La logique classique et la logique intuitionniste sont asymptotiquement identiques

Hervé Fournier [a]
Danièle Gardy [a]
Antoine Genitrini [a]
Marek Zaionc [b]

30 juin 2007

Résumé : Cet article considère les formules logiques construites sur le seul connecteur de l'implication et un nombre fini de variables. Quand le nombre de variables devient grand, on prouve les résultats quantitatifs suivants : *asymptotiquement, toutes les tautologies classiques sont des tautologies simples.* Il en découle qu'*asymptotiquement, toutes les tautologies classiques sont intuitionnistes.*

1. Introduction

On calcule la proportion de formules de taille n qui sont des tautologies parmi toutes les formules de taille n pour les formules propositionnelles construites sur l'implication et contenant k variables. Notre intérêt réside dans le fait de prouver l'existence et de calculer la limite du rapport quand n tend vers l'infini. Cette limite peut être appelée la densité de vérité pour la logique avec k variables. Après avoir isolé le cas particulier des formules appelées des tautologies simples, de densité $1/k + O(1/k^2)$, on exhibe quelques familles de non-tautologies dont la densité cumulée est $1 - 1/k + O(1/k^2)$. Il en découle que la proportion de tautologies, pour k grand, est très proche de la borne inférieure déterminée pour les tautologies simples. Une conséquence de cela est que la logique classique et la logique intuitionniste sont proches l'une de l'autre lorsque le nombre de variables propositionnelles est grand. Ce travail s'inscrit dans un domaine de recherche dans lequel la vraisemblance de la vérité est estimée pour la logique propositionnelle avec un nombre restreint de variables. Nous renvoyons à Gardy [4] pour un survol de la distribution de probabilité sur les fonctions booléennes induite par les expressions booléennes aléatoires. Pour la logique purement implicationnelle d'une variable, et en même temps pour les systèmes de type simple, la valeur exacte de la densité de vérité a été calculée dans un article de Moczurad, Tyszkiewicz et Zaionc [9]. La logique classique en une variable et avec les deux connecteurs que sont l'implication et la négation a été étudiée dans Zaionc [12]. Sur le même langage, la proportion exacte entre les logiques classique et intuitionniste a été déterminée dans Kostrzycka et Zaionc [6]. Quelques variantes dans lesquelles interviennent des formules avec d'autres connecteurs logiques ont aussi été considérées. Le cas des connecteurs et/ou a reçu beaucoup d'attention - voir Lefmann et Savický [7], Chauvin, Flajolet, Gardy et Gittenberger [1] et Gardy et Woods [5]. Matecki [8] a considéré le cas du connecteur d'équivalence.

On donne ensuite un couple de définitions. La section 2 présente brièvement l'utilisation de l'énumération via les fonctions génératrices et la combinatoire analytique, qui constitue l'outil prin-

[a] PRISM, CNRS UMR 8144, Université de Versailles Saint-Quentin en Yvelines, 45 av. des États-Unis, 78035 Versailles cedex, France.
Email: [herve. fournier, daniele.gardy, antoine.genitrini] @prism.uvsq.fr.
[b] Informatique théorique, Université Jagiellonian, Gronostajowa 3, 30-387 Cracovie, Pologne.
Email: zaionc@tcs.uj.edu.pl.
Traduction, Denise Vella-Chemla, décembre 2022 de l'article
https://webusers.imj-prg.fr/ herve.fournier/publications/tautologies-csl.pdf.

cipal que nous utiliserons. Les différentes classes de formules que nous considérons sont décrites dans la section 3, alors que la section 4 est consacrée à l'énumération de ces classes et au calcul de leur densité.

Définition 1 : *Soit* $\{x_1, x_2, \ldots, x_k\}$ *un enemble de variables propositionnelles booléennes. On définit* $\mathcal{F}_k$ *comme étant l'ensemble des expressions booléennes (ou formules) sur ces variables et le connecteur d'implication par la grammaire suivante :* $F := x_1 \mid \ldots \mid x_k \mid (F \to F)$.

De façon évidente, les expressions peuvent être représentées par des arbres planaires binaires, convenablement étiquetés : leurs nœuds internes sont étiquetés par le connecteur $\to$ et leurs feuilles par certaines variables booléennes. Par $\|\phi\|$, on désigne la *taille* de l'expression ϕ qu'on définit comme le nombre total d'occurrences des variables propositionnelles dans l'expression (ou les feuilles dans la représentation arborescente de l'expression). Les parenthèses qui sont quelquefois nécessaires et le signe d'implication lui-même ne sont pas inclus dans la taille de l'expression. Formellement,

$$\|x_i\| = 1 \text{ et } \|\psi \to \psi\| = \|\phi\| + \|\psi\|.$$

On dénote par $\mathcal{F}_k^n$ l'ensemble des expressions de $\mathcal{F}_k$ de taille n.

On peut maintenant définir la *forme canonique d'une expression.* Soit T une expression. Elle peut être décomposée selon sa branche droite - voir la figure 1. Par conséquent, elle est de la forme

$$A_1 \to (A_2(... \to (A_p \to r(T))...)) ;$$

on l'écrira

$$T = A_1, ..., A_p \to r(T).$$

Les formules A_i sont appelées les *prémisses* de T et $r(T)$, la feuille de l'arbre la plus à droite est appelée le *but* de T. Bien sûr, l'expression $T = A_1 \to (A_2 \to (... \to (A_p \to r(T))...))$ est logiquement équivalente à $\overline{A_1} \vee \overline{A_2} \vee ... \vee \overline{A_p} \vee r(T)$, où $\overline{A_i}$ représente la négation de A_i.

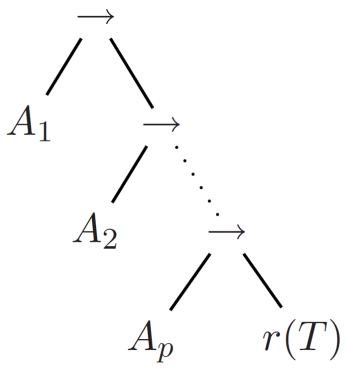

Figure 1: La décomposition canonique d'un arbre.

