nous suffit de remarquer que l'image du segment $[-\varepsilon, \varepsilon]$ par les itérées de la transformation homothétique $(z \mapsto (1+\eta)z)$ est $\mathbb{R}$ pour conclure la preuve. ■

**Lemme 8.5.4**

*Pour tout $t \neq 0$, $|\varphi_X(t)| < 1$ et $|\varphi_Y(t)| < 1$.*

**Démonstration :** Pour tout réel $t$, $|\varphi_X(t)| \leqslant 1$. Supposons qu'il existe $t_0 \in \mathbb{R}$ tel que $|\varphi_X(t_0)| = 1$. Soit $\theta_0 \in \mathbb{R}$ tel que $\mathbb{E}(\mathtt{e}^{it_0 X}) = e^{i\theta_0}$. Alors, presque sûrement, $\mathtt{e}^{it_0 X} = e^{i\theta_0}$, ce qui n'est possible que si $t_0 = 0$ (et $\theta_0 \in 2\pi\mathbb{Z}$) car $\mathtt{Supp}(X) = \mathbb{R}$. Le même argument s'applique pour la variable aléatoire $Y$. ■

**Lemme 8.5.5**

*Nous avons les limites suivantes : $\lim_{|t|\to+\infty} \varphi_X(t) = 0$ et $\lim_{|t|\to+\infty} \varphi_Y(t) = 0$.*

**Démonstration :** A partir du Système (8.5), nous pouvons montrer que les transformées de Fourier de $X$ et $Y$ sont solutions d'un système d'équations. Il suffit pour cela de conditionner par rapport à $V$ puis d'utiliser l'indépendance des $X^{(k)}$ et $Y^{(k)}$ du membre de droite, avant de déconditionner : nous obtenons, pour tout réel $t$,

$$\left\{ \begin{array}{l} \varphi_X(t) = \mathbb{E}\left( \displaystyle\prod_{k=1}^{a+1} \varphi_X(tV_k^\sigma) \prod_{k=a+2}^{S+1} \varphi_Y(tV_k^\sigma) \right) \\ \varphi_Y(t) = \mathbb{E}\left( \displaystyle\prod_{k=1}^{c} \varphi_X(tV_k^\sigma) \prod_{k=c+1}^{S+1} \varphi_Y(tV_k^\sigma) \right) \end{array} \right. \tag{8.11}$$

où $V = (V_1, \ldots, V_{S+1})$ est un vecteur aléatoire de loi de Dirichlet de paramètres $\left(\frac{1}{S}, \ldots, \frac{1}{S}\right)$. Comme les $V_1, \ldots, V_{S+1}$ sont non-nulles avec probabilité 1, le Système (8.11) implique via le Lemme de Fatou que

$$\left\{ \begin{array}{l} \limsup_{|t|\to+\infty} |\varphi_X(t)| \leqslant (\limsup_{|t|\to+\infty} |\varphi_X(t)|)^{a+1} \\ \limsup_{|t|\to+\infty} |\varphi_Y(t)| \leqslant (\limsup_{|t|\to+\infty} |\varphi_Y(t)|)^{d+1}. \end{array} \right.$$

Nous avons que, comme l'urne est grande, i.e. $m = a - c = d - b > S/2$, $a, d \geqslant 1$. Dès lors, la première inégalité du système ci-dessus nous assure que $\limsup_{|t|\to+\infty} |\varphi_X(t)| \in \{0, 1\}$. Montrons que $\limsup_{|t|\to+\infty} |\varphi_X(t)| = 0$.

Supposons par l'absurde que $\limsup_{|t|\to+\infty} \varphi_X(t) = 1$, et reprenons les idées de [Liu01]. Soit $t_0 \in ]0, +\infty[$ et $\varepsilon \in ]0, 1 - |\varphi_X(t_0)|[$. Comme $\varphi_X$ est continue, et comme $1 = |\varphi_X(0)| = \limsup_{|t|\to+\infty} |\varphi_X(t)|$, il existe $t_1 = t_1(\varepsilon)$ et $t_2 = t_2(\varepsilon)$ (pour plus de lisibilité, nous omettrons souvent le paramètre $\varepsilon$) tels que $0 < t_1 < t_0 < t_2 < +\infty$, $|\varphi_X(t_1)| = |\varphi_X(t_2)| = 1 - \varepsilon$ et $|\varphi_X(t)| \leqslant 1 - \varepsilon$ pour tout $t \in [t_1, t_2]$.

Rappelons que, pour tout $t \in \mathbb{R}$,

$$|\varphi_X(t)| \leqslant \mathbb{E}\left[ |\varphi_X(V^\sigma t)|^{a+1} \right].$$

Dès lors, si $A, A_1, \ldots, A_p$ sont $p$ copies indépendantes de $V^\sigma$, pour tout $p \geqslant 0$,

$$\begin{aligned} 1 - \varepsilon = |\varphi_X(t_2)| \leqslant & \mathbb{E}\left[ |\phi_X(A_1 \ldots A_p t_2)|^{(a+1)^p} \right] \\ \leqslant & (1-\varepsilon)^{(a+1)^p} \mathbb{P}(t_1 < (A_1 \ldots A_p t_2) \leqslant t_2) + (1 - \mathbb{P}(t_1 < (A_1 \ldots A_p t_2) \leqslant t_2)) \\ = & 1 - (1 - (1-\varepsilon)^{(a+1)^p}) \mathbb{P}(t_1 < (A_1 \ldots A_p t_2) \leqslant t_2), \end{aligned}$$

ce qui implique

$$\frac{1 - (1-\varepsilon)^{(a+1)^p}}{1 - (1-\varepsilon)} \mathbb{P}(t_1 < (A_1 \ldots A_p t_2) \leqslant t_2) \leqslant 1.$$

Comme $\lim_{\varepsilon\to 0}\frac{1-(1-\varepsilon)^{(a+1)^p}}{1-(1-\varepsilon)}=(a+1)^p$, $\lim_{\varepsilon\to 0} t_1(\varepsilon)=0$ et $\frac{t_1(\varepsilon)}{t_2(\varepsilon)}\leqslant\frac{t_1(\varepsilon)}{t_0}\to 0$ quand $\varepsilon$ tend vers 0, nous obtenons,

$$(a+1)^p\mathbb{P}(0\leqslant A_1\dots A_p\leqslant 1)=(a+1)^p\leqslant 1 \text{ pour tout } n\geqslant 0,$$

ce qui est impossible car $a+1\geqslant 1$, par hypothèse.

Nous avons donc montré que $\limsup_{t\to+\infty}|\varphi_X(t)|=0$. Nous pouvons montrer de la même manière que $\limsup_{t\to-\infty}|\varphi_X(t)|=0$. Enfin, nous pouvons montrer avec les même arguments que $\limsup_{|t|\to\infty}|\varphi_Y(t)|=0$ et conclure ainsi la preuve. ■

**Lemme 8.5.6**

*Pour tout $\rho\in]0,\frac{1}{m}[$, asymptotiquement quand $|t|$ tend vers $+\infty$, $\varphi_X(t)=\mathcal{O}(t^{-\rho})$ et $\varphi_Y(t)=\mathcal{O}(t^{-\rho})$.*

**Démonstration :** Soit $\varepsilon>0$. Le Lemme 8.5.5 nous assure qu'il existe $T>0$ tel que $|\varphi_X(t)|\leqslant\varepsilon$ et $|\varphi_Y(t)|\leqslant\varepsilon$ pour tout réel $t$ tel que $|t|\geqslant T$. Alors, au vu du Système (8.11), pour tout $t\in\mathbb{R}$,

$$|\varphi_X(t)|\leqslant\varepsilon^S\mathbb{E}|\varphi_X(V_{S+1}^\sigma t)|+\sum_{k=1}^{S}\mathbb{P}(V_k^\sigma|t|\leqslant T).$$

Comme, pour tout $k\in\{1,\dots,S+1\}$, $V_k^\sigma\overset{(loi)}{=}U^m$ où $U$ est une variable aléatoire uniforme sur $[0,1]$, cela implique que

$$|\varphi_X(t)|\leqslant\varepsilon^S\mathbb{E}|\varphi_X(U^m t)|+S\left(\frac{T}{|t|}\right)^{\frac{1}{m}},$$

et ce pour tout $t\in\mathbb{R}\backslash\{0\}$. Dès lors, pour tout $\rho\in]0,1/m[$, $\mathbb{E}\left(U^{-m\rho}\right)<+\infty$ et l'inégalité précédente implique qu'il existe une constante positive $C$ telle que pour tout réel non nul $t$,

$$|\varphi_X(t)|\leqslant\varepsilon^S\mathbb{E}|\varphi_X(U^m t)|+C\left(\frac{1}{|t|}\right)^{\rho}.$$

Dès lors, les variables aléatoires $X$ et $U$ satisfont les hypothèses du lemme *à la* Gronwall [Liu01, Lemme 3.2, page 93]) : nous pouvons itérer l'inégalité précédente, et, pour tout entier $p$,

$$|\varphi_X(t)|\leqslant\varepsilon^{pS}\mathbb{E}|\varphi_X(U_1^m\dots U_p^m t)|+C|t|^{-\rho}\sum_{k=0}^{p-1}\left(\varepsilon^S\mathbb{E}(U^{-m\rho})\right)^k,$$

ce qui induit

$$|\varphi_X(t)|\leqslant\frac{C|t|^{-\rho}}{1-\varepsilon^S\mathbb{E}\left(U^{-m\rho}\right)}$$

dès que $\varepsilon$ est tel que $1-\varepsilon^S\mathbb{E}\left(U^{-m\rho}\right)>0$. Le même raisonnement appliqué à $\varphi_Y$ permet de conclure la preuve. ■

**Démonstration du Théorème 8.5.1 :** Soit $\rho\in]0,\frac{a+1}{m}[$ et soit $\nu=\frac{\rho}{a+1}$. Le Lemme 8.5.6 nous assure qu'il existe $\kappa>0$ tel que $\varphi_X(t)\leqslant\kappa|t|^{-\nu}$ pour tout $t$ réel non nul. Via le Système (8.11), nous obtenons

$$|\varphi_X(t)|\leqslant\mathbb{E}\left(\prod_{k=1}^{a+1}|\varphi_X(V_k^\sigma t)|\right)\leqslant\frac{\kappa^{a+1}}{|t|^\rho}\mathbb{E}\left(\prod_{k=1}^{a+1}V_k^{-\sigma\nu}\right)$$

Comme le vecteur aléatoire $V=(V_1,\dots,V_{S+1})$ est de loi Dirichlet de paramètres $\left(\frac{1}{S},\dots,\frac{1}{S}\right)$, ses moments joints sont connus, et

$$\mathbb{E}\left(\prod_{k=1}^{a+1}V_k^{-\sigma\nu}\right)=\frac{\Gamma\left(1+\frac{1}{S}\right)}{\Gamma\left(1+\frac{1}{S}-(a+1)\sigma\nu\right)}\left(\frac{\Gamma\left(\frac{1}{S}-\sigma\nu\right)}{\Gamma\left(\frac{1}{S}\right)}\right)^{a+1}$$

est fini car $\sigma\nu < \frac{\sigma}{m} = \frac{1}{S}$ et $1 + \frac{1}{S} - (a+1)\sigma\nu > 1 - \frac{a}{S} > 0$ ($a < S = a + b$ car l'urne considérée n'est pas triangulaire). Nous avons donc prouvé l'assertion $(ii)$. L'assertion $(iii)$ est prouvée de manière similaire.

