

Expressions booléennes aléatoires

Danièle Gardy

PRiSM, Univ. de Versailles Saint-Quentin

Mars 2005

Plan

- Fonctions et formules Booléennes
- Les diverses représentations et leurs complexités
 - Circuits
 - Formules Booléennes et arbres
 - Diagrammes de décision binaires
- Quelques classes de fonctions Booléennes
- Dénombrements de classes d'équivalence
- Dénombrements d'arbres et lois de probabilité induites
 - La probabilité de *Vrai*
 - Le théorème de Drmota-Lalley-Woods
 - Quelques résultats complémentaires
- Complexité et probabilité

Fonctions et formules booléennes

Fonction Booléenne $f: \{0, 1\}^n \rightarrow \{0, 1\}$

(ou $\{Vrai, Faux\}^n \rightarrow \{Vrai, Faux\}$)

Ex: $n = 2$, fonction définie par

$$(0, 0) \mapsto 0$$

$$(0, 1) \mapsto 0$$

$$(1, 0) \mapsto 0$$

$$(1, 1) \mapsto 1$$

$$f(x_1, x_2) = x_1 \wedge x_2$$

Variables Booléennes : x_1, \dots, x_n

n : nombre de variables

Littéraux: les x_i ou leurs négations \bar{x}_i

Ensemble \mathcal{B}_n des fonctions Booléennes sur n variables

$$\text{Card}(\mathcal{B}_n) = 2^{2^n}$$

Décomposition de Shannon:

$$f(x_1, x_2, \dots, x_n) = x_1 f(1, x_2, \dots, x_n) + (1 - x_1) f(0, x_2, \dots, x_n)$$

Pour la coordonnée x_i :

$$f = x_i f|_{x_i=1} + (1 - x_i) f|_{x_i=0}.$$

Expression Booléenne :

sur les littéraux et des opérateurs: \neg , \wedge , \vee , \rightarrow , \leftrightarrow , \oplus , ...

$$x_1 \rightarrow (x_2 \wedge x_3)$$

$$\bar{x}_1 \vee (x_2 \wedge x_3)$$

$$(\bar{x}_1 \wedge (x_1 \vee \bar{x}_1)) \vee ((x_2 \wedge x_3) \wedge x_2)$$

$$x_1 \leftrightarrow (x_2 \wedge x_3)$$

$$x_1 \oplus x_2 \oplus x_3$$

On peut se limiter aux opérateurs \neg , \wedge et \vee .

Associativité des opérateurs :

$x_1 \vee (x_2 \vee x_3)$ et $x_1 \vee x_2 \vee x_3$ sont équivalentes

Commutativité des opérateurs :

$x_1 \vee x_2$ et $x_2 \vee x_1$ sont équivalentes

Forme normale conjonctive

$$\bigwedge_j (\bigvee_i t_{i,j})$$

Forme normale disjonctive

$$\bigvee_j (\bigwedge_i t_{i,j})$$

Satisfiabilité d'une expression:

Il existe une assignation des variables Booléennes qui donne la valeur *Vrai*. Elle est donc associée à une fonction autre que *Faux*.

Une expression est une **tautologie** si elle prend la valeur *Vrai* pour toute assignation des variables. Elle est donc associée à la fonction constante *Vrai*.

Quelques classes de fonctions Booléennes

- Fonction monotone: $x \leq y \Rightarrow f(x) \leq f(y)$

Formule monotone: sans négation, opérateurs \wedge et \vee seuls

Exemples: $x_1 \wedge x_2$, $x_1 \vee x_2$, $(x_1 \vee x_2) \wedge x_3 \wedge x_2$

- Formule réduite: arbre où les opérateurs \wedge et \vee alternent à chaque niveau, et sont d'arité non bornée.
- Formule rigide: comme les formules réduites, mais les fils de chaque sommet sont tous distincts.
- Fonction “read-once” : représentable par une formule dans laquelle chaque variable apparaît au plus une fois.

Une fonction “read-once” est représentée par une unique formule rigide.

Contre-exemple: $x_1 \oplus x_2$

Complexité d'une fonction Booléenne f :
taille minimale $L(f)$ d'une représentation de f

Représentations de f :

- expressions Booléennes /arbres
- circuits
- diagrammes de décision binaires (binary decision diagrams, ou BDD)

La mesure de complexité dépend de la représentation choisie.

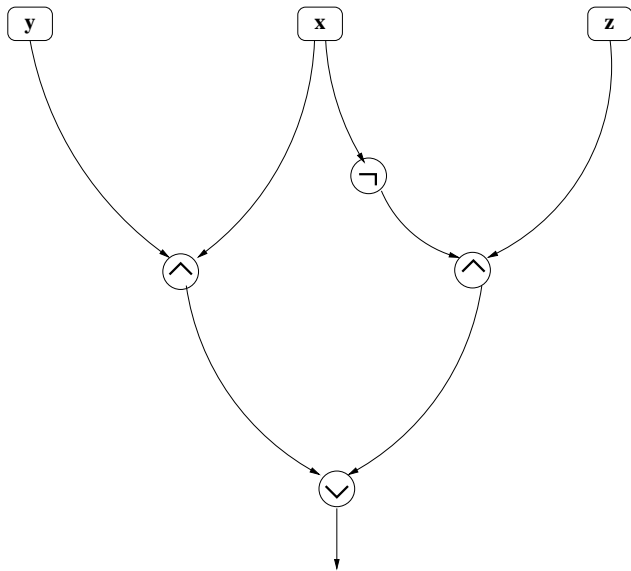
Circuits

Un circuit est un graphe orienté acyclique:

- noeuds sans arc entrant: *entrées*, étiquetées par 0, 1 ou un littéral;
 - noeuds internes, avec $k > 0$ arcs entrants: étiquetés par un opérateur Booléen d'arité k (en général, opérateur \wedge , \vee et \neg);
 - un noeud est la *sortie*
-
- *Taille* du circuit: c'est le nombre de ses noeuds internes
 - *Profondeur* : distance maximale d'une entrée à la sortie

Si le graphe sous-jacent au circuit est un arbre: on retrouve les formules Booléennes

$$(x \wedge y) \vee (\bar{x} \wedge z)$$



Circuit de taille 4 et de profondeur 3

Résultats sur la complexité $C(f)$ de circuits.