Pour un sous-ensemble $A \subseteq \mathcal{F}_k$, on définit la *densité* $\mu(A)$ comme :

$$\mu(\mathcal{A}) = \lim_{n\to\infty} \frac{|\{t \in \mathcal{F}_k \ : \ \|t\| = n\}|}{|\{t \in \mathcal{A} \ : \ \|t\| = n\}|}$$

si la limite existe. Le nombre $\mu(\mathcal{A})$ s'il existe est une probabilité asymptotique (selon la distribution uniforme) de trouver une formule de la classe $\mathcal{A}$ parmi toutes les formules de $\mathcal{F}_k$; cela peut être interprété comme la densité asymptotique de l'ensemble $\mathcal{A}$ dans l'ensemble $\mathcal{F}_k$. On peut immédiatement voir que la densité est finiment additive de telle façon que si $\mathcal{A}$ et $\mathcal{B}$ sont des classes disjointes de formules telles que $\mu(\mathcal{A})$ et $\mu(\mathcal{B})$ existe alors $\mu(\mathcal{A} \cup \mathcal{B})$ existent également et $\mu(\mathcal{A} \cup \mathcal{B}) = \mu(\mathcal{A}) + \mu(\mathcal{B})$.

2. Fonctions génératrices.

Dans cet article on recherche la proportion entre le nombre de formules de taille n qui sont des tautologies et le nombre de toutes les formules de taille n pour les formules propositionnelles du langage $\mathcal{F}_k$. Notre intérêt consiste à trouver la limite de cette fraction quand n tend vers l'infini. Dans ce but, l'analyse combinatoire a développé un outil extrêmement puissant, sous la forme des séries génératrices et des fonctions génératrices. On trouve un bel exposé de la méthode dans Wilf [11], ou dans Flajolet, Sedgewick [2, 3] ; voir également Gardy [4, 5.2] pour une application systématique de ces techniques aux densités pour les fonctions booléennes. Comme le lecteur peut s'y attendre, alors qu'on travaille en logique propositionnelle, on utilisera souvent l'analyse complexe, les fonctions analytiques et leurs singularités.

Soit $A = (A_0, A_1, A_2, ...)$ une séquence de nombres réels. La *série génératrice ordinaire* pour A est la série de puissances formelles $\sum_{n=0}^{\infty} A_n z^n$. Et, bien sûr, les séries de puissances formelles sont en bijection avec les séquences. Pourtant, en considérant z comme une variable complexe, cette série, comme on le sait en théorie des fonctions analytiques, converge uniformément vers une fonction $f_A(z)$ dans un certain disque ouvert $\{z \in \mathcal{C} : |z| < R\}$ de diamètre maximal, et $R \geqslant 0$ est appelé son rayon de convergence. Donc à la séquence A, on peut associer une fonction complexe $f_A(z)$, appelée la *fonction génératrice ordinaire* pour A, définie dans un voisinage de 0. Cette correspondance est bijective à nouveau (à moins que l'on ait $R = 0$), puisque, comme on le sait bien en théorie des fonctions analytiques, l'expansion d'une fonction complexe $f(z)$, analytique au voisinage de z_0, en une série de puissances $\sum_{n=0}^{\infty} A_n (z - z_0)^n$ est unique. Pour F une fonction de z analytique dans un voisinage de 0, on dénotera par $[z^n]F$ le coefficient de z^n dans l'expansion en série de F.

Beaucoup de questions concernant le comportement asymptotique de A peuvent être efficacement résolues en analysant le comportement de f_A sur le cercle complexe $|z| = R$. C'est l'approche que nous prenons pour déterminer la proportion de tautologies et de nombreuses autres classes de formules parmi toutes les formules d'une taille donnée.

Tout ensemble d'expressions est défini récursivement à partir d'ensembles plus simples ; on construit les fonctions génératrices qui énumèrent les éléments de ces ensembles par taille (nombre de feuilles), en utilisant des fonctions à une variable avec la variable z marquant les feuilles, et on obtient une fonction génératrice $\psi(z)$ pour l'ensemble considéré. On extrait alors le coefficient $[z^n]\psi(z)$ et on obtient la densité de l'ensemble étudié comme $\lim_{n\to\infty}[z^n]\psi(z)/[z^n]f_k(z)$, $f_k(z)$ étant la fonction génératrice pour l'ensemble de toutes les expressions de $\mathcal{F}_k$.

On rappelle maintenant trois constructions sur les classes d'objets combinatoires, et comment elles se traduisent en fonctions génératrices ordinaires. Soit A et B deux classes d'objets combinatoires,

avec les fonctions génératrices $f_A(z)$ et $f_B(z)$. La première construction, appelée *somme combinatoire*, capture l'union d'ensembles disjoints. La fonction génératrice de la somme combinatoire de A et B est $f_A(z) + f_B(z)$. La seconde construction appelée *produit cartésien* forme toutes les paires ordonnées possibles d'objets à partir de A et B - la taille de (a, b) étant la somme des tailles de a et b. La fonction génératrice énumérant cette classe est $f_A(z)f_B(z)$. Finalement, la *construction de séquences* fabrique toutes les séquences d'objets à partir de A. À nouveau, la taille d'une séquence d'objets est la somme de leurs tailles. La fonction génératrice énumérant cette classe est $1/(1 - f_A(z))$.

Le nombre de Catalan C_n est défini comme le onmbre d'arbres binaires complets (i.e. tout sommet a soit deux sommets fils soit aucun) avec n nœuds internes et $n + 1$ feuilles. Les résultats de base à propos des nombres de Catalan et de leur fonction génératrice sont résumés ci-dessous.