Comme $\frac{a+1}{m} = \frac{a+1}{a-c} > 1$, l'assertion $(ii)$ implique que la transformée de Fourier de $\varphi_X$ de $X$ est intégrable, nous assurant ainsi que $X$ admet une densité bornée et continue. Bien entendu, un argument similaire permet de traiter la variable aléatoire $Y$ et de conclure la preuve. ■

## 8.6 Conclusion

Grâce à l'étude de la structure arborescente de l'urne de Pólya, nous avons pu montrer différents résultats sur les variables $W^{CT}$ et $W^{DT}$. Nous avons montré que $W^{DT}$ admet une densité sur $\mathbb{R}$, propriété qui était déjà connue pour $W^{CT}$ [CPS11]. Remarquons au passage que la densité de $W^{DT}$ est continue bornée, ce qui n'est pas le cas de celle de $W^{CT}$ qui est minorée, à une constante près, par $|t|^{1/m-1}$ au voisinage de zéro. De plus, la transformée de Fourier de $W^{DT}$ est $L^1$, ce qui n'est pas le cas de celle de $W^{CT}$.

Nous avons aussi montré que, aussi bien en temps discret qu'en temps continu, les variables $W$ sont déterminées par leurs moments. Ce résultat permet de majorer les moments de ces variables aléatoires, même si l'ordre exact de ces moments n'est pas encore connu : une borne inférieure est connue pour $W^{CT}$ via la nullité du rayon de convergence de sa série de Laplace [CPS11]. Nous avons aussi démontré que la série de Laplace de la variable $W^{DT}$ a un rayon de convergence infini.

Ces différents résultats nous permettent de mieux appréhender les variables $W$. Nous comprenons mieux désormais les différences entre la loi de $W^{DT}$ et celle de $W^{CT}$. La loi de $W^{DT}$ semble plus *régulière* que celle de $W^{CT}$ : sa série de Laplace est convergente, sa transformée de Fourier est $L^1$, sa densité est continue, bornée.

De nombreux questions restent cependant ouvertes concernant ces variables aléatoires $W^{DT}$ et $W^{CT}$ : quel est l'ordre exact de leurs moments ? leurs queues sont-elles lourdes ?

Au delà de ces perspectives concernant les urnes à deux couleurs, il est très intéressant de pouvoir généraliser les méthodes développées dans ce chapitre à des urnes à $d$ couleurs, où de telles variables aléatoires $W$ apparaissent et sont encore plus méconnues que celles du cas à deux couleurs. C'est l'objet du Chapitre 9, dans lequel nous présentons des travaux en cours concernant cette généralisation naturelle.

# Chapitre 9

# Urnes à $d$ couleurs

## 9.1 Introduction

### 9.1.1 Motivations

L'objectif de ce chapitre est d'étendre les résultats du Chapitre 8 pour des urnes à $d$ couleurs. Une urne à $d$ couleurs est décrite comme suit. Fixons $\alpha_1, \ldots, \alpha_d \geqslant 0$ des entiers tels que $\alpha_1 + \ldots + \alpha_d > 0$ et $R = (a_{i,j})$ une matrice d'entiers naturels. À l'instant initial, il y a $\alpha_i$ boules de couleur $i$, pour tout $i \in \{1, \ldots, d\}$. À chaque étape, nous piochons uniformément au hasard une boule dans l'urne, regardons sa couleur (notons $i$ cette couleur), la remettons dans l'urne et ajoutons dans l'urne $a_{ij}$ boules de couleur $j$. Cette définition est la généralisation naturelle d'une urne à deux couleurs.

Les urnes à $d$ couleurs ont été très étudiées dans la littérature comme évoqué dans le Chapitre 7. Il est intéressant de noter que l'approche par combinatoire analytique, très efficace pour les urnes à deux couleurs, ne permet pas, sauf exceptions, de traiter les urnes à $d$ couleurs. Comme expliqué dans les travaux de thèse de Morcrette [Mor13], cette méthode repose sur la résolution d'un système de $d$ équations différentielles, qui n'est pas résoluble, en général.

Tout comme dans le chapitre précédent, nous ferons certaines hypothèses classiques sur les urnes que nous considèrerons. Nous supposerons que ces urnes sont *équilibrées*, i.e. qu'il existe un entier $S$ tel que, pour tout $i \in \{1, \ldots, d\}$, $\sum_{j=1}^{d} a_{i,j} = S$. Cet entier $S$ est appelé *balance* de l'urne, cela signifie qu'à chaque étape, le nombre de boules ajoutées dans l'urne est déterministe et égal à $S$. Tout comme dans le cas à deux couleurs, nous supposerons que la probabilité d'extinction de l'urne est nulle, c'est à dire que nous n'aboutissons jamais à une situation impossible (par exemple, il nous faut retirer une boule de couleur 1 de l'urne alors que l'urne ne contient pas de boule de couleur 1). Il est possible de se passer de cette hypothèse comme dans les travaux de Janson [Jan04], et l'étude peut alors être réalisée conditionnellement à la non-extinction de l'urne. Pour plus de simplicité, nous supposerons la *non-extinction* presque sûre de l'urne. Un travail supplémentaire, que nous ne ferons pas, serait nécessaire pour travailler conditionnellement à la non-extinction. Enfin, nous supposerons que l'urne est *irréductible*. Cette hypothèse est l'équivalent de la forme non-triangulaire de l'urne à deux couleurs ($bc \neq 0$ avec les notations du Chapitre 8) : nous supposons que, quelle que soit la composition initiale de l'urne, pour tout $i \in \{1, \ldots, d\}$, presque sûrement, il existe un entier $n$ tel que l'urne à l'étape $n$ contient au moins une boule de couleur $i$. Cette condition de *mélange* est nécessaire pour l'étude que nous ferons : l'absence de cette hypothèse mène à des résultats différents (voir par exemple [Pou08], ou [FDP06] dans le cas particulier des urnes à trois couleurs triangulaires).

Les urnes à $d$ couleurs ont le même comportement que les urnes à 2 couleurs, à ceci près que l'on ne peut plus parler de *grande* ou *petite* urne mais seulement de *grande* ou *petite* valeur propre

d'une urne. En effet, une urne à 2 couleurs est définie par une matrice de remplacement $R$ carrée de dimension 2. Cette matrice a donc deux valeurs propres : la balance $S$ et une seconde valeur propre notée $m$ dans le Chapitre 8 : rappelons que l'urne à 2 couleurs est petite si $\frac{m}{S} \leqslant \frac{1}{2}$ et grande si $\frac{m}{S} > \frac{1}{2}$. Dans le cas d'une urne à $d$ couleurs, la matrice $R$ est une matrice carrée de dimension $d$ ; outre $S$, elle admet un certain nombre d'autres valeurs propres, éventuellement complexes non réelles. On dira qu'une valeur propre $\lambda$ est grande si $1 > \sigma = \frac{\mathrm{Re}\lambda}{S} > \frac{1}{2}$, sinon, elle sera petite. Ainsi, une même urne de matrice de remplacement $R$ peut avoir différentes valeurs propres, certaines grandes et certaines petites. La valeur propre $S$ peut aussi être valeur propre multiple, ce qui n'était pas possible en dimension 2.

Le passage en dimension $d$ apporte donc deux difficultés : la multiplicité éventuelle de la valeur propre $S$, et l'apparition de plusieurs autres valeurs propres, éventuellement complexes et non réelles. Malgré cela, des théorèmes limites existent, montrés par Janson [Jan04], Pouyanne [Pou08] ou Gouet [Gou97], tout comme dans le cas d'urnes à 2 couleurs. Il s'agit de projeter le vecteur composition sur un sous-espace associé à la décomposition de Jordan de la matrice $R$. En temps continu, les lois limites sont gaussiennes dans le cas d'une petite valeur propre et font intervenir une variable aléatoire méconnue $W$ dans le cas d'une grande valeur propre. En temps discret, les résultats sont moins précis dans le cas des petites valeurs propres. Rappelons qu'il faut aussi traiter le cas des sous-espaces associés à la valeur propre $S$ à part. Tout comme dans le cas des urnes à deux couleurs, nous nous intéressons à la variable aléatoire $W$, en détail. Contrairement au cas deux couleurs, $W$ n'est à ce jour pas étudiée dans la littérature. L'intérêt de notre approche via la structure arborescente (cf. Chapitre 8) de l'urne est que cette approche semble facilement généralisable au cas de $d$ couleurs.

Tout comme dans le cas de deux couleurs, le processus se plonge en temps continu et nous avons donc deux variables $W$ à étudier : celle en temps discret $W^{DT}$ et celle en temps continu $W^{CT}$. De plus, cette variable aléatoire dépend de la composition initiale de l'urne.

Dans la Section 9.2, nous exploiterons la structure arborescente de l'urne en vue de réduire l'étude à un nombre fini de compositions initiales, et donc à un nombre fini de variables $W$ (plus précisément à l'étude de $d$ variables aléatoires), puis de montrer que ces $d$ variables aléatoires sont solutions d'un système de point fixe en loi. La suite du chapitre est consacrée à extraire un maximum d'information de ce système d'équations en loi. Nous montrerons, en Section 9.3, via le théorème de point fixe de Banach, que la solution de ce système est unique dans un espace de mesures approprié. Dans la Section 9.4, nous étudierons les moments des variables aléatoires $W$ et montrerons que ces variables sont déterminées par leurs moments, et que la variable aléatoire $W$ issue du processus en temps discret admet une série de Laplace de rayon de convergence infini. Puis, nous évoquerons en Section 9.5 quelques travaux en cours en vue de montrer que ces variables aléatoires admettent une densité.

Avant tout, rappelons les résultats de la littérature résumant le comportement asymptotique des urnes à $d$ couleurs.

### 9.1.2 Résultats antérieurs

Soit $U(n)$ le vecteur composition de l'urne à $d$ couleurs de matrice de remplacement $R = (a_{i,j})_{i,j\in\{1,\dots,d\}}$ et de composition initiale $U(0) =^t (\alpha_1, \dots, \alpha_d)$. **Dans toute la suite, nous supposerons que cette urne est équilibrée, irréductible et quil y a non-extinction presque sûre.**

La matrice $R$ se décompose sous la forme de blocs de Jordan, i.e. elle est semblable à une matrice

de la forme d'une matrice diagonale par blocs $\texttt{diag}(J_1, \dots, J_r)$ où chaque $J_i$ est de la forme

$$J = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \dots & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix},$$

avec $\lambda$ une valeur propre de $R$. Plusieurs blocs de Jordan peuvent être associés à la même valeur propre. Dans la suite, nous choisissons un bloc de Jordan, et non une valeur propre, et étudions le comportement de la projection du vecteur composition sur le sous-espace stable associé à ce bloc de Jordan.

Soit $E$ un sous-espace stable de $\mathbb{R}^d$ associé à un bloc de Jordan $J$ de la décomposition de Jordan de $R$ et à une valeur propre $\lambda$ de $R$. On note $\nu + 1$ la taille de ce bloc de Jordan et on appelle $v$ un vecteur propre de $E$ associé à la valeur propre $\lambda$.

**Définition 9.1.1**

*Soit $\sigma = \frac{\mathrm{Re}\lambda}{S}$. On dira que $\lambda$ est une* **grande valeur propre** *de $R$ si $\frac{1}{2} < \sigma < 1$, et une* **petite valeur propre** *si $\sigma \leqslant \frac{1}{2}$.*

Le comportement asymptotique de la projection du vecteur composition $U(n)$ sur le sous-espace $E$ est décrit par le théorème suivant :

**Théorème 9.1.2 (cf. [Pou08])**

*Si $\frac{1}{2} < \sigma < 1$, alors,*

$$\lim_{n\to\infty} \frac{\pi_E(U(n))}{n^{\lambda/S} \ln^{\nu} n} = \frac{1}{\nu!} W^{DT} v,$$

*p.s. et dans tout $L^p$ ($p \geqslant 1$), où $\pi_E(U(n))$ est la projection du vecteur composition au temps $n$ sur $E$ (définie par la décomposition de Jordan de $R$).*

Il s'avère que le comportement des projections du vecteur composition en temps discret sur les blocs de Jordan associés à des petites valeurs propres n'est pas connu à ce jour en toute généralité, mais seulement si toutes les valeurs propres (excepté $S$) sont petites, et dans le cas où la plus grande valeur propre vérifie $\sigma = \frac{1}{2}$, seulement selon le plus grand bloc de Jordan associé à cette valeur propre (cf. [Jan04, Théorèmes 3.22 et 3.23]). Nous nous intéressons dans ce mémoire aux seules grandes valeurs propres, mais montrer un Théorème concernant tous les blocs de Jordan associés à des petites valeurs propres en temps discret (même quand d'autres valeurs propres de la matrice sont grande) permettrait de compléter notre connaissance des urnes à $d$ couleurs. Nous verrons ultérieurement que ce comportement gaussien selon les petites valeurs propres est connu en temps continu [Jan04].