- Borne supérieure simple:

$$C(f) \leq 2^{n+1} - 3$$

- Loi de probabilité uniforme sur \mathcal{B}_n . Alors, p.s. pour $n \rightarrow +\infty$ (Shannon)

$$C(f) > \frac{2^n}{n}$$

- Pour toutes les fonctions $f \in \mathcal{B}_n$ (Lupanov)

$$C(f) \leq \frac{2^n}{n} (1 + o(1))$$

Effet de Shannon: La complexité de la plupart des fonctions est de l'ordre de la complexité maximale.

Bornes inférieures?

Fonction *seuil* (vaut 1 ssi au moins k des n variables valent 1).

Pour $k = 2$, la taille minimale d'un circuit (avec \wedge et \vee binaires) la représentant est $\geq 2n - 4$

Représentations arborescentes

Les règles de construction des formules définissent différents types d'arbres

- Opérateurs non associatifs \Rightarrow arbres binaires
- Opérateurs associatifs \Rightarrow arbres d'arité quelconque
- Opérateurs non commutatifs \Rightarrow arbres planaires
- Opérateurs commutatifs \Rightarrow arbres non planaires
- Etiquetage des noeuds internes: par les opérateurs acceptés
- Etiquetage des feuilles: par les variables Booléennes ou leurs négations
- La négation est souvent repoussée aux feuilles

Opérateurs binaires \wedge et \vee , et négation sur les variables \Rightarrow *arbres et/ou*: arbres de Catalan, noeuds internes étiquetés par \wedge et \vee , feuilles étiquetées par les littéraux.

Opérateurs binaires \wedge et \vee , et négation non restreinte aux variables \Rightarrow arbres unaires-binaires, noeuds internes étiquetés par \wedge et \vee , ou par \neg , feuilles étiquetées par les variables.

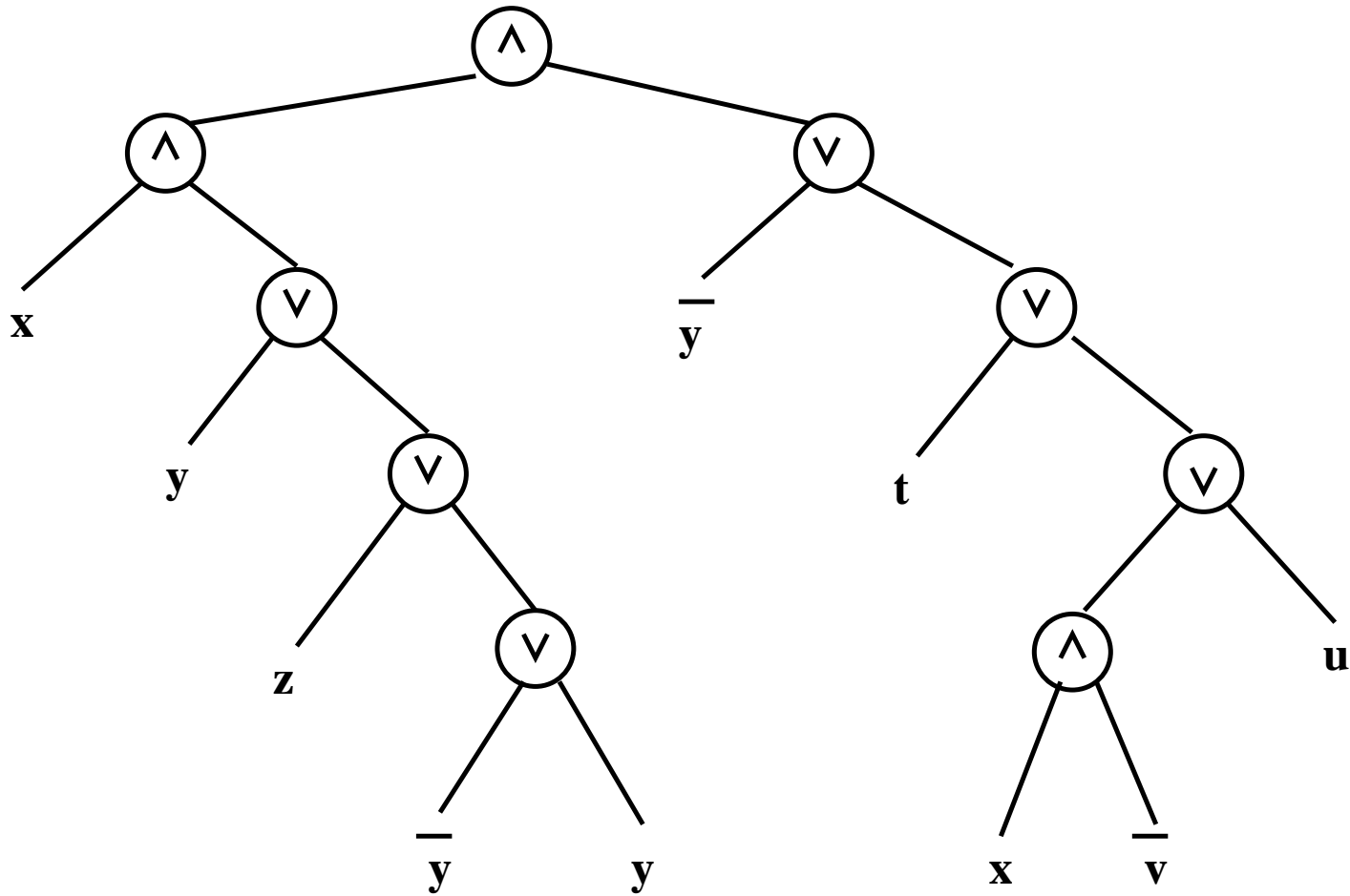
Opérateurs \wedge et \vee d'arité variable et commutatifs, alternant par niveaux, et négation restreinte aux variables \Rightarrow arbres généraux non planaires; il suffit de choisir l'étiquette de la racine et celles des feuilles (littéraux).