Proposition 2. *Soit $C(z)$ la fonction génératrice énumérant les arbres binaires complets selon leur nombre de feuilles ; elle satisfait :*

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

et est égale à :

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

Ses coefficients sont

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

Il s'ensuit que le nombre d'expressions booléennes de taille n sur k variables est $k^n C_{n-1}$, puisqu'une telle expression est obtenue en étiquetant les n feuilles avec n'importe laquelle des variables $x_1, ..., x_k$.

Comme exemple, dans le reste de cette section, on montre comment on peut obtenir la fonction génératrice $f_k(z)$ pour l'ensemble de toutes les expressions construites sur k variables et avec le connecteur d'implication, avant de définir quelques sous-ensembles d'expressions dans la section 3 et de calculer leurs fonctions génératrices dans la section 4.

Proposition 3. *La fonction génératrice énumérant l'ensemble $\mathcal{F}_k$ de toutes les expressions booléennes sur k variables est*

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

Preuve : En utilisant la forme canonique d'une expression, on sait qu'un arbre est une séquence (possiblement vide) d'arbres, suivie par une feuille - voir la figure 1. La fonction $f_k(z)$ satisfait ainsi

$$f_k(z) = \frac{kz}{1 - f_k(z)}, \text{ i.e. } f_k(z) = kz + f_k(z)^2.$$

Résoudre l'équation et choisir entre les deux possibilités ($f_k(0) = 0$) donne la solution.
La proposition 3 donne une autre manière d'obtenir le nombre d'expressions de taille n en extrayant les coefficients de $f_k(z)$. Dans le reste de l'article, on abrègera f_k par f.

Finalement, les calculs basiques suivants seront utilisés intensivement dans le reste de l'article. Remarquons d'abord que pour tout j.

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

De plus,

$$[z^n]\sqrt{1-4kz}=(4k)^n[z^n]\sqrt{1-z}=-2k^nC_{n-1}.$$

3. Tautologies et non-tautologies

Définissons maintenant quelques classes d'expressions, toutes étant des types particuliers soit de tautologies soit de non-tautologies.

Définition 4. *On définit les sous-ensembles de $\mathcal{F}_k$ suivants :*

- Cl_k est l'ensemble de toutes les *tautologies classiques* i.e. les formules qui sont vraies selon n'importe quelle valuation.
- Int_k est l'ensemble de toutes les *tautologies intuitionistes* i.e. les formules pour lesquelles il y a des λ-termes fermés (des preuves constructives) de type identique à la formule.
- $Pierce_k$ est l'ensemble de toutes les *expressions de Pierce* i.e. des tautologies classiques qui ne sont pas des tautologies intuitionnistes.
- SN_k est l'ensemble des *expressions simples qui ne sont pas des tautologies classiques*, défini par

  $$T=A_1,...,A_p\to r(T),$$

  tel que pour tout i, $r(A_i)\neq r(T)$.
- G_k est l'ensemble des *tautologies simples* i.e. des expressions qui peuvent s'écrire

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

  où il existe i tel que A_i est une variable égale à $r(T)$.
- LN_k est l'ensemble des *expressions moins simples qui ne sont pas des tautologies classiques*, défini par l'ensemble des arbres de la forme

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

  tels que

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

  où $r(C)=r(T), q\geqslant 1$, et

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

  où $r(D)\neq r(T), s\geqslant 0$, et l'assertion suivante est vérifiée : pour tout j, $r(B_j)\notin\{r(T),r(D)\}$ et $r(D_j)\notin\{r(T),r(D)\}$.

Ajouter un exposant n aux ensembles que nous venons juste de définir signifie qu'on considère seulement les expressions de taille exactement égale à n (l'arbre qui représente l'expression a n feuilles).

Notons que les tautologies simples sont des tautologies intuitionnistes puisque l'une de leurs prémisses est égale au but. Les relations évidentes entre les classes ci-dessus sont les suivantes :

$$\begin{array}{rcl} SN_k \cup LN_k & \subset & \mathcal{F}_k \backslash Cl_k \\ SN_k \cap LN_k & = & \varnothing \\ G_k & \subsetneq & Int_k \subsetneq Cl_k \subsetneq \mathcal{F}_k \backslash (SN_k \cup LN_k) \\ Pierce_k & = & Cl_k \backslash Int_k \end{array}$$

Notre objectif dans la suite de cet article est de calculer les densités de ces ensembles. Les résultats sont résumés dans la figure 2 ; les preuves sont données dans la section suivante. Comme conséquence, on obtient le résultat suivant, donnant une réponse positive à la conjecture de [9, page 593].

Théorème 5. *Asymptotiquement (pour k un grand nombre de variables booléennes), toutes les tautologies sont simples i.e.*

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

Preuve : On sait que pour tout k, la densité de la logique classique avec k variables propositionnelles $\mu(Cl_k)$ existe. Un tel résultat est obtenu par des techniques standards d'analyse des algorithmes ; on saute les détails et on renvoie le lecteur intéressé à Flajolet et Sedgewick [3] ou à [4].

Puisque $G_k \subset Cl_k \subset \mathcal{F}_k \backslash (SN_k \cup LN_k)$, et à partir des densités obtenues dans les propositions 7, 8 et 9, on a

$$\frac{4k+1}{(2k+1)^2} = \mu(G_k) \leqslant \mu(Cl_k) \leqslant 1 - \left(\frac{k(k-1)}{(k+1)^2} + \frac{2k(k-1)^2}{(k+2)^4} \right).$$

Les bornes supérieure et inférieure sont asymptotiquement identiques, égales à $1/k + O(1/k^2)$. □

En utilisant exactement le même argument, on peut aussi obtenir un résultat reliant le comportement asymptotique de la logique classique versus celui de la logique intuitionniste.

Corollaire 6. *Asymptotiquement (pour k un grand nombre de variables booléennes), les tautologies classiques sont intuitionnistes i.e.*

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

où $\mu^-(Int_k) = \liminf_{n\to\infty} \frac{|Int_k^n|}{|\mathcal{F}_k^n|}$.