Plongeons le processus d'urnes à $d$ couleurs en temps continu comme dans le Chapitre 8 : chaque boule est associée à un réveil qui sonne au bout d'un temps aléatoire de loi exponentielle de paramètre 1, indépendamment des autres réveils. Quand un réveil associé à une boule de couleur $i$ sonne, alors elle meurt et donne naissance à $a_{i,j} + \delta_{i,j}$ boules de couleur $j$, et ce pour tout $j \in \{1, \dots, d\}$ (où $\delta_{i,j} = 1$ si $i = j$ et 0 sinon). Tout comme pour les urnes à 2 couleurs, on note $\tau_n$ la date de la $n^{\text{ième}}$ sonnerie dans l'urne, et si l'on note $U^{CT}(t)$ le vecteur composition de l'urne en temps continu au temps $t$. On a la connexion suivante : presque sûrement,

$$(U(n))_{n\geqslant 0} = (U^{CT}(\tau_n))_{n\geqslant 0}. \tag{9.1}$$

De plus, le processus $(U(n))_{n\geqslant 0}$ est indépendant de la suite de temps d'arrêt $(\tau_n)_{n\geqslant 0}$.

En temps continu, le comportement du vecteur composition projeté sur un sous-espace stable associé à une grande valeur propre de $R$ vérifie, avec les mêmes notations que précédemment,

**Théorème 9.1.3 (cf. [Jan04])**

*Si $\frac{1}{2} < \sigma < 1$, alors, presque sûrement,*

$$\lim_{t\to+\infty} \frac{\pi_E(U(t))}{t^\nu \mathsf{e}^{\lambda t}} = \frac{1}{\nu!} W^{CT} v,$$

*où $\pi_E$ est la projection sur le sous-espace $E$. De plus, la variable aléatoire $W^{CT}$ admet des moments de tous ordres.*

Nous nous intéressons dans la suite de ce chapitre à l'étude des variables aléatoires $W^{DT}$ et $W^{CT}$ définies dans les théorèmes limites 9.1.2 et 9.1.3. Ces variables aléatoires dépendent en fait de la composition initiale de l'urne $\alpha = (\alpha_1, \dots, \alpha_d)$. Nous noterons donc $W^{DT}_{\boldsymbol{\alpha}}$ (resp. $W^{CT}_{\boldsymbol{\alpha}}$) la variable $W$ associée à la composition initiale $\boldsymbol{\alpha}$.

La connexion (9.1) permet de montrer des relations entre les variables aléatoires $W$ issues du processus en temps discret et du processus en temps continu. Pour cela nous avons besoin du résultat suivant :

$$\lim_{n\to\infty} n\mathsf{e}^{-S\tau_n} = \xi, \tag{9.2}$$

presque sûrement, où $\xi$ est une variable aléatoire de distribution Gamma$(\frac{\alpha_1+\dots+\alpha_d}{S})$. Ce résultat est en fait démontré pour une urne à 2 couleurs dans [CPS11], mais la preuve reste la même pour une urne à $d$ couleurs car le seul paramètre important est la balance $S$.

Nous avons

$$\frac{\pi_E\left(U^{CT}(\tau_n)\right)}{\tau_n^\nu \mathsf{e}^{\lambda\tau_n}} = \frac{\pi_E\left(U^{DT}(n)\right)}{n^{\lambda/S}\ln^\nu n} \cdot \frac{n^{\lambda/S}\ln^\nu n}{\tau_n^\nu \mathsf{e}^{\lambda\tau_n}}.$$

De plus, au vu de l'Équation (9.2) qui implique $\frac{\ln n}{\tau_n} \to S$ quand $n$ tend vers $+\infty$,

$$\frac{n^{\lambda/S}\ln^\nu n}{\tau_n^\nu \mathsf{e}^{\lambda\tau_n}} = \left(\frac{\ln n}{\tau_n}\right)^\nu (n\mathsf{e}^{-S\tau_n})^{\lambda/S} \to S^\nu \xi^{\lambda/S},$$

Ce qui implique [Jan04] que, pour toute composition initiale $\boldsymbol{\alpha}$, nous avons la connexion suivante :

$$W^{CT}_{\boldsymbol{\alpha}} \overset{(loi)}{=} S^\nu \xi^{\lambda/S} W^{DT}_{\boldsymbol{\alpha}}, \tag{9.3}$$

où $\xi$ est une variable aléatoire de loi Gamma$\left(\frac{\alpha_1+\dots+\alpha_d}{S}\right)$, et où $\xi$ et $W^{DT}$ sont indépendantes.

Par ailleurs, $(U^{DT}(n(t)))_{t\geqslant 0} = (U^{CT}(t))_{t\geqslant 0}$ presque sûrement, où $n(t)$ est le nombre de boules dans l'urne au temps $t$. Nous pouvons donc en déduire que, pour toute composition initiale $\boldsymbol{\alpha}$,

$$W^{DT}_{\boldsymbol{\alpha}} \overset{(loi)}{=} S^{-\nu} \xi^{-\lambda/S} W^{CT}_{\boldsymbol{\alpha}}, \tag{9.4}$$

où $\xi$ est une variable aléatoire de loi Gamma$\left(\frac{\alpha_1+\dots+\alpha_d}{S}\right)$ mais où $\xi$ et $W^{CT}_{\boldsymbol{\alpha}}$ *ne* sont *pas* indépendantes, ce qui peut être constaté via un calcul de covariance.

Dans la Section suivante, nous allons montrer que, grâce à la structure arborescente de l'urne, nous pouvons nous ramener à l'étude de seulement $d$ compositions initiales au lieu d'une infinité. Il nous suffira de considérer $W_{\boldsymbol{e_1}}, \dots, W_{\boldsymbol{e_d}}$ où, pour tout $i \in \{1, \dots, d\}$, $\boldsymbol{e_i}$ est le vecteur dont toutes les coordonnées sont 0 sauf la $i^{\text{ième}}$ qui vaut 1. Il nous suffit donc de considérer les $d$ compositions initiales différentes composées d'une seule boule. Nous montrerons ensuite que ces $d$ variables aléatoires $W_{\boldsymbol{e_1}}, \dots, W_{\boldsymbol{e_d}}$ vérifient un système de $d$ équations en loi. Ce cheminement est similaire à celui du Chapitre 8.

## 9.2 Arborescence de l'urne

### 9.2.1 Décomposition

Le raisonnement étant le même que dans le Chapitre 8, nous serons plus rapides.

Tout comme dans le cas d'une urne à 2 couleurs, le processus d'urne en temps discret peut être vu comme une forêt dont les feuilles peuvent être de $d$ couleurs différentes. A l'étape initiale, la forêt est composée de $\alpha_i$ racines de couleur $i$, pour tout $i \in \{1, \dots, d\}$. À chaque étape, on pioche une feuille de la forêt uniformément au hasard (n.b. une racine est aussi une feuille). Cette feuille devient alors un nœud interne qui a $S+1$ enfants, dont $a_{i,j} + \delta_{i,j}$ sont de couleur $j$ pour tout $j \in \{1, \dots, d\}$ où $i$ est la couleur de la feuille piochée au hasard.

La composition de l'urne est décrite par l'ensemble des feuilles de la forêt. Dès lors, si l'on note $D_p(n)$ le nombre de feuilles dans le $p^{ième}$ sous-arbre de la forêt, alors, le $p^{\text{ième}}$ sous-arbre de la forêt au temps $n$ est une forêt qui était initialement réduite à une seule racine, et qui est prise au *temps interne* $\frac{D_p(n)-1}{S}$. Nous obtenons donc

$$U_{\boldsymbol{\alpha}}(n) \overset{(loi)}{=} \sum_{c=1}^{d} \sum_{p=\beta_{c-1}+1}^{\beta_c} U_{\boldsymbol{e_c}}^{(p)} \left( \frac{D_p(n)-1}{S} \right), \tag{9.5}$$

où $\beta_0 = 0$, et, pour tout $c \geqslant 1$, $\beta_c = \sum_{i=1}^{c} \alpha_i$ et où les processus d'urnes $U_{\boldsymbol{e_c}}^{(p)}$ sont des copies indépendantes des processus $U_{\boldsymbol{e_c}}$.

Rappelons aussi que le vecteur $(D_1(n), \dots, D_{\alpha_1+\dots+\alpha_d}(n))$ est de même loi que le vecteur composition d'une urne de Pólya-Eggenberger de composition initiale $(1, \dots, 1)$ et de matrice de remplacement $SI_{\alpha_1+\dots+\alpha_d}$. Rappelons que (cf. Théorème 7.3.1), asymptotiquement quand $n$ tend vers $+\infty$, presque sûrement,

$$\frac{1}{nS} \left( D_1(n), \dots, D_{\beta_d}(n) \right) \to Z = (Z_1, \dots, Z_{\alpha_1+\dots+\alpha_d})$$

où $Z$ est un vecteur aléatoire de loi de Dirichlet$\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$.

Dès lors, en divisant l'Équation (9.5) par $n^{\lambda/S} \ln^{\nu} n$ et en prenant son image par la projection $\pi_E$ sur $E$, nous obtenons :

$$\frac{\pi_E(U_{\boldsymbol{\alpha}}(n))}{n^{\lambda/S} \ln^{\nu} n} \overset{(loi)}{=} \sum_{c=1}^{d} \sum_{p=\beta_{c-1}+1}^{\beta_c} \frac{\pi_E\left( U_{\boldsymbol{e_c}}^{(p)} \left( \frac{D_p(n)-1}{S} \right) \right)}{n^{\lambda/S} \ln^{\nu} n}.$$

Puis, par passage à la limite, via les Théorèmes 7.3.1 et 9.1.2, nous obtenons

**Théorème 9.2.1**

*Pour toute composition initiale $\boldsymbol{\alpha}$,*

$$W_{\boldsymbol{\alpha}}^{DT} \overset{(loi)}{=} \sum_{c=1}^{d} \sum_{p=\beta_{c-1}+1}^{\beta_c} Z_p^{\lambda/S} W_{\boldsymbol{e_c}}^{(p)},$$

*où $Z = (Z_1, \dots, Z_{\beta_d})$ ($\beta_d = \alpha_1 + \dots + \alpha_d$) est un vecteur aléatoire de loi de Dirichlet de paramètres $\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$, et où les $W_{\boldsymbol{e_c}}^{(p)}$ sont des copies indépendantes de $W_{\boldsymbol{e_c}}^{DT}$, indépendantes entre elles.*

En temps continu, nous pouvons aussi faire le parallèle avec une forêt, mais le raisonnement est encore plus simple car les horloges exponentielles nous assurent que les sous-arbres de la forêt sont indépendants. Nous pouvons donc montrer que

$$U_{\boldsymbol{\alpha}}^{CT}(t) \overset{(loi)}{=} \sum_{c=1}^{d} \sum_{p=\beta_{c-1}+1}^{\beta_c} U_{\boldsymbol{e_c}}^{(p)}(t-\tau_1 Fson),$$

où $\beta_0 = 0$ et $\beta_c = \sum_{j\leqslant c} \alpha_j$, et où les $U_{\boldsymbol{e_c}}^{(p)}(t)$ sont des copies indépendantes de $U_{\boldsymbol{e_c}}^{CT}(t)$.