Fonction Booléenne

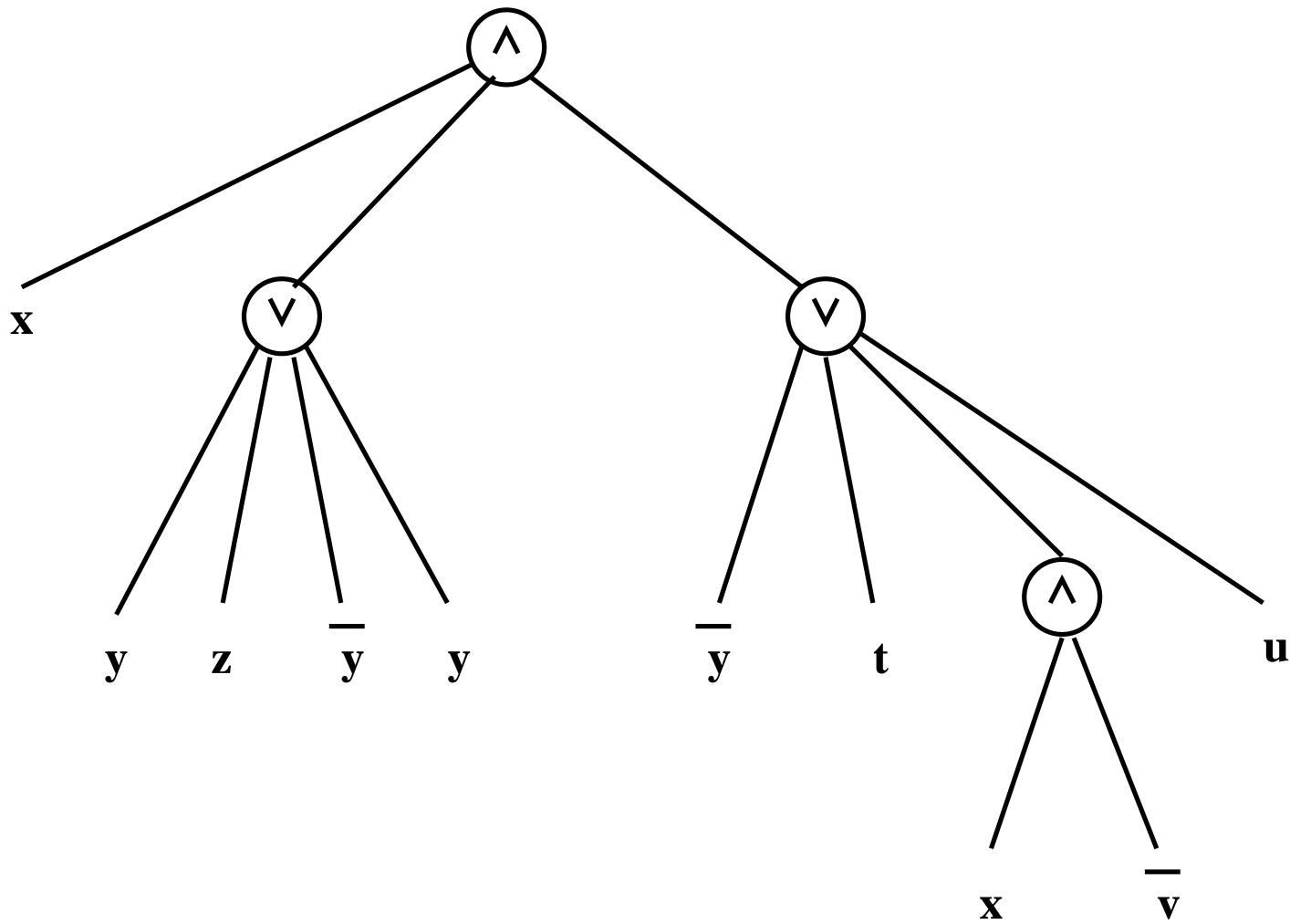
$$x \wedge (y \vee z \vee \bar{y} \vee y) \wedge (\bar{y} \vee t \vee (x \wedge \bar{v}) \vee u)$$

- Arbre et/ou : binaire
- Formule réduite et arbre associé
- Formule rigide

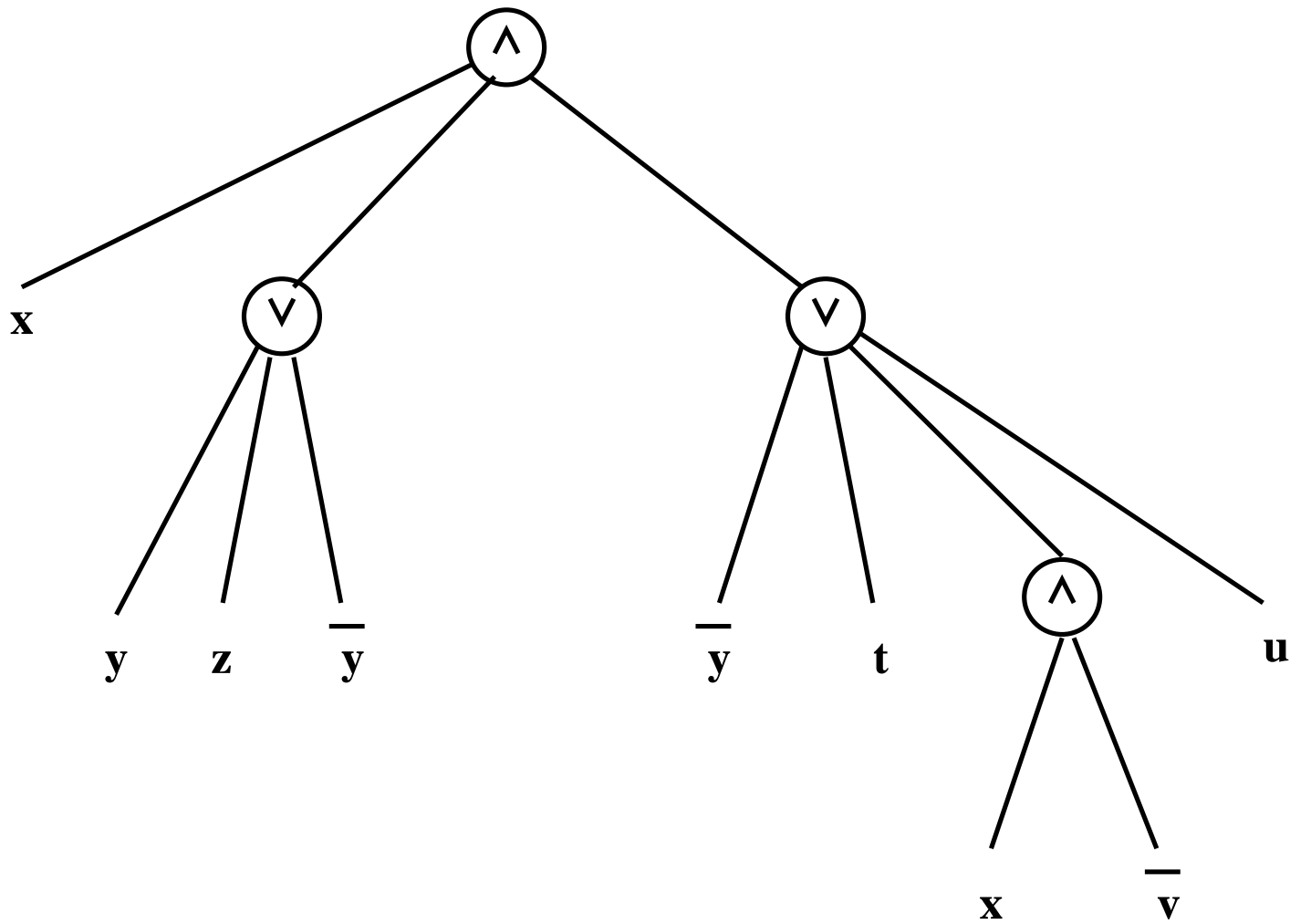
$$\mathbf{x \wedge (y \vee z \vee \bar{y} \vee y) \wedge (\bar{y} \vee t \vee (x \wedge \bar{v}) \vee u)}$$