Preuve : À partir du fait que $G_k \subset Int_k \subset Cl_k$, on a

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

Le résultat découle du fait qu'à la fois $\mu(G_k)$ et $\mu(Cl_k)$ sont égaux à $1/k+O(1/k^2)$. □

Ce résultat permet également d'estimer la taille de la différence entre les logiques classique et intuitionniste (appelées les formules de Pierce). Des détails sont donnés en section 4.4.

4. Énumération des classes

On calcule maintenant les densités des trois ensembles SN_k, G_k et LN_k. Le calcul de ces densités est fait de façon systématique. D'abord, chaque ensemble d'expressions est défini récursivement à partir d'ensembles plus simples ; cela permet de construire les fonctions génératrices énumérant les éléments de ces ensembles par leur taille (le nombre de feuilles), et d'obtenir une fonction génératrice pour la classe considérée. Alors on extrait le coefficient $[z^n]\psi(z)$ et on obtient la densité de l'ensemble étudié comme $\lim_{n\to\infty}[z^n]\psi(z)/[z^n]f(z)$ - on rappelle que f dénote la fonction génératrice de toutes les formules.

La dernière partie traite les formules de Pierce. Bien que nous ne sachions pas si cet ensemble de formules a une densité, on donne quelques bornes et on montre que leur ordre de grandeur est $\Theta(1/k^2)$.

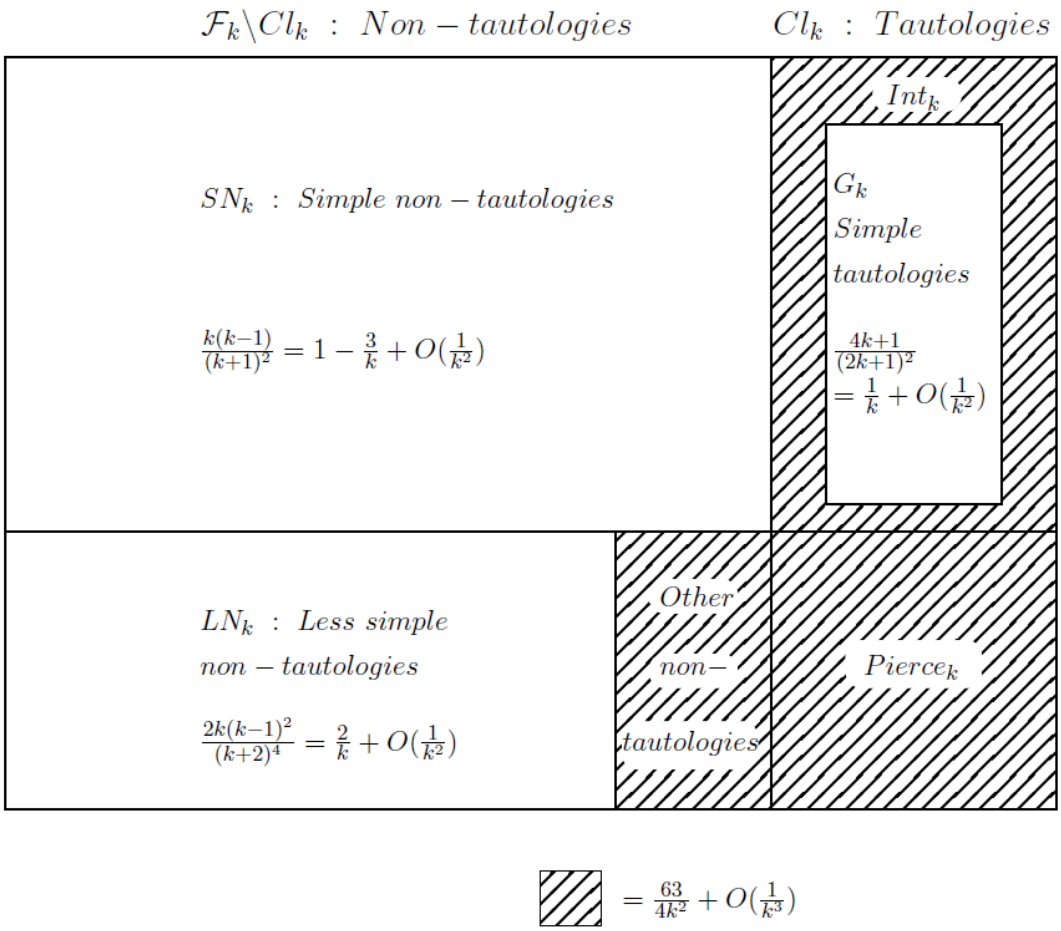

Figure 2: Densité des tautologies simples, des non-tautologies simples et moins simples.

4.1. Non-tautologies simples

On considère d'abord l'ensemble SN_k des expressions simples qui sont des non-tautologies. Si $T \in SN_k$ alors T est du type suivant

$$T = A_1, ..., A_p \to r(T),$$

tel que pour tout i, $r(A_i) \neq r(T)$. On vérifie d'abord que ceci n'est en effet pas une tautologie. Considérons juste la valuation suivante des variables propositionnelles. Définissons $r(T)$ comme valant faux et tous les $r(A_i)$ comme valant vrai. Selon cette valuation, l'expression entière est fausse. Calculons ensuite la fonction génératrice $SN(z)$ associée à SN_k.

Fixons d'abord une variable booléenne α et considérons tous les arbres avec $r(T) = \alpha$. Un tel arbre est une non-tautologie simple si et seulement si toutes ses prémisses A_i satisfont $r(A_i) \neq \alpha$. La fonction génératrice de toutes les prémisses possibles est $\frac{k-1}{k} f(z)$. Comme une non-tautologie simple de but α est une séquence de telles prémisses suivie par la feuille α, la fonction génératrice SN^α des non-tautologies simples de but α est égale à

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

Puisque α peut être choisi arbitrairement parmi les k litéraux, on a $SN(z) = k \cdot SN^\alpha(z)$, ce qui donne

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

Proposition 7. *La densité des non-tautologies simples existe et est égale à*

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

Pour k grand, cette densité est $1 - 3/k + O(1/k^2)$.