En divisant cette égalité en loi par $t^\nu \mathsf{e}^{\lambda t}$ puis en projetant sur $E$ via $\pi_E$, via le Théorème 9.1.3, nous obtenons

**Théorème 9.2.2**

*Pour toute composition initiale $\boldsymbol{\alpha}$,*

$$W_{\boldsymbol{\alpha}}^{CT} \overset{(loi)}{=} U^\lambda \sum_{c=1}^{d} \sum_{p=\beta_{c-1}+1}^{\beta_c} W_{\boldsymbol{e_c}}^{(p)},$$

*où $U$ est une variable aléatoire uniforme sur $[0,1]$ et où les $W_{\boldsymbol{e_c}}^{(p)}$ sont des copies indépendantes de $W_{\boldsymbol{e_c}}^{CT}$, indépendantes entre elles et indépendantes de $U$.*

Les Théorèmes 9.2.1 et 9.2.2 permettent, aussi bien en temps discret qu'en temps continu, de se ramener à l'étude de $d$ variables aléatoires $(W_{\boldsymbol{e_1}}, \dots, W_{\boldsymbol{e_d}})$ au lieu d'une infinité. Toute information obtenue concernant ces $d$ variables aléatoires pourra a priori être traduite pour n'importe quelle composition initiale $\boldsymbol{\alpha}$.

### 9.2.2 Dislocation

Dans cette partie, nous réutilisons la structure arborescente de l'urne de façon à montrer que les $d$ variables aléatoires $W_{\boldsymbol{e_1}}, \dots, W_{\boldsymbol{e_d}}$ sont solutions d'un système de $d$ équations en loi.

Plaçons-nous tout d'abord en temps discret, et considérons l'urne contenant à l'étape initiale une unique boule, de couleur $c \in \{1, \dots, d\}$. La première étape est déterministe : la première boule piochée est la boule de couleur $c$ et l'urne contient donc après cette première étape, $S+1$ boules, dont $a_{c,i} + \delta_{c,i}$ de couleur $i$, pour tout $i \in \{1, \dots, d\}$. Autrement dit, la composition de l'urne peut être représentée par les feuilles d'une forêt initialement composée de $S+1$ racines, dont $a_{c,i} + \delta_{c,i}$ de couleur $i$ (cf. Figure 9.1).

Si l'on note $J_p(n)$ le nombre de feuilles dans le $p^{\text{ième}}$ sous-arbre de la forêt, alors,

$$U_{\boldsymbol{e_c}}(n) \overset{(loi)}{=} \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} U_{\boldsymbol{e_i}}^{(p)}\left(\frac{J_p(n)-1}{S}\right), \tag{9.6}$$

où $\gamma_0 = 0$ et $\gamma_i = \sum_{j=1}^{i}(a_{c,j}+\delta_{j,c})$ pour tout $i \in \{1, \dots, d\}$ et où les $U_{\boldsymbol{e_i}}^{(p)}$ sont des copies indépendantes du processus $U_{\boldsymbol{e_i}}$.

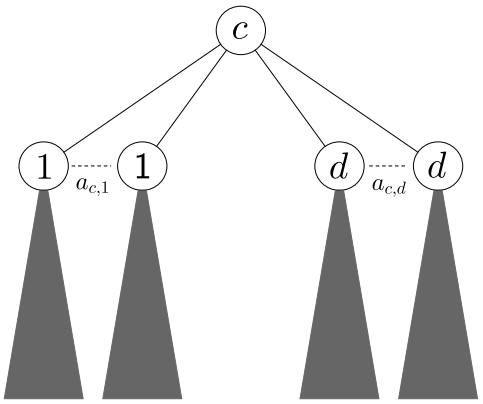
Le Théorème 7.3.1 nous assure que, presque sûrement, asymptotiquement quand $n$ tend vers $+\infty$,

$$\frac{1}{nS}(J_1(n), \dots, J_{S+1}(n)) \to V = (V_1, \dots, V_{S+1}),$$

où $V$ est un vecteur aléatoire de loi de Dirichlet de paramètres $\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$.

Nous obtenons ainsi, via le Théorème 9.1.2, après renormalisation et projection de l'Équation (9.6), le

FIGURE 9.1 – *Arborescence issue d'une boule de couleur c, dans le cas particulier où $c \neq 1$ et $c \neq d$.*




**Théorème 9.2.3**

*Pour toute couleur $c \in \{1, \dots, d\}$,*

$$W_{\boldsymbol{e_c}}^{DT} \overset{(loi)}{=} \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} V_p^{\lambda/S} W_{\boldsymbol{e_i}}^{(p)}, \tag{9.7}$$

*où $\gamma_0 = 0$ et $\gamma_i = \sum_{j=1}^{i}(a_{c,j} + \delta_{c,j})$ pour tout $i \in \{1, \dots, d\}$ ; où les $W_{\boldsymbol{e_i}}^{(p)}$ sont des copies indépendantes de $W_{\boldsymbol{e_i}}^{DT}$, indépendantes les unes des autres ; et où le vecteur $V = (V_1, \dots, V_{\gamma_d})$ est de loi de Dirichlet$\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$ et indépendant des $W$.*

Nous obtenons un résultat identique en temps continu, c'est le Théorème 3.9 de [Jan04] :

**Théorème 9.2.4 ([Jan04])**

*Pour toute couleur $c \in \{1, \dots, d\}$,*

$$W_{\boldsymbol{e_c}}^{CT} \overset{(loi)}{=} U^{\lambda} \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} W_{\boldsymbol{e_i}}^{(p)}, \tag{9.8}$$

*où $\gamma_0 = 0$, $\gamma_i = \sum_{j \leqslant i}(a_{c,j} + \delta_{c,j})$, où $U$ est une variable aléatoire uniforme sur $[0, 1]$ et où les $W_{\boldsymbol{e_i}}^{(p)}$ sont des copies indépendantes de $W_{\boldsymbol{e_i}}^{CT}$, indépendantes les unes des autres et indépendantes de $U$.*

### 9.2.3 Connexion

Tout comme dans le cas des urnes à deux couleurs, il est intéressant de noter que l'on peut déduire le système vérifié par $(W_{\boldsymbol{e_i}}^{CT})_{i \in \{1,\dots,d\}}$ de celui vérifié par $(W_{\boldsymbol{e_i}}^{DT})_{i \in \{1,\dots,d\}}$, et ce via la connexion (9.3) :

**Proposition 9.2.5**

*Soient $X_1, \dots, X_d$ solutions du Système (9.7) et soit $\xi$ une variable aléatoire de loi Gamma$\left(\frac{1}{S}\right)$. Alors les variables aléatoires $S^{\nu}\xi^{\frac{\lambda}{S}}X_1, \dots, S^{\nu}\xi^{\frac{\lambda}{S}}X_d$ sont solutions du Système (9.8).*

Cette proposition, via la connexion (9.3), nous assure que le Théorème 9.2.4 peut être vu comme une conséquence du Théorème 9.2.3. L'implication réciproque, si elle existe, n'est pas encore connue. Cette proposition est une conséquence du lemme suivant, que nous ne redémontrons pas tant il est similaire au Lemme 8.2.5 :

**Lemme 9.2.6**

*Considérons les deux équations en loi suivantes d'inconnues $X, X_1, \dots, X_{S+1}$ :*

$$X \overset{(loi)}{=} \sum_{k=1}^{S+1} V_k^{\lambda/S} X_k \tag{9.9}$$

*où $V = (V_1, \dots, V_{S+1})$ est un vecteur aléatoire de loi de Dirichlet de paramètre $\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$ indépendant des $X_1, \dots, X_{S+1}$ ; et*

$$X \overset{(loi)}{=} V^{\lambda/S} \sum_{k=1}^{S+1} X_k \tag{9.10}$$

*où $V$ est une variable aléatoire de loi Bêta$\left(\frac{1}{S}, 1\right)$ indépendante des $X_1, \dots, X_{S+1}$.*

*Soient $V, \xi_1, \dots, \xi_{S+1}$ des variables aléatoires indépendantes telles que $\xi_1, \dots, \xi_{S+1}$ sont de loi Gamma$\left(\frac{1}{S}\right)$ et $V$ est de loi Bêta$\left(\frac{1}{S}, 1\right)$. On pose*

$$\xi = V \sum_{i=1}^{S+1} \xi_i,$$

*et pour tout $k \in \{1, \dots, S+1\}$,*

$$V_k = \frac{\xi_k}{\sum_{i=1}^{S+1} \xi_i}.$$

*Dès lors,*

*(i) la variable aléatoire $\xi$ est de loi Gamma$\left(\frac{1}{S}\right)$,*

*(ii) le vecteur aléatoire $(V_1, \dots, V_{S+1})$ est indépendant de $\xi$ et de loi Dirichlet$\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$,*

*(iii) si $X, X_1, \dots, X_{S+1}$ satisfont l'Équation (9.9), si $X_1, \dots, X_{S+1}$ sont indépendants de $V, \xi_1, \dots, \xi_{S+1}$, et si $X$ est indépendant de $\xi$, alors $S^\nu \xi^{\lambda/S} X$, $S^\nu \xi_1^{\lambda/S} X_1$, …, $S^\nu \xi_{S+1}^{\lambda/S} X_{S+1}$ satisfont l'Équation (9.10).*

## 9.3 Unicité des solutions

L'objectif de cette partie est de montrer que, dans un espace approprié, les solutions respectives des systèmes (9.7) et (9.8) sont uniques. Dans le cas continu, ce résultat est déjà montré dans [Jan04] : nous ne détaillerons donc pas de preuve pour le cas continu, mais seulement pour le cas discret.

Nous nous plaçons dans l'espace $\mathcal{M}_2$ des mesures de probabilités sur $\mathbb{C}$ de carré intégrable et pour tout complexe $A$, nous notons $\mathcal{M}_2^{\mathbb{C}}(A)$ le sous-ensemble des mesures de $\mathcal{M}_2$ de moyenne $A$. Rappelons que la distance de Wasserstein est définie pour tout couple de mesures $\mu, \nu$ de $\mathcal{M}_2^{\mathbb{C}}(A)$ par

$$d_W(\mu, \nu) = \min_{X \sim \mu, Y \sim \nu} \|X - Y\|_2,$$

où $\|\cdot\|_2$ est la norme $L^2$ sur $\mathbb{C}$.

Pour tout $A_1, \dots, A_d \in \mathbb{C}$, nous définissons la distance de Wasserstein sur l'espace produit $\times_{i=1}^d \mathcal{M}_2^{\mathbb{C}}(A_i)$ comme suit : pour tout $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ et $\boldsymbol{\nu} = (\nu_1, \dots, \nu_d)$ deux éléments de $\times_{i=1}^d \mathcal{M}_2^{\mathbb{C}}(A_i)$,

$$d(\boldsymbol{\mu}, \boldsymbol{\nu}) = \max_{1 \leqslant i \leqslant d} \{d_W(\mu_i, \nu_i)\}.$$

Nous savons que $(\mathcal{M}_2^{\mathbb{C}}(A), d_W)$ et donc $\times_{i=1}^d \mathcal{M}_2^{\mathbb{C}}(A_i)$ sont des espaces métriques complets (cf. [Dud02]).