$$\mathbf{x \wedge (y \vee z \vee \bar{y} \vee y) \wedge (\bar{y} \vee t \vee (x \wedge \bar{v}) \vee u)}$$



$$\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z} \vee \overline{\mathbf{y}} \vee \mathbf{y}) \wedge (\overline{\mathbf{y}} \vee \mathbf{t} \vee (\mathbf{x} \wedge \overline{\mathbf{v}}) \vee \mathbf{u})$$



Complexité de la représentation arborescente

Soit $L(f)$ la complexité de la représentation arborescente de f

1. Presque sûrement (pour la distribution uniforme sur \mathcal{B}_n),

$$L(f) \geq \frac{2^n}{\log n} \left(1 - \frac{1}{\log n}\right)$$

2. Pour toute fonction $f \in \mathcal{B}_n$,

$$L(f) \leq \frac{2^n}{\log n} (1 + o(1))$$

On a encore un *effet de Shannon*: presque toutes les fonctions Booléennes ont une complexité en arbre proche de la complexité maximale, i.e. d'ordre $2^n / \log n$.

Peut-on dire que presque tous les arbres de taille “assez grande” définissent des fonctions Booléennes de complexité d'ordre $2^n / \log n$?

- Si on modifie l'ensemble des connecteurs (en gardant le pouvoir d'exprimer toutes les fonctions de \mathcal{B}_n), $L(f)$ varie polynomialement.
- Borne inférieure pour des formules avec les connecteurs \wedge , \vee et \neg : $L(f) \geq \Omega(n^{5/2-\epsilon})$ (pour tout $\epsilon > 0$)
- Borne inférieure pour des formules avec les 16 opérateurs binaires: $L(f) \geq \Omega(n^2 / \log n)$

Relations entre les complexités des circuits $C(f)$ (taille minimale) et $D(f)$ (profondeur minimale) et la complexité des arbres $L(f)$?

- $C(f) \leq L(f)$
- $D(f) = \Theta(\log L(f))$: $L(f)$ est polynomiale ssi f est calculable par un circuit de profondeur logarithmique.
- Peut-on être plus précis en restreignant les formules, par exemple en se limitant aux formules monotones?
 \exists problème tel que $C(f)$ soit polynomial, $D(f)$ en $O((\log n)^2)$ et $L(f)$ superpolynomial, en $n^{O(\log n)}$.

Presque toutes les 2^{2^n} fonctions Booléennes ont une complexité exponentielle... mais les autres?

Savicky-Woods: Que peut-on dire sur les fonctions de complexité “modérée”?

Pour L “petit”: $L = o(2^n/n^k)$, le nombre $B(n, L)$ de fonctions de complexité L est donné par

$$\log B(n, L) = L \log n + L \left(\log \frac{2}{\ln 4 - 1} - o(1) \right)$$

On a donc $B(n, L) = (Cn)^L e^{o(L)}$.

Pour les fonctions de complexité $L(f)$ polynomiale, la complexité $C(f)$ est du même ordre: Si $L(f) \leq n^\alpha$ avec $\alpha > 1$, alors p.s. $C(f) > L(f)/\alpha$.

Techniques employées:

- énumération du nombre de fonctions “read-once”:

$$r_n \sim \sqrt{2(\ln 4 - 1)} \left(\frac{2}{e(\ln 4 - 1)} \right)^n n^{n-1}$$

- énumération de certains types d’arbres correspondant aux formules réduites ou rigides, et analyse asymptotique

Diagrammes de décision binaires

MDA: Minimal Deterministic Automaton: arbre de décision binaire où on compacte les chemins égaux \Rightarrow graphe orienté acyclique avec une racine et 2 feuilles.

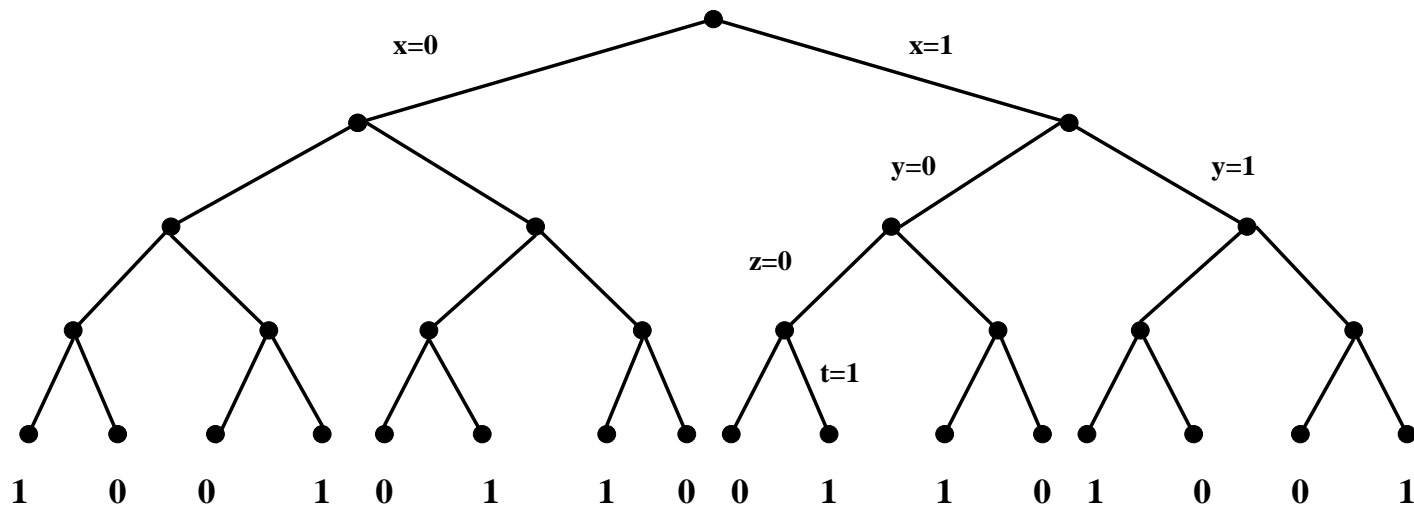
Les 2^n chemins allant de la racine à une feuille sont de longueur exactement n .