Preuve : Ce résultat était déjà donné dans l'article [9, page 586], avec une preuve différente. On donne une preuve alternative ici. Si elle existe, la densité est donnée par la formule suivante :

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

Après modification du dénominateur et du numérateur de la fonction génératrice $SN(z)$, on obtient :

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

Le dénominateur de la fraction rationnelle $SN(z)$ a un unique zéro $\rho = -1/(k-1)^2$. Pourtant cette valeur élimine aussi le numérateur de l'expression puisque

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

Ainsi ρ n'est pas un pôle effectif. Par conséquent la seule singularité qui compte asymptotiquement est $z = 1/4k$. En mettant à part le terme d'erreur, on obtient

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

Cela donne

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

par conséquent la densité de SN_k existe et est égale à $k(k-1)/(k+1)^2$. □

4.2. Tautologies simples

Si T est une tautologie simple, alors T peut s'écrire

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

avec l'un des A_i égal à $r(T)$. Il est trivial de vérifier que T est en effet une tautologie, puisque T est logiquement équivalent à

$$T \sim \overline{A_1} \vee \ldots \vee \overline{r(T)} \vee \ldots \vee \overline{A_p} \vee r(T).$$

qui a de façon évidente la valeur *vrai*.

Calculons maintenant la fonction génératrice des tautologies simples. Un arbre T n'est pas une tautologie simple si et seulement si toutes ses prémisses sont différentes de $r(T)$ - voir la figure 3. La fonction génératrice pour $\mathcal{F}_k \backslash G_k$ est donc égale à $kz/(1-(f(z)-z))$. Il découle de cela que la fonction génératrice de G_k est

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

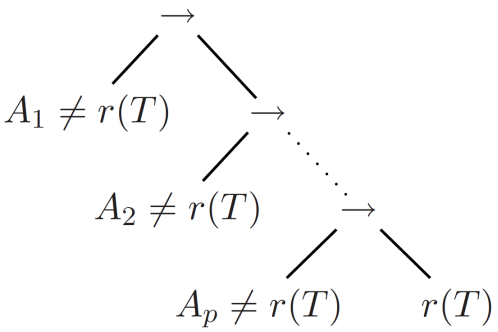

Figure 3: Arbres qui ne sont pas des tautologies simples.

Proposition 8. *La densité limite des tautologies simples sur k variables existe et est égale à*

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

Pour k grand, cette densité est asymptotiquement égale à $1/k - 3/4k^2 + O(1/k^3)$.

Preuve : Une autre preuve antérieure de ce résultat est donnée dans l'article [9, page 584]. Nous en donnons ici une preuve alternative. La fonction génératrice $G(z)$ peut s'écrire comme

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

avec $P(z)$ un polynôme convenable. Soit ρ son pôle ; $\rho = -1-k$. Mais ρ est plus grand que la singularité algébrique $1/(4k)$; par conséquent, $1/(4k)$ est la singularité dominante de $G(z)$.

Finalement, on obtient (au terme d'erreur près)

$$\begin{aligned}[z^n]G(z) &= -\frac{2k}{(2k+1)^2}[z^n]\sqrt{1-4kz} - \frac{2k}{(2k+1)^2}[z^{n-1}]\sqrt{1-4kz} \\ &= \frac{4k}{(2k+1)^2}k^nC_{n-1} + \frac{4}{(2k+1)^2}k^nC_{n-2}.\end{aligned}$$

Démontrons l'existence et calculons la valeur de la densité de G_k^n.

$$\begin{aligned}\mu(G_k) &= \lim_{n\to\infty}\frac{|G_k^n|}{|F_k^n|} = \lim_{n\to\infty}\left(\frac{4k}{(2k+1)^2}k^nC_{n-1} + \frac{4}{(2k+1)^2}k^nC_{n-2}\right)\cdot\frac{1}{k^nC_{n-1}} \\ &= \frac{4k}{(2k+1)^2} + \frac{4}{(2k+1)^2}\cdot\lim_{n\to\infty}\frac{C_{n-2}}{C_{n-1}}.\end{aligned}$$

Donc $\mu(G_k)$ existe effectivement, et est égal à $(4k+1)/(2k+1)^2$. □

4.3. Les non-tautologies moins simples

Dans la famille SN_k des non-tautologies simples, on n'a pas autorisé une prémisse à avoir une feuille la plus à droite égale à $r(T)$. Mais ici on va considérer des arbres avec exactement une telle prémisse.

On rappelle qu'un arbre T définit une non-tautologie moins simple s'il est du type

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

où $C = C_1, ..., C_q \to r(C)$, avec $r(C) = r(T), q \geqslant 1$, et $C_1 = D_1, D_2, \ldots, D_s \to r(D)$ est tel que $r(D) \neq r(T), s \geqslant 0$, et l'assertion suivante est vérifiée : pour tout j, $r(B_j) \notin \{r(T), r(D)\}$ et $r(D_j) \notin \{r(T), r(D)\}$. Voir la figure 4 pour la forme générale de l'arbre et la figure 5 pour le sous-arbre C ; dans ces figures, si un sous-arbre A est souligné, cela signifie qu'il respecte la contrainte $r(A) \notin \{r(T), r(D)\}$.

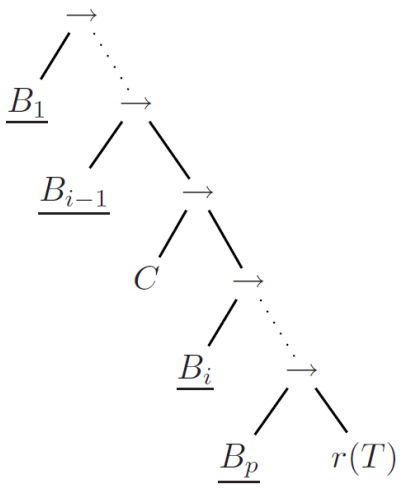

Figure 4: Non-tautologies moins simples.