Concentrons-nous sur le modèle d'urne en temps discret. Le vecteur aléatoire $(W_{\boldsymbol{e_1}}^{DT}, \dots, W_{\boldsymbol{e_d}}^{DT})$ est solution du Système (9.7) :

$$W_{\boldsymbol{e_c}} \overset{(loi)}{=} \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} V_p^{\lambda/S} W_{\boldsymbol{e_i}}^{(p)}.$$

Pour tout $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d) \in \times_{i=1}^d \mathcal{M}_2^{\mathbb{C}}(m_i)$, pour tout $c \in \{1, \dots, d\}$, posons

$$K_c(\boldsymbol{\mu}) = \mathcal{L}\left( \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} V_p^{\lambda/S} x_i^{(p)} \right),$$

où $\gamma_0 = 0$, $\gamma_i = \sum_{j \leqslant i} (a_{c,j} + \delta_{c,j})$ et pour tout $i \in \{1, \dots, d\}$, les $(x_i^{(p)})_{1 \leqslant k \leqslant S+1}$ sont des variables aléatoires indépendantes de loi $\mu_i$, indépendantes les unes des autres et du vecteur $V$ qui est de loi de Dirichlet de paramètres $\left(\frac{1}{S}, \dots, \frac{1}{S}\right)$.

On définit l'application $K$ comme

$$K(\boldsymbol{\mu}) = (K_1(\boldsymbol{\mu}), \dots, K_d(\boldsymbol{\mu})).$$

Montrons le théorème suivant via le théorème de point fixe de Banach :

**Théorème 9.3.1**

*Pour toute grande valeur propre $\lambda$ de la matrice de remplacement $R$, pour tout $\boldsymbol{A} = (A_1, \dots, A_d) \in \mathbb{C}^d$ tel que $\boldsymbol{A} \in \texttt{Ker}(R - \lambda I_d)$, l'application $\Phi$ est une contraction de $\times_{i=1}^d \mathcal{M}_2^{\mathbb{C}}(A_i)$ dans lui-même. Dès lors, la loi de $(W_{\boldsymbol{e_1}}^{DT}, \dots, W_{\boldsymbol{e_d}}^{DT})$ est l'unique solution de (9.7) à moyenne fixée.*

**Démonstration :** Remarquons tout d'abord que

$$\mathbb{E}K_c(\boldsymbol{\mu}) = \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} \mathbb{E}V_p^{\lambda/S} \mathbb{E}x_i^{(k)}$$

car $(V_1, \dots, V_{S+1})$ est indépendant de $(x_i^{(1)}, \dots, x_i^{(S+1)})_{1 \leqslant i \leqslant d}$. Comme pour tout $p \in \{1, \dots, S+1\}$, pour tout $i \in \{1, \dots, d\}$, $\mathbb{E}x_i^{(p)} = A_i$, nous avons

$$\begin{aligned}
\mathbb{E}K_c(\boldsymbol{\mu}) &= \sum_{i=1}^{d} A_i \sum_{p=\gamma_{i-1}+1}^{\gamma_i} \mathbb{E}V_p^{\lambda/S} \\
&= \frac{\sum_{i=1}^{d} A_i(\gamma_i - \gamma_{i-1})}{1+\lambda} \\
&= \frac{\sum_{i=1}^{d} A_i(a_{c,i} + \delta_{c,i})}{1+\lambda}
\end{aligned}$$

car $V_p^{\lambda/S} \overset{(loi)}{=} U^\lambda$ pour tout $p \in \{1,\dots,S+1\}$, où $U$ est une variable aléatoire uniforme sur $[0,1]$, et car $\mathrm{Re}\lambda > 1/2$, ce qui nous assure que $\lambda \neq -1$. Comme $\lambda$ est valeur propre de la matrice de remplacement $R$, le système suivant est vérifié par tout $\boldsymbol{A} = (A_1,\dots,A_d)$ tel que $\boldsymbol{A} \in \mathtt{Ker}(R - \lambda I_d)$ :

$$\sum_{i=1}^{d} a_{c,i} A_i = \lambda A_c$$

pour tout $1 \leqslant c \leqslant d$, ce qui implique

$$\mathbb{E}K_c(\boldsymbol{\mu}) = A_c$$

pour tout $\boldsymbol{\mu} \in \times_{i=1}^{d} \mathcal{M}_2^{\mathbb{C}}(A_i)$. De plus $K(\boldsymbol{\mu})$ est de carré intégrable, ce qui nous assure que $K$ est bien une application de $\times_{i=1}^{d} \mathcal{M}_2^{\mathbb{C}}(A_i)$ dans lui-même, pour tout $\boldsymbol{A} \in \mathtt{Ker}(R - \lambda I_d)$.

Montrons désormais que $K$ est une contraction pour la distance de Wasserstein. Nous réutilisons les méthodes utilisées dans le Chapitre 8, Section 8.3, l'idée clef étant d'utiliser la loi de la variance totale. Soit $(x_1^{(p)},\dots,x_d^{(p)})_{1\leqslant p\leqslant S+1}$ une suite de variables aléatoires indépendantes de lois marginales communes $\boldsymbol{\mu} = (\mu_1,\dots,\mu_d) \in \times_{i=1}^{d} \mathcal{M}_2^{\mathbb{C}}(A_i)$, et $(y_1^{(p)},\dots,y_d^{(p)})_{1\leqslant p\leqslant S+1}$ une suite de variables indépendantes de lois marginales communes $\boldsymbol{\nu} = (\nu_1,\dots,\nu_d) \in \times_{i=1}^{d} \mathcal{M}_2^{\mathbb{C}}(A_i)$, et soit $(V_1,\dots,V_{S+1})$ un vecteur aléatoire de loi de Dirichlet de paramètres $\left(\frac{1}{S},\dots,\frac{1}{S}\right)$. Nous avons, pour tout $c \in \{1,\dots,d\}$,

$$\begin{aligned}
d_W(K_c(\boldsymbol{\mu}),K_c(\boldsymbol{\nu}))^2 &\leqslant \left\| \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} V_p^{\lambda/S}(x_i^{(p)} - y_i^{(p)}) \right\|_2^2 \\
&= \mathbb{E}\left| \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} V_p^{\lambda/S}(x_i^{(p)} - y_i^{(p)}) \right|^2 \\
&= \mathbb{E}\left[ \mathbb{E}\left( \left| \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} V_p^{\lambda/S}(x_i^{(p)} - y_i^{(p)}) \right|^2 \ | (V_1,\dots,V_{S+1}) \right) \right].
\end{aligned}$$

Comme $(V_1,\dots,V_{S+1})$ est indépendant de $(x_1^{(p)},\dots,x_d^{(p)})_{1\leqslant p\leqslant d}$, nous obtenons :

$$\begin{aligned}
d_W(K_c(\boldsymbol{\mu}),K_c(\boldsymbol{\nu}))^2 &\leqslant \sum_{i=1}^{d} \sum_{p=\gamma_{i-1}+1}^{\gamma_i} \mathbb{E}|V_p^{2\lambda/S}| \mathbb{E}|x_i^{(p)} - y_i^{(p)}|^2 \\
&\leqslant \sum_{i=1}^{d} \mathbb{E}|x_i - y_i|^2 \sum_{p=\gamma_{i-1}+1}^{\gamma_i} \mathbb{E}|V_p^{2\lambda/S}| \\
&= \sum_{i=1}^{d} \frac{\gamma_i - \gamma_{i-1}}{2\mathrm{Re}\lambda + 1} \|x_i - y_i\|_2^2,
\end{aligned}$$

et ce pour tout $(x_1,\dots,x_d) \sim \boldsymbol{\mu}$ et $(y_1,\dots,y_d) \sim \boldsymbol{\nu}$. Dès lors,

$$\begin{aligned}
d_W(K_c(\boldsymbol{\mu}),K_c(\boldsymbol{\nu}))^2 &\leqslant \sum_{i=1}^{d} d_W(\mu_i,\nu_i)^2 \frac{a_{c,i} + \delta_{c,i}}{2\mathrm{Re}\lambda + 1} \\
&\leqslant d(\boldsymbol{\mu},\boldsymbol{\nu})^2 \frac{S+1}{2\mathrm{Re}\lambda + 1} \leqslant cst \cdot d(\boldsymbol{\mu},\boldsymbol{\nu})^2
\end{aligned}$$

où $cst = \frac{S+1}{2\mathrm{Re}\lambda+1} < 1$ car $\frac{\mathrm{Re}\lambda}{S} = \sigma > \frac{1}{2}$. Pour conclure,

$$d(K(\boldsymbol{\mu}),K(\boldsymbol{\nu})) = \max_{1\leqslant c\leqslant d} \{d_W(K_c(\boldsymbol{\mu}),K_c(\boldsymbol{\nu}))\} \leqslant \sqrt{cst}\, d(\boldsymbol{\mu},\boldsymbol{\nu}),$$

et l'application $K$ est bien une contraction de $\times_{i=1}^{d} \mathcal{M}_2^{\mathbb{C}}(A_i)$. ■

## 9.4 Moments

Dans cette section, nous nous intéressons aux moments des variables $(W^{DT}_{\boldsymbol{e_i}})_{i\in\{1,\dots,d\}}$ et $(W^{CT}_{\boldsymbol{e_i}})_{i\in\{1,\dots,d\}}$. Via la convergence dans tous les $L^p$ ($p \geqslant 1$) des Théorèmes 9.1.2 et 9.1.3, nous savons que ces variables aléatoires admettent des moments de tous ordres. De plus,

**Théorème 9.4.1**

*(i) Pour toute composition initiale $\boldsymbol{\alpha}$, la variable aléatoire $W^{CT}_{\boldsymbol{\alpha}}$ est déterminée par ses moments.*

*(ii) Pour toute composition initiale $\boldsymbol{\alpha}$, la série de Laplace de $W^{DT}_{\boldsymbol{\alpha}}$ a un rayon de convergence infini.*

Tout comme pour les urnes à deux couleurs, ce théorème se montre via le critère de Carleman grâce au lemme suivant :

**Lemme 9.4.2**

*Si $X_1, \dots, X_d$ sont des solutions du Système (9.8) admettant des moments de tous ordres, alors, les suites $\left(\frac{\mathbb{E}|X_i|^p}{p!\ln^p p}\right)^{\frac{1}{p}}$, pour tout $i \in \{1,\dots,d\}$ sont bornées.*

Une fois ce lemme montré pour les variables $W$ en temps continu, nous pourrons utiliser les équations de décomposition pour l'étendre à toute composition initiale, puis la Proposition 9.2.5 pour conclure quant aux $W^{DT}$.

**Démonstration :** La preuve de ce lemme est très similaire à celle du Lemme 8.4.2. Soit $X_1, \dots, X_d$ solutions du Système (9.8), soit $\varphi(p) := \ln^p(p+2)$ et soit, pour tout $i \in \{1,\dots,d\}$,

$$u_p^{(i)} := \frac{\mathbb{E}|X_i|^p}{p!\varphi(p)}.$$

Montrons par récurrence sur $p \geqslant 1$ que, pour tout $i \in \{1,\dots,d\}$, la suite $\left(\frac{\mathbb{E}|X_i|^p}{p!\varphi(p)}\right)^{\frac{1}{p}}$ est bornée. Élevons les $d$ équations du Système (9.8) à la puissance $p$. Comme $\mathbb{E}|U^{\lambda p}| = \frac{1}{p\mathrm{Re}\lambda+1}$, pour tout $c \in \{1,\dots,d\}$, comme les coefficients $(a_{c,i})_{1\leqslant c,i\leqslant d}$ sont positifs,

$$\mathbb{E}|X_c|^p \leqslant \frac{1}{p\mathrm{Re}\lambda+1}\left(\sum_{i=1}^{d}(a_{c,i}+\delta_{c,i})\mathbb{E}|X_i|^p + \sum_{\substack{p_1+\dots+p_{S+1}=p\\ p_j\leqslant p-1}} \frac{p!}{p_1!\dots p_{S+1}!}\prod_{i=1}^{d}\mathbb{E}|X_i|^{p_{\gamma_{i-1}+1}}\dots\mathbb{E}|X_i|^{p_{\gamma_i}}\right),$$

ou encore,

$$p\mathrm{Re}\lambda\mathbb{E}|X_c|^p \leqslant \sum_{i=1}^{d} a_{c,i}\mathbb{E}|X_i|^p + \sum_{\substack{p_1+\dots+p_{S+1}=p\\ p_j\leqslant p-1}} \frac{p!}{p_1!\dots p_{S+1}!}\prod_{i=1}^{d}\mathbb{E}|X_i|^{p_{\gamma_{i-1}+1}}\dots\mathbb{E}|X_i|^{p_{\gamma_i}},$$

ce qui implique, pour tout $c \in \{1,\dots,d\}$,

$$p\mathrm{Re}\lambda u_p^{(c)} \leqslant \sum_{i=1}^{d} a_{c,i}u_p^{(i)} + \sum_{\substack{p_1+\dots+p_{S+1}=p\\ p_j\leqslant p-1}} \frac{\varphi(p_1)\dots\varphi(p_{S+1})}{\varphi(p)}\prod_{i=1}^{d}u^{(i)}_{p_{\gamma_{i-1}+1}}\dots u^{(i)}_{p_{\gamma_i}} \tag{9.11}$$

Contrairement au cas à deux couleurs, nous ne pouvons prouver que l'inverse de $(p\mathrm{Re}\lambda I_d - R)$ est à coefficients positifs ou nuls pour montrer que les variables $X$ et $Y$ admettent des moments de tous ordres dès lors qu'elles sont intégrables. C'est pourquoi nous avons supposé qu'elles admettent en effet tous leurs moments.