Diagramme de décision binaire: on part du MDA; on fusionne deux noeuds s'ils sont au même niveau et si leurs successeurs sont identiques.

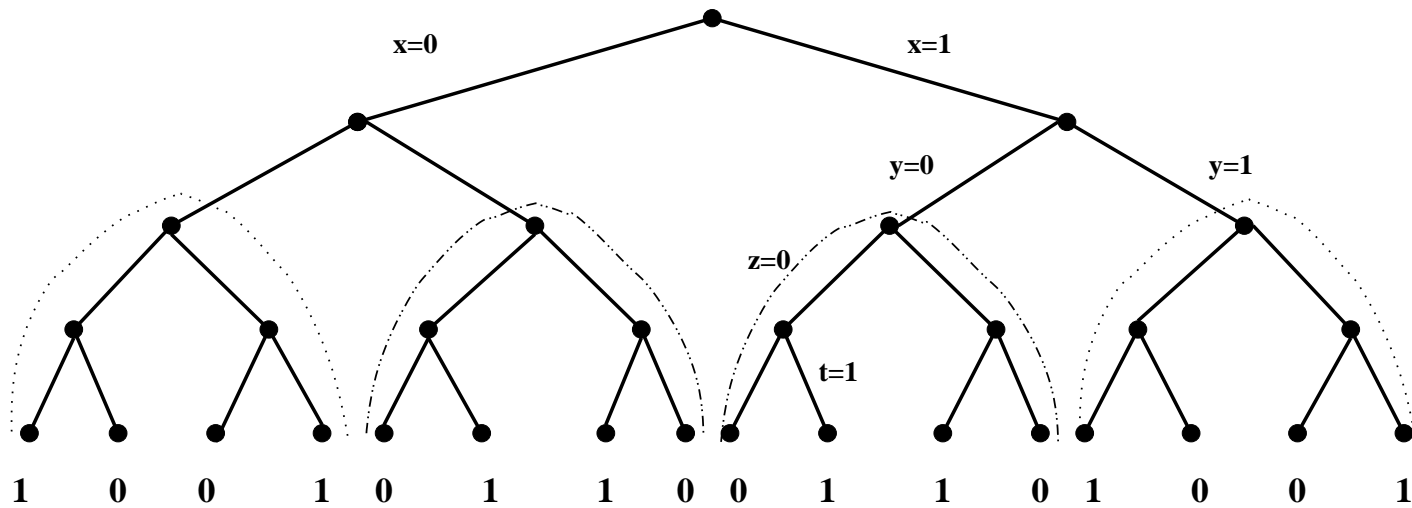
C'est parfois appelé un BDD quasi-réduit.

Fonction $(x \Leftrightarrow y) \Leftrightarrow (z \Leftrightarrow t)$

Arbre de décision pour l'ordre x, y, z, t



Reconnaître des sous-arbres communs



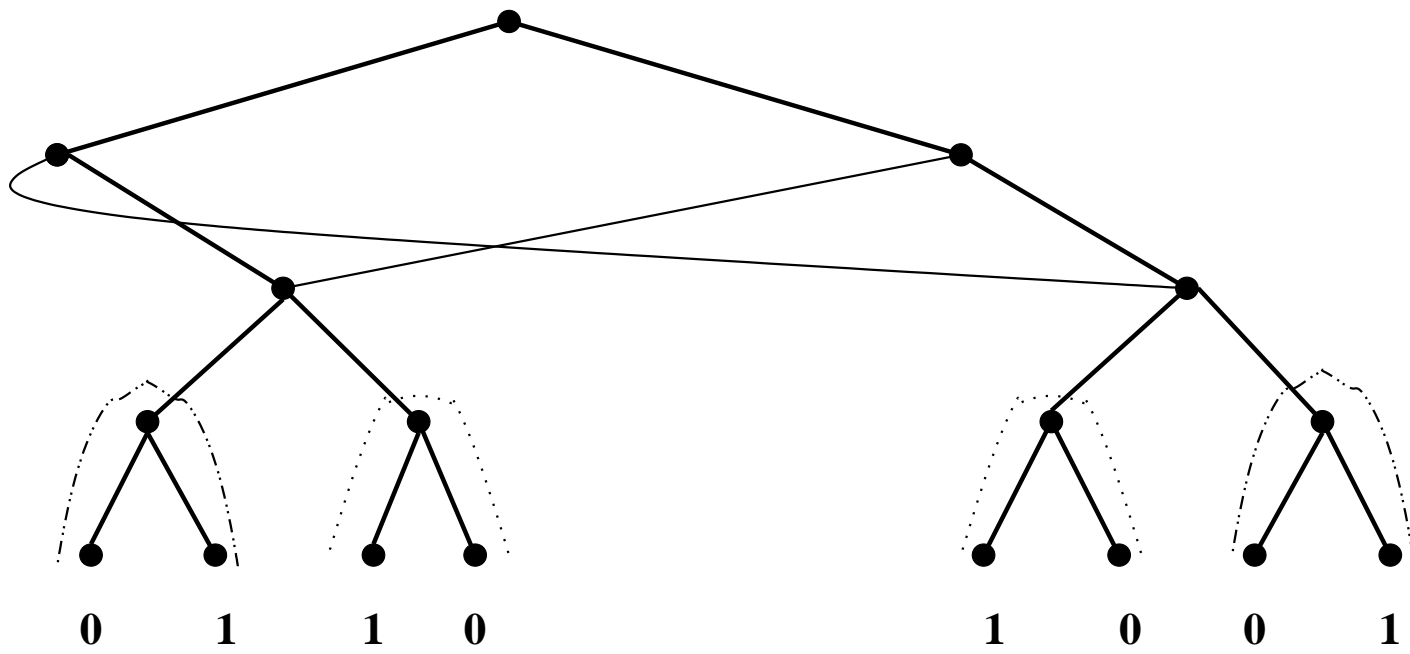


Diagramme *ordonné* : ordre fixé pour l'examen des variables.