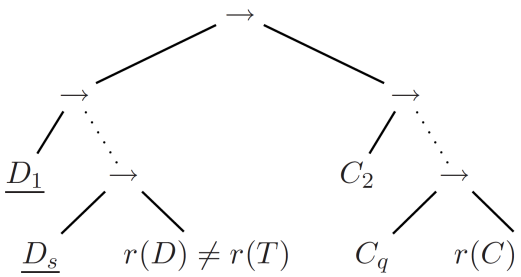

Figure 5: Sous-arbre C d'une non-tautologie moins simple.

Démontrons d'abord qu'un tel arbre n'est pas une tautologie. Pour cela, considérons la valuation selon laquelle toutes les variables sont *vraies*, exceptées $r(T)$ et $r(D)$ qui sont *fausses* ; selon cette valuation, la totalité de l'expression a pour valeur *faux* ; pour vérifier cela, remarquons simplement que la fonction calculée par un tel arbre peut être développée en une conjonction de termes, l'un d'eux étant $\bigwedge_i \overline{r(B_i)} \wedge r(T) \wedge \bigwedge_i \overline{r(D_i)} \wedge r(D)$.

Nous calculerons maintenant la fonction génératrice de LN_k. Fixons α et β, deux variables distinctes. Nous calculerons d'abord $\psi(z)$, les fonctions génératrices de tous les arbres $LN_k^{\alpha,\beta}$ à partir de LN_k tel que $r(T) = \alpha$ et $r(D) = \beta$. Par symétrie, $\psi(z)$ est indépendant du choix de α et β.

Soit $b(z)$ la fonction génératrice de tous les arbres $T \in \mathcal{F}_k$ telle que $r(T) \notin \{\alpha, \beta\}$. Bien sûr, $b(z) = (k-2)/k \cdot f(z)$. Cette fonction génératrice énumère les sous-arbres possibles B_j mais également les sous-arbres possibles D_j. Ainsi, la fonction génératrice de tous les arbres possibles pour D est $d(z) = z/(1 - b(z))$, puisque c'est une séquence d'arbres D_j telle que $r(D_j) \notin \{\alpha, \beta\}$, suivie par la feuille β. De la même manière, la fonction génératrice pour le sous-arbre C est $c(z) = d(z) \cdot 1/(1 - f(z)) \cdot z$.

Notons qu'un arbre de $LN_k^{\alpha,\beta}$ est construit comme une séquence d'arbres B_j avec $r(B_j) \notin \{\alpha, \beta\}$, puis un sous-arbre C comme décrit ci-dessus, puis une autre séquence d'arbres B_j avec $r(B_j) \notin \{\alpha, \beta\}$, suivie par la feuille α. De plus, cette décomposition est unique. La fonction génératrice

pour $LN_k^{\alpha,\beta}$ est ainsi égale à

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

Maintenant, on peut facilement voir que LN_k est l'union disjointe des $LN_k^{\alpha,\beta}$. En effet, étant donné un arbre $T \in LN_k$, alors α est égal à $r(T)$ et la prémisse C de T est déterminée de manière unique parce que c'est la seule prémisse de T avec comme but $r(T)$. Ainsi, β est déterminée de manière unique également puisque c'est le but de la première prémisse de C. Il découle de cela que $\phi(z) = k(k-1)\psi(z)$.

Proposition 9. *La densité des non-tautologies moins simples est égale à*

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

Pour k grand, elle est égale à $2/k + O(1/k^2)$.

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

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

où $P(z)$ est un polynôme convenable. Le dénominateur de la fraction rationnelle $\phi(z)$ a un zéro $\rho = -2/(k-2)^2$. Pourtant cette valeur annule également le numérateur (et les deux premières dérivées) de l'expression, et ce n'est pas un pôle effectif de ϕ. Par conséquent, la seule singularité qui importe asymptotiquement est $z = 1/4k$. En mettant à part le terme d'erreur, on obtient :

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

Démontrons l'existence et calculons la valeur de la densité de LN_k^n :

$$\begin{aligned}
\mu(LN) &= \lim_{n\to\infty} \frac{|LN_k^n|}{|F_k^n|} \\
&= \lim_{n\to\infty} \left(\frac{k^{n+1}(k-1)}{\left(2+\frac{(k-2)^2}{4k}\right)^3} \frac{C_{n-3}}{k^n C_{n-1}} - \frac{k^{n-2}(k-1)(2k^3-6k^2+8)}{\left(2+\frac{(k-2)^2}{4k}\right)^3} \frac{C_{n-4}}{k^n C_{n-1}} \right) \\
&= \frac{64k^4(k-1)}{(k+2)^6} \cdot \lim_{n\to\infty} \frac{C_{n-3}}{C_{n-1}} - \frac{64k(k-1)(2k^3-6k^2+8)}{(k+2)^6} \cdot \lim_{n\to\infty} \frac{C_{n-4}}{C_{n-1}} \\
&= \frac{4k^4(k-1)-k(k-1)(2k^3-6k^2+8)}{(k+2)^6} = \frac{2k(k-1)^2}{(k+2)^4}
\end{aligned}$$

La densité existe effectivement, et est égale à :

$$2k(k-1)^2/((k+2)^4).$$

Pour k grand, la densité est asymptotiquement égale à $2/k + O(1/k^2)$.