Soit

$$\Phi(p) := \sum_{\substack{p_1+\cdots+p_{S+1}=p \\ p_j \leqslant p-1}} \frac{\varphi(p_1)\dots\varphi(p_{S+1})}{\varphi(p)}. \tag{9.12}$$

Nous savons, au vu du Lemme 8.4.3, que $\Phi(p) \leqslant (1+8\ln(p+2))^{S+1}$, pour tout $p \geqslant 2$.

Notons $\Delta_p$ le déterminant de la matrice $p\mathrm{Re}\lambda I_d - R$ (ce déterminant est non nul pour tout $p \geqslant 2$ car $2\mathrm{Re}\lambda > S$), et $\Delta_p(j,i)$ le déterminant de la matrice $p\mathrm{Re}\lambda I_d - R$ privée de sa colonne $i$ et de sa ligne $j$. Pour tout $1 \leqslant i,j \leqslant d$, le polynôme $\Delta_p(j,i)$ est de degré au plus $d-1$ en $p$. Dès lors,

$$\sup_{1\leqslant i,j\leqslant d} \frac{|\Delta_p(j,i)|}{|\Delta_p|} = \mathcal{O}\left(\frac{1}{p}\right),$$

et il existe une constante $\eta > 0$ et un entier $p_0 \geqslant 1$ tel que, pour tout $p \geqslant p_0$,

$$\sup_{1\leqslant i,j\leqslant d} \frac{|\Delta_p(j,i)|}{|\Delta_p|} \leqslant \frac{\eta}{p}.$$

Pour toute matrice carrée $M$ de dimension $d$ à coefficients réels, on notera $\|M\|_\infty = \sup_{1\leqslant i,j\leqslant d}|M_{i,j}|$ et $|||M||| = \sup_{\|x\|_\infty=1}\|Mx\|_\infty$ (la norme $\|x\|_\infty$ d'un vecteur $x$ est le maximum des modules de ses coordonnées) respectivement la norme sup et la norme d'opérateur de la matrice $M$. Comme nous travaillons en dimension finie, ces deux normes sont équivalentes et il existe une constante $\kappa > 0$ telle que, pour tout matrice $M$ réelle de dimension $d$,

$$|||M||| \leqslant \kappa\|M\|_\infty.$$

Par ailleurs, notons $\Delta_p(c)$ le déterminant de la matrice obtenue en remplaçant la $c^{\text{ième}}$ colonne de $p\mathrm{Re}\lambda I_d - R$ par une colonne de 1. Nous savons que $\Delta_p$ est de degré $d$ en $p$ alors que $\Delta_p(c)$ est un polynôme de degré au plus $d-1$ en $p$. Dès lors, il existe un entier $p_1 \geqslant p_0$ tel que, pour tout $p \geqslant p_1$, pour tout $c \in \{1,\dots,d\}$,

$$\frac{\Delta_p(c)}{\Delta_p}(1+8\ln(p+2))^{S+1} \leqslant \frac{1}{2\kappa^2\eta\mathrm{Re}\lambda}.$$

Enfin, comme $\frac{\|p\mathrm{Re}\lambda I_d - R\|_\infty}{p\mathrm{Re}\lambda} \to 1$ quand $p$ tend vers $+\infty$, il existe $p_2 \geqslant p_1$ tel que, pour tout $p \geqslant p_2$,

$$\frac{\|p\mathrm{Re}\lambda I_d - R\|_\infty}{p\mathrm{Re}\lambda} \leqslant 2.$$

Posons maintenant

$$A := \max\{(u_q^{(i)})^{\frac{1}{q}}, 1 \leqslant q \leqslant p_2, 1 \leqslant i \leqslant d\}.$$

Montrons par récurrence sur $p \geqslant p_2$ que, pour tout $q \leqslant p$ pour tout $c \in \{1,\dots,d\}$, $(u_q^{(c)})^{\frac{1}{q}} \leqslant A$. Supposons que cette affirmation soit vraie pour $p \geqslant p_2$. L'Équation (9.11) implique

$$p\mathrm{Re}\lambda\, u_p^{(c)} \leqslant \sum_{i=1}^{d} a_{c,i}u_p^{(i)} + A^p\Phi(p).$$

Soit $v_1,\dots,v_d$ la solution du système

$$p\mathrm{Re}\lambda v_c = \sum_{i=1}^{d} a_{c,i}v_i + A^p\Phi(p).$$

En résolvant ce système de Cramer, nous obtenons

$$v_c = A^p \Phi(p) \frac{\Delta_p(c)}{\Delta_p}.$$

Pour tout $p \geqslant p_2$,

$$(p\mathrm{Re}\lambda I_d - R)\boldsymbol{u}^{(p)} \leqslant (p\mathrm{Re}\lambda I_d - R)\boldsymbol{v} \leqslant |||p\mathrm{Re}\lambda I_d - R|||A^p \frac{1}{2\kappa^2\eta\mathrm{Re}\lambda}\mathbb{1},$$

où $\boldsymbol{u}^{(p)}$ et $\boldsymbol{v}$ sont les vecteurs de coordonnées respectives $(u_p^{(i)})_{1\leqslant i\leqslant d}$ et $(v_i)_{1\leqslant i\leqslant d}$, où $\mathbb{1}$ est le vecteur dont toutes les coordonnées sont égales à 1, et où le signe $\leqslant$ entre vecteurs se lit coordonnée par coordonnée. Dès lors, si on considère la norme $\|.\|_\infty$ sur $\mathbb{R}^d$,

$$\|(p\mathrm{Re}\lambda I_d - R)\boldsymbol{u}^{(p)}\|_\infty \leqslant A^p \frac{|||p\mathrm{Re}\lambda I_d - R|||}{2\kappa^2\eta\mathrm{Re}\lambda} \leqslant A^p \frac{\|p\mathrm{Re}\lambda I_d - R\|_\infty}{2\kappa\eta\mathrm{Re}\lambda} \leqslant A^p \frac{p}{\kappa\eta}.$$

Notons $M = p\mathrm{Re}\lambda I_d - R$. Les coefficients de $M^{-1}$ sont exprimables comme suit :

$$(M^{-1})_{i,j} = (-1)^{i+j} \frac{\Delta_p(j,i)}{\Delta_p},$$

où $\Delta_p$ est le déterminant de $M$, et $\Delta_p(j,i)$ le déterminant de la matrice $M$ privée de la ligne $j$ et de la colonne $i$. Par définition de $p_2$, pour tout $p \geqslant p_2$,

$$\|M^{-1}\|_\infty = \sup_{1\leqslant i,j\leqslant d} |(M^{-1})_{i,j}| \leqslant \frac{\eta}{p},$$

ce qui implique, pour tout $p \geqslant p_2$,

$$\|\boldsymbol{u}^{(p)}\|_\infty \leqslant |||M^{-1}|||A^p \frac{p}{\kappa\eta} \leqslant \kappa\|M^{-1}\|_\infty A^p \frac{p}{\kappa\eta} \leqslant A^p.$$

Dès lors, pour tout $c \in \{1,\dots,d\}$,

$$u_c^{(p)} \leqslant A^p,$$

ce qui conclut le raisonnement par récurrence. ■

**Démonstration du Théorème 9.4.1 :** *(i)* Le lemme 9.4.2 permet de montrer que, pour tout $i \in \{1,\dots,d\}$, la variable aléatoire $W^{CT}_{\boldsymbol{e_i}}$, qui admet des moments de tous ordres au vu du Théorème 9.1.3 vérifie le critère de Carleman. Tout comme pour les urnes à deux couleurs, nous pouvons désormais utiliser le Théorème 9.2.2 pour généraliser le Lemme 9.4.2 à toute composition initiale. Pour toute composition initiale $\boldsymbol{\alpha}$, $W^{CT}_{\boldsymbol{\alpha}}$ vérifie le critère de Carleman.

*(ii)* Nous avons l'inégalité suivante : il existe une constante $C$ telle que, pour tout entier $p$,

$$\frac{\mathbb{E}|W^{CT}|^p}{p!} \leqslant C^p \ln^p p.$$

Dès lors, via l'Équation (9.3), il existe une constante $D$ telle que, pour tout entier $p$,

$$\frac{\mathbb{E}|W^{DT}|^p}{p!} \leqslant D^p \frac{\ln^p p}{\Gamma\left(p\mathrm{Re}\lambda + \frac{1}{S}\right)},$$

ce qui implique que la série de Laplace de $W^{DT}$ est de rayon de convergence infini. Le résultat se généralise à toute composition initiale $\boldsymbol{\alpha}$ via le Théorème 9.2.1. ■

Il est intéressant de noter que la démonstration du Théorème 9.4.1 nous donne une borne supérieure pour les moments des variables aléatoires $W^{CT}$ et $W^{DT}$. Nous n'avons par contre aucune information concernant une borne inférieure.

## 9.5 Densité

Dans cette dernière section, nous présentons une ébauche de preuve de l'existence d'une densité pour les variables $W^{DT}$ et $W^{CT}$. Grâce à l'Équation (9.3), nous savons que si $W^{DT}$ admet une densité, alors $W^{CT}$ en admet une elle aussi. Il paraît donc raisonnable de se consacrer à l'étude du Système (9.7), même si celui-ci semble plus délicat que le système en temps continu.

Le problème de l'existence de la densité des variables aléatoires $(W^{DT}_{\boldsymbol{e_i}})$ est un problème ouvert. Une stratégie possible est d'adapter la preuve développée dans le Chapitre 8, Section 8.5. Sans détailler les preuves, nous allons détailler un plan de preuve envisageable et pointer les difficultés à surmonter pour généraliser la preuve du Théorème 8.5.2.

Pour tout $c \in \{1, \dots, d\}$, pour tout $z \in \mathbb{C}$, on note $\varphi_c(z) = \mathbb{E}\mathbf{e}^{i\mathrm{Re}(W_{\boldsymbol{e_i}}\bar{z})}$. La preuve de l'existence de la densité peut se faire en cinq étapes, tout comme dans le cas des urnes à deux couleurs.