Diagramme *réduit*: on réduit les chemins (si un noeud a ses deux successeurs égaux entre eux, on supprime ce noeud)

ROBDD: diagramme de décision binaire réduit et ordonné.

Représentation canonique: Pour toute fonction Booléenne f , il y a exactement *un* ROBDD sous l'ordre x_1, x_2, \dots, x_n qui la représente.

Conséquence: on peut tester en temps constant si un ROBDD représente une tautologie.

Complexité sur les BDDs

Fonction f tirée uniformément dans \mathcal{B}_n .

Arbre de décision: avec 2^n noeuds.

Taille moyenne du diagramme obtenu en réduisant l'arbre de décision $\sim C 2^n / n$.

C fluctue avec n .

(Flajolet-Kieffer-Yang 99, Gropl-Promel-Srivastav 98,
Vuillemin-Beal 04)

Dans l'arbre de décision complet associé à f :

- Au premier niveau, un seul sommet
- Au second niveau, 2 sommets
- ... Au j ème niveau, $J = 2^{j-1}$ sommets

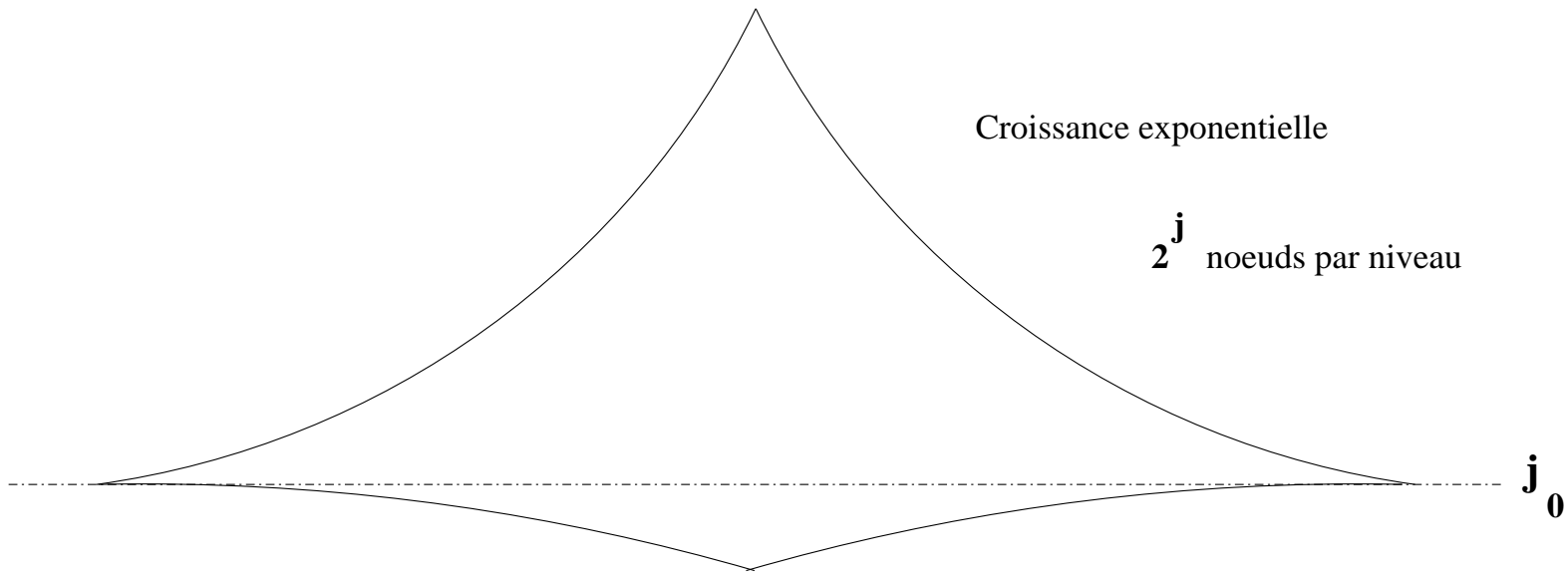
Quand on réduit:

- Toujours un sommet au premier niveau
- Deux sommets au niveau 2, *sauf si* $f|_{x_1=0} = f|_{x_1=1}$, ce qui arrive avec une probabilité $2^{2^{n-1}}/2^{2^n}$: le nombre moyen de sommets est $2 - 2^{-2^{n-1}}$
- Au niveau 2, la probabilité que les 4 sommets de l'arbre de décision soient présents est $\sim 1 - 6/2^{2^{n-2}}$.
- Au niveau j , les J sommets sont présents avec une probabilité $\sim 1 - J!/2^{2^{n-j}}$.

⇒ En partant du haut de l'arbre, le nombre moyen de sommets à profondeur j est $\sim 2^j$: croissance exponentielle.

- Au niveau n : il y a deux fonctions: 0 et 1
- Au niveau $n - 1$: 4 fonctions d'une seule variable x_n
- Au niveau $n - 2$: 16 fonctions de deux variables x_{n-1}, x_n

⇒ En partant du bas de l'arbre, le nombre moyen de sommets au niveau $n - p$ est $\sim 2^{2^p}$: croissance double exponentielle.



Croissance exponentielle

2^j noeuds par niveau

Croissance double exponentielle

2^{2^p} noeuds par niveau

j_0

Niveau critique : j_0 tel que

$$2^{j_0} = 2^{2^n - j_0}$$

On trouve que $j_0 \sim n - \log n$

- Haut de l'arbre : j_0 niveaux “presque” complets, donc $\sim 2^{j_0} \sim 2^n/n$ noeuds
 - Bas de l'arbre : $n - j_0$ niveaux, donnant aussi de l'ordre de $2^n/n$ noeuds
- \Rightarrow la taille du diagramme est donc $O(2^n/n)$

Taille moyenne du diagramme

C'est $\sum_j E[S(n, j)]$, où $S(n, j)$ est le nombre de sommets à profondeur j .

A profondeur j : on tire $2^j - 1$ fonctions parmi les $2^{2^{n-j}}$ fonctions Booléennes sur $n - j$ variables

Quel est le nombre moyen $E[S(n, j)]$ d'éléments distincts obtenus?

⇒ problème du *collectionneur de coupons!*

$$E[S(n, j)] = 2^{2^{n-j}} \left[1 - \left(1 - 2^{-2^{n-j}} \right)^{2^j} \right]$$

Pour résumer: représentations d'une fonction Booléenne

1. Table de vérité ou arbre de décision: taille 2^n
2. Arbres (formules): taille moyenne $2^n / \log n$
3. Circuits: taille moyenne $2^n / n$
4. Diagrammes de décision binaires: taille moyenne $2^n / n$

La probabilité de *Vrai*

Quelle est la probabilité qu'une formule logique définisse une tautologie (soit toujours vraie)?