4.4 Formules de Pierce

Nous sommes prêts à estimer le nombre de formules de Pierce. Bien que l'on ne sache pas si l'ensemble des formules de Pierce a une densité, on donnera des bornes sur $\limsup_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|}$ et $\liminf_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|}$. Une borne supérieure simple pour $Pierce_k$ peut être obtenue à partir de

$$Pierce_k = Cl_k \backslash Int_k \subset F_k \backslash (SN_k \cup LN_k \cup G_k).$$

Puisque SN_k, LN_k et G_k sont disjoints, on a une estimation supérieure simple basée sur les propositions 7, 8 et 9 :

$$\limsup_{n\to\infty} \frac{|Pierce_k^n|}{|\mathcal{F}_k^n|} \leqslant 1 - \frac{k(k-1)}{(k+1)^2} - \frac{2k(k-1)^2}{(k+2)^4} - \frac{4k+1}{(2k+1)^2} = \frac{63}{4k^2} + O\left(\frac{1}{k^3}\right).$$

Pourtant, on peut obtenir une borne plus précise sur le nombre de formules de Pierce. Pour cela, on bornera ensuite la densité des tautologies qui ne sont pas simples - cette densité existe puisque nous savons déjà qu'à la fois la densité de toutes les tautologies et la densité des tautologies simples existent. Notons que ce résultat fournit une preuve alternative pour le théorème 5.

Lemme 10. *La densité des tautologies non simples T telles qu'exactement une prémisse a un but égal à $r(T)$ est bornée supérieurement par $5/k^2 + O(1/k^3)$.*

Preuve : Soit A une tautologie non simple de but $r(A) = \alpha$. Soit p le nombre de prémisses de A. On appelle B la prémisse de A dont le but est $r(A)$ et $\alpha_1, ..., \alpha_{p-1}$ les buts des prémisses autres

que B. Par hypothèse, $\alpha_i \neq \alpha$ pour tout $i \in \{1, ..., p-1\}$. Bien sûr, B ne peut pas être réduit à une feuille (sinon A serait une tautologie simple). Décomposons $B = (B_1, ..., B_m, \alpha)$, avec $m \geqslant 1$. Comme $\overline{B} = B_1 \wedge ... \wedge B_m \wedge \overline{\alpha}$, en développant l'expression A, on obtient que nécessairement, pour tout $j \in \{1, ..., m\}$,

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

a la valeur *vrai*. Dénotons par $C_{(\alpha_1 ..., \alpha_{p-1}, \alpha)}$ l'ensemble d'arbres tel que

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

a la valeur *vrai*. Soit $C \in C_{(\alpha_1 ..., \alpha_{p-1}, \alpha)}$.

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

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

  doit prendre la valeur *vrai*. Il s'ensuit que $\alpha \in \{\gamma_1, \ldots, \gamma_s\}$ ou $\gamma \in \{\gamma_1, \ldots, \gamma_s, \alpha_1, \ldots, \alpha_{p-1}\}$.

On va maintenant calculer une fonction génératrice $c_{(\alpha_1 ..., \alpha_{p-1}, \alpha)}$ donnant une borne supérieure sur le nombre d'arbres de $C_{(\alpha_1 ..., \alpha_{p-1}, \alpha)}$. Définissons

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

le premier terme correspondant au premier point ci-dessus, le second terme correspondant au cas $\alpha \in \{\gamma_1, ..., \gamma_s\}$ et le troisième au cas $\gamma \in \{\gamma_1, ..., \gamma_s, \alpha_1, \ldots, \alpha_{p-1}\}$. Cette fonction génératrice dépend seulement de p ; ainsi, on va maintenant la dénoter par c_p. Définissons maintenant

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

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

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

prend la valeur *vrai*. Bien sûr

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

On définit maintenant

$$a_p(z) = p \cdot ((k-1)/k \cdot f(z))^{p-1} \cdot \bar{b}_p(z) \cdot z \cdot k.$$

La fonction génératrice a_p donne une borne supérieure sur le nombre de tautologies non simples A avec p prémisses, l'une d'entre elles exactement ayant un but égal à $r(A)$. En effet, z correspond à $r(A) = \alpha$, k correspond au choix de α parmi les litéraux et p correspond à la position de l'unique

prémisse qui a pour but α.

On définit maintenant $a(z) = \sum_{p=1}^{\infty} a_p(z)$. Cette fonction limite le nombre de tautologies non simples A avec seulement une prémisse de but $r(A)$. Le calcul basé sur la fonction génératrice définie ci-dessus amène à une densité asymptotique $5/k^2+O(1/k^3)$. □

Lemme 11. *La densité des tautologies non simples T telles qu'exactement deux prémisses ont un but égal à $r(T)$ est $O(1/k^3)$.*

Preuve : considérons une tautologie non simple A avec exactement deux prémisses B_1 et B_2 ayant un but égal à $r(A)$. Soient $\alpha_1, ..., \alpha_{p-2}$ les buts des autres prémisses. Puisque A n'est pas simple, à la fois B_1 et B_2 ne sont pas réduits à une feuille. Soit C la première prémisse de B_1, et D la première prémisse de B_2. Soit γ le but de C et $\gamma_1, ..., \gamma_s$ les buts de ses prémisses (avec $s \geqslant 0$). On définit $\delta, \delta_1, ..., \delta_t$, les litéraux correspondant pour l'arbre D. Puisque A est une tautologie, on peut argumenter comme dans le lemme précédent et on obtient que nécessairement

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

prend la valeur *vrai*. La même méthode que dans le précédent lemme (non détaillée ici) amène à une densité $O(1/k^3)$. □

Lemme 12. *La densité asymptotique des arbres T tels qu'au moins trois prémisses ont un but égal à $r(T)$ est $O(1/k^3)$.*

Preuve : La fonction génératrice de cette famille d'arbres est égale à

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

On obtient une densité $O(1/k^3)$. □

Proposition 13. *La densité asymptotique des tautologies non simples est bornée supérieurement par $5/k^2 + O(1/k^3)$.*

Preuve : Une tautologie n'est pas réduite à une feuille. De plus, une tautologie T a (au moins) une prémisse avec un but $r(T)$: sinon, ce serait une non-tautologie simple. La densité des tautologies non simples est ainsi bornée supérieurement par la somme des trois densités obtenues dans les lemmes 10, 11 et 12. Par conséquent, elle est bornée supérieurement par $5/k^2+O(1/k^3)$. □