- **Étape 1 :** pour tout $c \in \{1, \dots, d\}$, il existe $\varepsilon > 0$ tel que

$$D(0, \varepsilon) \subseteq \mathtt{Supp}(W_{\boldsymbol{e_c}}).$$

- **Étape 2 :** pour tout $c \in \{1, \dots d\}$, pour tout $z \neq 0$, $|\varphi_c(z)| < 1$. Cette seconde étape se montre en utilisant le résultat de lÉtape 1 par la même preuve que le Lemme 8.5.4.
- **Étape 3 :** pour tout $c \in \{1, \dots, d\}$, $\lim_{|z|\to+\infty} \varphi_c(z) = 0$. Cette étape de montre via des arguments similaires à ceux utilisés dans la preuve du Lemme 8.5.5 en utilisant l'hypothèse d'irréductibilité de l'urne.
- **Étape 4 :** pour tout $c \in \{1, \dots, d\}$, pour tout $\rho \in \left]0, \frac{1}{\mathrm{Re}\lambda}\right[$, $|\varphi_c(z)| = \mathcal{O}(|z|^{-\rho})$ quand $z$ tend vers $+\infty$. Cette étape se montre comme le Lemme 8.5.6.
- **Étape 5 :** pour conclure la preuve, il ne nous resterait plus qu'à adapter la preuve du Théorème 8.5.2. Ce dernier point est difficile car nous pouvons montrer que $|\varphi_c(z)| = \mathcal{O}\left(|z|^{-\rho}\right)$ pour tout $\rho \in \left]0, \frac{a_c c+1}{\mathrm{Re}\lambda}\right[$, mais rien ne nous assure que $\frac{a_c c+1}{\mathrm{Re}\lambda} > 1$. Il nous faut donc certainement être un peu plus fin dans notre analyse.

Les points difficiles à surmonter pour prouver l'existence de la densité de $W^{DT}_{\boldsymbol{e_i}}$ (pour tout $i \in \{1, \dots, d\}$) via ce plan de preuve sont l'Étape 1 et l'Étape 5. Ces travaux sont actuellement en cours.

## 9.6 Conclusion

Nous avons montré comment exploiter la structure arborescente d'une urne de Pólya peut aider à son étude. Nous avons notamment pu montrer que la variable $W^{CT}_{\boldsymbol{\alpha}}$ est déterminée par ses moments et que $W^{DT}_{\boldsymbol{\alpha}}$ a une série de Laplace convergente. Par ailleurs, ces méthodes serviront à démontrer dans des travaux ultérieurs l'existence d'une densité pour ces deux variables.

Il reste encore du chemin avant de comprendre en profondeur ces variables $W^{DT}_{\boldsymbol{\alpha}}$ et $W^{CT}_{\boldsymbol{\alpha}}$ : quel est l'ordre exact de leurs moments, comment se comporte leur transformée de Fourier au voisinage de l'origine ? en l'infini ? leur série de Laplace converge-t-elle ?

Par ailleurs, au delà de l'étude des projections du vecteur composition de l'urne sur les espaces stables associés à des blocs de Jordan dont la valeur propre est grande, comment se comportent, en temps discret, les projections sur les espaces stables associés à de petites valeurs propres ? Il paraît naturel de conjecturer un comportement gaussien (puisque c'est la cas en temps continu), mais il serait intéressant d'établir un tel résultat afin de rendre l'étude asymptotique des urnes à $d$ couleurs exhaustive.

# Urnes de Pólya : conclusion

Nous avons montré dans ce chapitre comment exploiter la structure arborescente sous-jacente des urnes de Pólya pour mieux les étudier. Nous nous sommes concentrés sur les grandes urnes de Pólya à deux couleurs et sur les grandes valeurs propres d'urnes de Pólya à $d$ couleurs. Le résultat principal de cette partie est d'avoir montré que la variable aléatoire $W^{CT}$ est déterminée par ses moments et la variable $W^{DT}$ a une série de Laplace convergente, aussi bien pour les grandes urnes à deux couleurs que pour les grandes valeurs propres d'urnes à $d$ couleurs. Cette approche nous a aussi permis, en dimension deux, de montrer l'existence de la densité de $W^{DT}$ et $W^{CT}$ (résultat déjà connu pour cette dernière). Nous conjecturons que des arguments semblables nous permettront de montrer l'existence de cette densité aussi en dimension $d$.

Il est intéressant de noter que la variable $W^{DT}$ semble plus régulière que $W^{CT}$ : sa série de Laplace est convergente alors que celle de $W^{CT}$ a un rayon de convergence nul, et, pour les urnes à deux couleurs, la transformée de Fourier de $W^{DT}$, contrairement à celle de $W^{CT}$, est intégrable et sa densité est continue bornée, alors que celle de $W^{CT}$ est infini en zéro.

Beaucoup de problèmes restent ouverts, aussi bien en dimension deux qu'en dimension $d$. Quel est l'ordre exact des moments de ces variables ? Quel est le comportement précis de leur transformée de Fourier au voisinage de zéro ou à l'infini ? Quels est le comportement de leurs queues de distribution ?

En temps discret, le comportement général des urnes à $d$ couleurs est encore assez mystérieux : nous nous sommes appliqués ici à étudier le comportement des projections du vecteur composition sur des espaces stables associés à des grandes valeurs propres. Mais que peut-on dire de la projection de ce vecteur composition sur un espace stable associé à un bloc de Jordan d'une petite valeur propre ? Comme signalé auparavant, il est démontré [Jan04] que ce comportement est gaussien, mais seulement dans le cas où $S$ est valeur propre simple de $R$ et où les autres valeurs propres de $R$ sont toutes petites, et seulement concernant les projections sur les espace stables associé à la valeur propre de plus grande partie réelle. Peut-on établir un résultat plus général ? Il semble en effet raisonnable de conjecturer un comportement gaussien le long de tout espace stable associé à une petite valeur propre, et les récents travaux de Knape et Neininger [KN13] laissent penser qu'une approche par systèmes de point fixe permettra de montrer ce résultat.

Par ailleurs, lorsque la valeur propre $S$ n'est pas simple, comment se comporte la projection du vecteur composition sur un espace stable associé à un bloc de Jordan de $S$ ? La connaissance des urnes de Pólya-Eggenberger nous pousse à conjecturer une convergence vers un vecteur aléatoire de loi de Dirichlet : est-ce vrai ?

Les réponses à toutes ces questions permettront de mieux cerner le comportement des urnes de Pólya équilibrées et irréductibles.

# Bibliographie


[AB05] D. Aldous et A. Bandyopadhyay : A survey of max-type recursive distributional equations. *Annals of Applied Probability*, 15(2):1047–1110, 2005.

[ABM12] G. Alsmeyer, J.D. Biggins et M. Meiners : The functional equation of the smoothing transform. *Annals of Probability*, 40(5):2069–2105, 2012.

[AK68] K. Athreya et S. Karlin : Embedding of urn schemes into continuous time markov branching process and related limit theorems. *Annals of Mathematical Statistics*, 39: 1801–1817, 1968.

[AN72] K. B. Athreya et P. E. Ney : *Branching Processes*. Springer-Verlag, 1972.

[Ath69] K.B. Athreya : On a characteristic property of Pólya's urn. *Studia Scientiarum Mathematicarum Hungarica*, 4:31–36, 1969.

[Ben74] E. A. Bender : Asymptotic methods in enumeration. *SIAM Review*, 16(4):485–515, 1974.

[Ber06] J. Bertoin : *Random fragmentation and coagulation processes*. Cambridge University Press, 2006.

[BFS92] François Bergeron, Philippe Flajolet et Bruno Salvy : *Varieties of increasing trees*. Springer, 1992.

[BK64] David Blackwell et David Kendall : The Martin boundary for Pólya's urn scheme, and an application to stochastic population growth. *Journal of Applied Probability*, 1(2):284–296, 1964.

[BK05] J. D. Biggins et A. Kyprianou : Fixed points of the smoothing transform : The boundary case. *Electronic Journal of Probability*, 10:609–631, 2005.

[BM10] Jin X. Barral, J. et B. Mandelbrot : Convergence of complex multiplicative cascades. *Annals of Applied Probability*, 20(4):1219–1252, 2010.

[BP85] A. Bagchi et A. Pal : Asymptotic normality in the generalized Pólya-Eggenberger urn model with applications to computer data structures. *SIAM Journal on Algebraic and Discrete Methods*, 6:394–405, 1985.

[CDM93] R. Casas, J. Díaz et C. Martínez : Average-case analysis on simple families of trees using a balanced probability model. *Theoret. Comput. Sci.*, 117:99–112, 1993.

[CFGG04] B. Chauvin, Ph. Flajolet, D. Gardy et B. Gittenberger : And/Or trees revisited. *Combinatorics, Probability and Computing*, 13(4-5):475–497, Juillet-Septembre 2004.

[CLP12a] B. Chauvin, Q. Liu et N. Pouyanne : Limit distributions for multitype branching processes of m-ary search trees. *À paraître aux Annales de l'Institut Henri Poincaré*, 2012. `http://arxiv.org/abs/1112.0256`.

[CLP12b] B. Chauvin, Q. Liu et N. Pouyanne : Support and density of the limit m-ary search trees distribution. *Discrete Mathematics and Theoretical Computer Science*, AQ:191–200, 2012.

[CLR89] T. H. Cormen, C. E. Leiserson et R. L. Rivest : *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, 1989.

[Com74] L. Comtet : *Advanced Combinatorics : The Art of Finite and Infinite Expansions*. Reidel, 1974.

[CP04] B. Chauvin et N. Pouyanne : m-ary search trees when m>26 : a strong asymptotics for the space requirements. *Random Structures & Algorithms*, 24(2):133–154, 2004.

[CPS11] B. Chauvin, N. Pouyanne et R. Sahnoun : Limit distributions for large Pólya urns. *The annals of Applied Probability*, 21(1):1–32, 2011.

[Dev98] L. Devroye : Branching processes and their applications in the analysis of tree structures and tree algorithms. pages 249–314, 1998.

[Drm97] M. Drmota : Systems of functional equations. *Random Structures & Algorithms*, 10(1-2):103–124, 1997.

[Drm09] M. Drmota : *Random Trees*. Springer Verlag, 2009.

[Dud02] R.M. Dudley : *Real Analysis and Probability*. Cambridge University Press, 2002.

[EP23] F. Eggenberger et G. Pólya : Über die statistik verketetter vorgäge. *Zeitschrift für Angewandte Mathematik und Mechanik*, 1:279–289, 1923.

[FDP06] Ph. Flajolet, Ph. Dumas et V. Puyhaubert : Some exactly solvable models of urn process theory. *In Proceedings of the 4th Colloquium on Mathematics and Computer Science*, volume AG, pages 59–118. DMTCS Proceedings, 2006.

[FGG09] H. Fournier, D. Gardy et A. Genitrini : Balanced And/Or trees and linear threshold functions. *In 5th SIAM Workshop on Analytic and Combinatorics (ANALCO)*, pages 51–57, 2009.

[FGGG08] H. Fournier, D. Gardy, A. Genitrini et B. Gittenberger : Complexity and limiting ratio of Boolean functions over implication. *In Proceedings of the 33rd International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 347–362, Torun, Pologne, Août 2008. LNCS 5162.

[FGGG12] H. Fournier, D. Gardy, A. Genitrini et B. Gittenberger : The fraction of large random trees representing a given Boolean function in implicational logic. *Random Structures & Algorithms*, 20(7):875–887, 2012.

[FGGZ07] H. Fournier, D. Gardy, A. Genitrini et M. Zaionc : Classical and intuitionistic logics are asymptotically identical. *In Proceedings of the 16th Annual Conference on Computer Science Logic (EACSL)*, pages 177–193. LNCS 4646, 2007.

[FGGZ10] H. Fournier, D. Gardy, A. Genitrini et M. Zaionc : Simple tautologies over implication with negative literal. *Journal Mathematical Logic Quarterly*, 56(4):388–396, 2010.

[FGP05] Ph. Flajolet, J. Gabarró et H. Pekari : Analytic urns. *Annals of Probability*, 33(3):1200–1233, 2005.