- Si toutes les fonctions sur n variables sont équiprobables, c'est $1/|\mathcal{B}_n| = 2^{-2^n}$
- Si on construit une formule "au hasard" sur n variables, est-ce qu'on obtient n'importe quelle fonction (donc une tautologie) avec une probabilité uniforme?
- Est-ce raisonnable de comparer des formules de longueur différente?

1. On prend des formules toutes de même taille “grande”, et on regarde la distribution de probabilité $P(.)$ induite sur \mathcal{B}_n
 2. On construit un arbre suivant certaines règles (par ex. processus de branchement) et on l’étiquette: distribution de probabilité $\pi(.)$ induite sur \mathcal{B}_n
- Comment varie la probabilité de *Vrai* si on change l’ensemble des connecteurs autorisés?
 - Et si le nombre de variables Booléennes est infini?

Loi de probabilité induite sur \mathcal{B}_n par l'équiprobabilité des arbres de taille fixée

Loi de probabilité uniforme sur les arbres de type donné (binaire ou non, planaire ou non...) et de taille m donnée \Rightarrow loi P_m induite sur \mathcal{B}_n :

- $T_m(f)$ nombre d'arbres de taille m représentant f
- $T_m = \sum_f T_m(f)$: nombre total d'arbres de taille m

$$P_m(f) = \frac{T_m(f)}{T_m}$$

Puis m tend vers l'infini \Rightarrow loi limite $P(f)$

Fonctions génératrices de dénombrement $T_f(z) = \sum_m T_m(f)z^m$ et $T(z) = \sum_m T_m z^m$:

$$P_m(f) = \frac{[z^m]T_f(z)}{[z^m]T(f)}$$

Si on connaît les équivalents asymptotiques de $T_m(f)$ et de T_m pour $m \rightarrow +\infty$, alors on aura le comportement de $P_m(f)$ pour $m \rightarrow +\infty$, et donc on pourra décider de l'existence d'une limite, et de sa valeur éventuelle.

Processus de branchement et loi induite sur \mathcal{B}_n

- On engendre un arbre binaire planaire par un processus de branchement: on part de la racine; à chaque noeud on décide de s'arrêter ou d'avoir deux fils, avec équiprobabilité.
- L'arbre obtenu est p.s. fini.
- On étiquette les noeuds internes par \wedge et \vee (équiprobables).
- On étiquette les feuilles par les $2n$ littéraux (équiprobables).

On obtient ainsi une distribution sur les arbres et/ou finis, qui induit une *autre* distribution de probabilité $\pi(f)$ sur \mathcal{B}_n .

$$\pi(f) = \sum_{\tau \text{ calcule } f} \text{Proba}(\tau)$$

On montre que $\pi(f)$ s'exprime à l'aide des valeurs des fonctions génératrices $T_f(z)$ et $T(z)$, prises en leur singularité dominante (algébrique) commune ρ :

$$\pi(f) = \frac{T_f(\rho)}{T(\rho)}$$

Un exemple simple: arbres et/ou, n=1

Les connecteurs sont \vee , \wedge et \neg

\mathcal{T} ensemble de tous les arbres: arbres de Catalan; noeuds internes étiquetés par \vee et \wedge , feuilles étiquetées par x ou $\neg x$.

$$T(z) = \frac{1 - \sqrt{1 - 16z}}{4}$$

Singularité $\rho = 1/16$

$$T_m = 2^{2m+1} C_m$$

Quatre fonctions: *Vrai*, *Faux*, x et $\neg x$.

Calcul des séries génératrices de dénombrement $T_f(z)$

\mathcal{T}_f ensemble des arbres représentant la fonction f

$$\begin{aligned}\mathcal{T}_{Vrai} = & (\wedge, \mathcal{T}_{Vrai}, \mathcal{T}_{Vrai}) \oplus (\vee, \mathcal{T}_x, \mathcal{T}_{\bar{x}}) \oplus (\vee, \mathcal{T}_{\bar{x}}, \mathcal{T}_x) \\ & \oplus (\vee, \mathcal{T}_{Vrai}, \mathcal{T}) \oplus (\vee, \mathcal{T}, \mathcal{T}_{Vrai}) \setminus (\vee, \mathcal{T}_{Vrai}, \mathcal{T}_{Vrai})\end{aligned}$$

$$\begin{aligned}\mathcal{T}_x = & \{x\} \oplus (\wedge, \mathcal{T}_x, \mathcal{T}_x) \oplus (\wedge, \mathcal{T}_x, \mathcal{T}_{Vrai}) \oplus (\wedge, \mathcal{T}_{Vrai}, \mathcal{T}_x) \\ & \oplus (\vee, \mathcal{T}_x, \mathcal{T}_{Faux}) \oplus (\vee, \mathcal{T}_{Faux}, \mathcal{T}_x) \oplus (\vee, \mathcal{T}_x, \mathcal{T}_x)\end{aligned}$$

Equations similaires sur \mathcal{T}_{Faux} et $\mathcal{T}_{\bar{x}}$

Par symétrie, $T_{\bar{x}} = T_x$ et $T_{Faux} = T_{Vrai}$.

Les fonctions génératrices satisfont un système algébrique:

$$\begin{aligned}T_{Vrai}(z) &= 2zT_x(z)^2 + 2zT_{Vrai}(z)T(z) \\T_x(z) &= 1 + 2zT_x(z)^2 + 4zT_x(z)T_{Vrai}(z)\end{aligned}$$

On résoud et on obtient

$$\begin{aligned}T_{Vrai}(z) &= \frac{1}{8z} \left(2 - \sqrt{2 + 16z + 2\sqrt{1 - 16z}} \right), \\T_x(z) &= \frac{-1}{8z} \left(1 + \sqrt{1 - 16z} - \sqrt{2 + 16z + 2\sqrt{1 - 16z}} \right).\end{aligned}$$

Calcul de la loi $P(\cdot)$

$T_x(z)$ et $T_{Vrai}(z)$ ont la même singularité algébrique $\rho = 1/16$ que $T(z)$. Un lemme de transfert donne

$$[z^m]T_x(z) \sim 2^{2m+2} \frac{\sqrt{3}-1}{\sqrt{3}} C_{m-1}$$

d'où

$$P_m(x) \sim \frac{\sqrt{3}-1}{\sqrt{3}} \cdot \frac{m+1}{2m-1} \rightarrow \frac{\sqrt{3}-1}{2\sqrt{3}} = P(x) = 0.2113\dots$$