On peut obtenir une borne inférieure pour les formules de Pierce par l'argument suivant. Considérons les formules spéciales de $\mathcal{F}_k$ de la forme $((a \to T) \to a) \to a$ où $T = A_1, ..., A_p \to r(T)$ est une non-tautologie simple prise dans $\mathcal{F}_k$ (voir la section 4.1) et où la variable a diffère de $r(T)$. On observe que $((a \to T) \to a) \to a$ doit être une formule de Pierce. C'est évidemment une tautologie classique. Supposons que $((a \to T) \to a) \to a$ est aussi une tautologie intuitionniste. Cela signifie qu'il doit exister un terme fermé du type $((a \to T) \to a) \to a$. La forme normale longue de ce terme est de la forme $\lambda p_{(a \to T) \to a} \lambda q_a.t$ où t est un terme de type T avec pour seules variables libres

p et q. Considérons un terme fermé $\lambda p_{(a\to T)\to a}\lambda q_a.t$. Le type de ce terme est la formule implicative

$$((a \to T) \to a) \to (a \to T).$$

Mais ce type est à nouveau une non-tautologie simple puisque les variables a et $r(T)$ sont différentes. Ainsi la formule n'est pas prouvable classiquement et par conséquent, de façon intuitionniste non plus ; contradiction. Pour plus de détails à propos de la relation entre la logique intuitionniste et le lambda calcul consulter par exemple Sørensen, Urzyczyn [10].

Maintenant nous devons compter les éléments de cette famille. Le nombre de telles formules est $(k-1)\cdot|SN_k^{n-3}|$. Ainsi la densité de cet ensemble particulier de formules de Pierce existe et est égale à

$$\lim_{n\to\infty}\frac{(k-1)\cdot|SN_k^{n-3}|}{|\mathcal{F}_k^n|} = \lim_{n\to\infty}\frac{(k-1)\cdot|SN_k^{n-3}|}{|\mathcal{F}_k^{n-3}|}\cdot\frac{|\mathcal{F}_k^{n-3}|}{|\mathcal{F}_k^n|} = \frac{1}{64k^2}\,\frac{(k-1)^2}{(k+1)^2}$$

puisque $\lim_{n\to\infty}|\mathcal{F}_k^{n-3}|/|\mathcal{F}_k^n| = 1/(4k)^3$.

Proposition 14. *On a les bornes suivantes sur le nombre de formules de Pierce :*

$$\frac{1}{64k^2} - O\left(\frac{1}{k^3}\right) \leqslant \liminf_{n\to\infty}\frac{|Pierce_k^n|}{|\mathcal{F}_k^n|} \leqslant \limsup_{n\to\infty}\frac{|Pierce_k^n|}{|\mathcal{F}_k^n|} \leqslant \frac{5}{k^2} + O\left(\frac{1}{k^3}\right).$$

Preuve : La borne inférieure vient de la discussion précédente. Puisque les formules de Pierce sont des tautologies non simples, la borne supérieure est une conséquence de la proposition 13. □

5. Dernières remarques

On a montré qu'asymptotiquement, toutes les tautologies avec implication sont simples, i.e. l'une de leurs prémisses est égale à leur but. La méthode développée dans cet article étend la logique de l'implication à des litéraux à la fois positifs et négatifs. Dans ce nouveau paradigme, à nouveau, on peut démontrer que la plupart des tautologies, quand le nombre des variables devient grand, exhibent une structure très simple ; plus précisément, la plupart des tautologies ont l'une de leurs prémisses égale au but (comme précédemment), ou ont deux de leurs prémisses qui sont des litéraux opposés.

Quelques questions subsistent à propos de l'ensemble des formules de Pierce. On conjecture que pour tout k, les densités $\mu(Int_k)$ et $\mu(Pierce_k)$ existent. Si c'était le cas, il serait intéressant d'évaluer les densités asymptotiques de ces ensembles.

Références

[1] B. Chauvin, P. Flajolet, D. Gardy, B. Gittenberger. And/Or trees revisited, *Combinatorics, Probability and Computing*, 13(4-5):475-497, 2004.

[2] P. Flajolet, R. Sedgewick. *Analytic combinatorics: functional equations, rational and algebraic functions*, INRIA, Number 4103, 2001.

[3] P. Flajolet, R. Sedgewick. *Analytic combinatorics.* Livre en préparation, consultable ici : http://algo.inria.fr/flajolet/Publications/books.html, 2007.

[4] D. Gardy. Random Boolean expressions, *Colloquium on Computational Logic and Applications*, Chambéry (France), Juin 2005. Proceedings in DMTCS, pp 1-36, 2006.

[5] D. Gardy, A. Woods. And/or tree probabilities of Boolean function, *Discrete Mathematics and Theoretical Computer Science*, pp 139-146, 2005.

[6] Z. Kostrzycka, M. Zaionc. Statistics of intuitionnistic versus classical logic, *Studia Logica*, 76(3):307-328, 2004.

[7] H. Lefmann, P. Savický. Some typical properties of large And/Or Boolean formulas, *Random Structures and Algorithms*, vol 10, pp 337-351, 1997.

[8] G. Matecki. Asymptotic density for equivalence, *Electronic Notes in Theoretical Computer Science*, 140:81-91, 2005.

[9] M. Moczurad, J. Tyszkiewicz, M. Zaionc. Statistical properties of simple types, *Mathematical Structures in Computer Science*, 10(5):575-594, 2000.

[10] M. Sørensen, P. Urzyczyn. *Lectures on the Curry-Howard Isomorphism*, volume 149 of *Studies in Logic and the Foundations of Mathematics*. Elsevier Science, 2006.

[11] H. Wilf. *Generatingfunctionology*, seconde édition. Academic Press, Boston, 1994.

[12] M. Zaionc. On the asymptotic density of tautologies in logic of implication and negation, *Reports on Mathematical Logic*, vol 39, pp 67-87, 2005.

[13] M. Zaionc. Probability distribution for simple tautologies, *Theoretical Computer Science*, 355(2):243-260, 2006.