[FK05] J. Fill et N. Kapur : Transfer theorems and asymptotic distributional results for m-ary search trees. *Random Structures & Algorithms*, 26(4):359–391, 2005.

[FO90] Ph. Flajolet et A. M. Odlyzko : Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216–240, 1990.

[For05] D. J. Ford : Probabilities on cladograms : introduction to the alpha model. 2005. http://arxiv.org/abs/math/0511246.

[Fre65] D. Freedman : Bernard Friedman's urn. *Annals of Mathematical Statistics*, 36:956–970, 1965.

[Fri49] B. Friedman : A simple urn model. *Communications on Pure and Applied Mathematics*, 2:59–70, 1949.

[FS09] Ph. Flajolet et R. Sedgewick : *Analytic Combinatorics*. Cambridge University Press, 2009.

[Gar02] D. Gardy : Occupancy urn models in analysis of algorithms. *Journal of Statistical Planning and Inference, special issue on the Fourth International Conference on Lattice Paths Combinatorics and Applications*, 101(1-2):95–105, February 2002.

[Gar06] D. Gardy : Random Boolean expressions. *In Proc. Colloquium on Computational Logic and Applications*, volume AF, pages 1–36. DMTCS Proceedings, 2006.

[Gen09] A. Genitrini : *Expressions booléennes aléatoires : probabilité, complexité et comparaison quantitative des logiques propositionnelles*. Thèse de doctorat, Université de Versailles St-Quentin-en-Yvelines, 2009. http://www-apr.lip6.fr/~genitrini/publi/these_genitrini.pdf.

[GG10] A. Genitrini et B. Gittenberger : No Shannon effect on probability distributions on Boolean fonctions induced by Boolean expressions. *In Proceedings of of Analysis of Algorithms*, volume AM, pages 303–316, Vienne, Autriche, Juillet 2010. DMTCS Proceedings.

[GGKM12] A. Genitrini, B. Gittenberger, V. Kraus et C. Mailler : Probabilities of Boolean functions given by random implicational formulas. *Electronic Journal of Combinatorics*, 19(2):P37, 20 pages, 2012.

[GGKM13] A. Genitrini, B. Gittenberger, V. Kraus et C. Mailler : Associative and commutative tree representations for Boolean functions. *Soumis à Random Structures & Algorithms*, 2013. http://arxiv.org/abs/1305.0651.

[GK12] A. Genitrini et J. Kozik : In the full propositional logic, 5/8 of classical tautologies are intuitionistically valid. *Annals of Pure and Applied Logic*, 163(7):875–887, 2012.

[GKP94] R.L. Graham, D.E. Knuth et O. Patashnik : *Concrete Mathematics : a Foundation for Computer Science*, volume 2. Addison-Wesley Reading, MA, 1994.

[GKZ07] A. Genitrini, J. Kozik et M. Zaionc : Intuitionistic vs. classical tautologies, quantitative comparison. *In* Springer-Verlag, éditeur : *TYPES*, pages 100–109, Cividale del Friuli, Italie, Mai 2007.

[Gou93] R. Gouet : Martingale functional central limit theorems for a generalized Pólya urn. *Annals of Probability*, 21:1624–1639, 1993.

[Gou97] R. Gouet : Strong convergence of proportions in a multicolor Pólya urn. *Journal of Applied Probability*, 34:426–435, 1997.

[Har49] G. H. Hardy : *Divergent Series*. Oxford University Press, 1949.

[HMPW08] B. Haas, G. Miermont, J. Pitman et M. Winkel : Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability*, 36(5):1790–1837, 2008.

[Jan04] S. Janson : Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications*, 110:177–245, 2004.

[Jan06] S. Janson : Limit theorems for triangular urn schemes. *Probability Theory and Related Fields*, 134(3):417–452, 2006.

[JK97] N. L. Johnson et S. Kotz : *Urn Models and their Application*. Wiley and Sons, 1997.

[Juk12] S. Jukna : *Boolean function complexity : advances and frontiers*. Springer Verlag, 2012.

[KN13] M. Knape et R. Neininger : Pólya urns via the contraction method. 2013. `http://arxiv.org/abs/1301.3404`.

[Koz08] J. Kozik : Subcritical pattern languages for And/Or trees. *In Fifth Colloquium on Mathematics and Computer Science*, numéro 1, Blaubeuren, Allemagne, Septembre 2008. DMTCS Proceedings.

[KP76] J-P. Kahane et J. Peyrière : Sur certaines martingales de Benoît Mandelbrot. *Advances in Mathematics*, 22:131–145, 1976.

[Kra11] V. Kraus : *Asymptotic study of unlabelled trees and other unlabelled graph structures*. Thèse de doctorat, Université Technique de Vienne, Autriche, 2011. `http://krausver.bplaced.net/diss.pdf`.

[Lal93] S. P. Lalley : Finite range random walk on free groups and homogeneous trees. *The Annals of Probability*, 21(4):2087–2130, 1993.

[Liu99] Q. Liu : Asymptotic properties of supercritical age-dependent branching processes and homogeneous branching random walks. *Stochastic Processes and their Applications*, 82(1):61–87, 1999.

[Liu01] Q. Liu : Asymptotic properties and absolute continuity of laws stable by random weight mean. *Stochastic Processes and their Applications*, 95:83–107, 2001.

[LS97] H. Lefmann et P. Savický : Some typical properties of large And/Or Boolean formulas. *Random Structures & Algorithms*, 10:337–351, 1997.

[Mah98] H. Mahmoud : On rotations in fringe-balanced binary trees. *Information Processing Letters*, 65:41–46, 1998.

[Mah08] H. Mahmoud : *Pólya urn models*. CRC Press, 2008.

[Man74] B. Mandelbrot : Multiplications aléatoires itérées et distributions invariantes par moyenne pondérée aléatoire. *Comptes Rendus de l'Académie des Sciences de Paris, Série A*, 278:289–292, 1974.

[Mor13] B. Morcrette : *Combinatoire analytique et modèles d'urnes*. Thèse de doctorat, UPMC, Paris, 2013. `http://www-apr.lip6.fr/~morcrette/these.pdf`.

[Ngu04] M. Nguyen The : *Distributions de valuations sur les arbres*. Thèse de doctorat, École Polytechnique, Palaiseau, 2004.

[NR06] R. Neininger et L. Rüschendorf : A survey of multivariate aspects of the contraction method. *Discrete Mathematics and Theoretical Computer Science*, 8:31–56, 2006.

[Pit84] B. Pittel : On growing random binary trees. *Journal of Mathematical Analysis and Applications*, 103(2):461–480, 1984.

[Pou08] N. Pouyanne : An algebraic approach to Pólya processes. *Annales de l'institut Henri Poincaré*, 44(2):293–323, 2008.

[PR87] G. Pólya et R. C. Read : *Combinatorial enumeration of Groups, Graphs and Chemical Compounds*. Springer Verlag, New York, 1987.

[PVW94] J. B. Paris, A. Vencovská et G. M. Wilmers : A natural prior probability distribution derived from the propositional calculus. *Annals of Pure and Applied Logic*, 70:243–285, 1994.

[Rös92] U. Rösler : A fixed point theorem for distributions. *Stochastic Processes and their Applications*, 42:195–214, 1992.

[RR01] U. Rösler et L. Rüschendorf : The contraction method for recursive algorithms. *Algorithmica*, 29(1-2), 2001.

[RS42] J. Riordan et C. E. Shannon : The number of two terminal series-parallel networks. *Journal of Mathematical Physics*, 21:83–93, 1942.

[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 System Technical*, 28:59–98, 1949.

[Sib88] M. Sibuya : Log-concavity of Stirling numbers and unimodality of Stirling distributions. *Annals of the Institute of Statistical Mathematics*, 40(4):693–714, 1988.

[Smy96] R. Smythe : Central limit theorems for urn models. *Stochastic Processes and Their Applications*, 65:115–137, 1996.

[Weg87] I. Wegener : *The Complexity of Boolean Functions*. John Wiley & Sons, 1987.

[Weg05] I. Wegener : *Complexity Theory : Exploring the Limits of Efficient Algorithms*. Springer Verlag, 2005.

[Woo97] A. R. Woods : Coloring rules for finite trees, and probabilities of monadic second order sentences. *Random Structures & Algorithms*, 10(4):453–485, 1997.

**Résumé**

Cette thèse étudie deux objets aléatoires discrets : les arbres booléens aléatoires et les urnes de Pólya. Ces deux objets, tous deux en lien avec l'informatique fondamentale, sont étudiés dans ce mémoire via des méthodes de combinatoire analytique et de probabilités.

Les arbres booléens sont des arbres étiquetés de façon à représenter des expressions booléennes. Chaque arbre booléen représente donc une fonction booléenne. Dans la première partie de cette thèse, nous définissons et comparons plusieurs distributions de probabilité sur l'ensemble des fonctions booléennes via leur représentation par des arbres booléens. Il s'avère que toutes ces distributions chargent préférentiellement les fonctions booléennes de *faible complexité*, et que certaines d'entre elles sont *dégénérées* au sens où elles ne chargent qu'un petit nombre de fonctions booléennes. L'étude de ces modèles se fait principalement par des outils de combinatoire analytique, mais nous utilisons aussi des méthodes probabilistes, comme le plongement en temps continu, ou *poissonisation*, pour certaines de ces distributions.

Une urne de Pólya est un processus discret aléatoire qui modélise, en particulier, de nombreux objets issus de l'informatique comme les arbres $m$-aires de recherche, les arbres $2-3$, les AVL, entre autres. Nous étudions dans la seconde partie de ce mémoire des urnes de Pólya équilibrées, irréductibles et à coefficients positifs. Le comportement asymptotique d'une urne, ainsi que celui de son plongement en temps continu, font intervenir des variables aléatoires $W$ assez méconnues à ce jour. Nous étudions ces variables aléatoires $W$ via la structure arborescente de l'urne et montrons qu'elles sont solutions de systèmes d'équations en loi, ce qui nous permet notamment d'établir que ces variables aléatoires sont déterminées par leurs moments, et surtout d'aborder cette étude aussi bien pour des urnes à deux couleurs que pour des urnes à $d$ couleurs.

**Mots-clefs :** arbres aléatoires, fonctions booléennes, urnes de Pólya, combinatoire analytique, plongement en temps continu, équations de point fixe en loi, détermination par les moments.

**Abstract**

This thesis deals with two random objects : random Boolean trees and Pólya urns. These two objects, which are both related to theoretical computer science, are studied in this thesis by analytic combinatorics and probabilistic methods.

Boolean trees are labelled trees which represent Boolean expressions. Each Boolean tree represents a Boolean function. In the first part of the thesis, we define and compare different probability distributions on the set of Boolean functions via their tree representation. We show that all these distribution weight mostly low *complexity* functions and that some of them are even *degenerated*, meaning that they only weight very few boolean functions. We study these models mainly by analytic combinatorics but we also use probabilistic methods such as embedding in continuous time to study some of these distributions.

A Pólya urn is a discrete random process, that can be used to modelize numerous objects of theoretical computer science such as, $m$-ary search trees, $2-3$ trees, AVL, among others. In the second part of this thesis, we study balanced, irreducible Pólya urns with nonnegative coefficients. Random variables $W$ arise from the asymptotic behaviour of such an urn, and of its embedding in continuous time ; these random variables are quite unknown up to now. We study them in discrete and continuous time via the underlying tree-structure of the urn, and prove that they are solutions of systems of equations in law. This new approach allows us to prove that the variables $W$ are moment-determined, and, above all, to study the variables $W$ both for two-coloured and $d$-coloured Pólya urns.

**Key-words :** random trees, Boolean functions, Pólya urns, analytic combinatorics, embedding in continuous time smoothing equations, moment-determined laws.