De même, on montre que

$$P_m(Vrai) \rightarrow \frac{1}{2\sqrt{3}} = P(Vrai) = 0.2886\dots$$

Complexité moyenne: 0.577

Calcul de la loi $\pi(\cdot)$

$$T_{Vrai}(z) = \frac{1}{8z} \left(2 - \sqrt{2 + 16z + 2\sqrt{1 - 16z}} \right),$$

$$T_x(z) = \frac{-1}{8z} \left(1 + \sqrt{1 - 16z} - \sqrt{2 + 16z + 2\sqrt{1 - 16z}} \right).$$

$$\pi(Vrai) = \frac{T_{Vrai}(1/16)}{T(1/16)} = \frac{2 - \sqrt{3}}{2} = 0.1339\dots$$

$$\pi(x) = \frac{T_x(1/16)}{T(1/16)} = \frac{\sqrt{3} - 1}{2} = 0.3660\dots$$

Complexité moyenne: 0.268

Pour un nombre quelconque de variables

On cherche à étendre les calculs faits pour $n = 1$.

Soit $f \in \mathcal{B}_n$; on peut écrire une équation sur l'ensemble \mathcal{T}_f

$$\mathcal{T}_f = 1_{\{f \text{ literal}\}} \oplus \sum_{g,h:f=g\vee h} (\vee, \mathcal{T}_g, \mathcal{T}_h) \oplus \sum_{g,h:f=g\wedge h} (\wedge, \mathcal{T}_g, \mathcal{T}_h).$$

se traduisant en équation sur la fonction génératrice $T_f(z)$

$$\begin{aligned} T_f(z) = & 1_{\{f \text{ literal}\}} + z \sum_{g,h:g\vee h=f} T_g(z) T_h(z) \\ & + z \sum_{g,h:g\wedge h=f} T_g(z) T_h(z) \end{aligned}$$

On fait cela pour les 2^n fonctions \Rightarrow système algébrique de taille 2^{2^n}

Pour $n = 2$, système de taille 16

Des symétries permettent de regrouper les fonctions Booléennes ayant mêmes séries génératrices:

- permutation de variables
- négation d'une variable
- négation de la formule toute entière

On se ramène ainsi à 4 fonctions Booléennes “intrinsèquement différentes”: $Vrai$, x , $x \wedge y$, $x \oplus y$.

On a donc un système de 4 équations quadratiques à 4 inconnues, qu'on peut résoudre explicitement.

Pour $n = 3$, le système est de taille 256

Des symétries ramènent de même à 14 fonctions Booléennes, i.e. à un système de taille 14. On peut écrire explicitement le système, puis le résoudre (par ex. par itération).

Pour $n = 4$: le système est de taille 65 536

Nombre de fonctions “intrinsèquement différentes”?

Et pour $n \geq 5$?

Classes d'équivalence de fonctions Booléennes

Les fonctions définies par les formules $x \wedge y$, $y \wedge x$, $x \wedge \bar{y}$, $\bar{x} \wedge \bar{y}$, $\bar{x} \vee \bar{y} = \neg(x \wedge y)$ “se ressemblent” \Rightarrow elles auront la même série génératrice.

On les regroupe dans une même classe d'équivalence.

Peut-on dénombrer ces classes?

En d'autres termes, quel est le nombre de fonctions “essentiellement différentes” sur n variables Booléennes?
(Polya 1940, Harrison vers 1960-70)

Relation d'équivalence sur les fonctions Booléennes de n variables:

f et g sont équivalentes si on passe de f à g par une suite d'opérations parmi:

- permutation de variables,
- complémentation d'une variable,
- négation de la formule toute entière.

Pour évaluer la taille du système à résoudre, on est donc amené à dénombrer des classes d'équivalence

L'outil pour faire cela: *théorie de Polya*

n	1	2	3	4	5	6
$ \mathcal{B}_n $	2	16	256	65 536	$4.2 \cdot 10^9$	$1.8 \cdot 10^{19}$
Nbre classes	2	4	14	222	616 126	$2 \cdot 10^{14}$

Un outil pour l'énumération de familles d'arbres: le théorème de Drmota-Lalley-Woods

Famille d'arbres \leftrightarrow langages algébriques \leftrightarrow système d'équations algébriques.

Drmota 97, Lalley 93, Woods 97

Système polynomial non linéaire

$$\{y_j = \Phi_j(z, y_1, \dots, y_m)\}, 1 \leq j \leq m$$

1. *a-positivité*: Tous les termes des séries $\Phi_j(\vec{y})$ sont positifs ou nuls.
2. *a-propre*: Φ est une contraction, i.e. satisfait une condition de Lipschitz ($K < 1$)

$$\begin{aligned} & d(\Phi(y_1, \dots, y_m), \Phi(y'_1, \dots, y'_m)) \\ & < K d((y_1, \dots, y_m), (y'_1, \dots, y'_m)) \end{aligned}$$

3. *a-irréductibilité*: le *graphe de dépendances* est fortement connexe.
Graphe de dépendances: sommets $1, 2, \dots, m$; arcs $k \rightarrow j$ si y_j apparaît dans ϕ_k .
4. *a-apériodicité*: z (et non z^2 ou z^3 ou...) est la variable pertinente
Pour chaque ϕ_j , il existe trois monômes z^a, z^b et z^c tels que $b - a$ et $c - a$ sont premiers entre eux.

Alors

1. Toutes les coordonnées y_j de la solution ont même rayon de convergence $\rho < \infty$.
2. Il existe des fonctions h_j analytiques autour de l'origine, t.q. ($1 \leq j \leq m$)

$$y_j = h_j \left(\sqrt{1 - z/\rho} \right) \quad (z \rightarrow \rho^-)$$

3. Toutes les autres singularités dominantes sont de la forme $\rho \omega$ avec ω racine de l'unité.
4. Si le système est a-apériodique, alors les y_j ont ρ pour unique singularité dominante, et les coefficients ont un développement asymptotique complet de la forme

$$[z^n]y_j(z) \sim \rho^{-n} \left(\sum_{k \geq 1} d_k n^{-1-k/2} \right).$$

Retour aux probabilités P et π sur \mathcal{B}_n

Loi de probabilité $P(f)$ ou $\pi(f)$ induite par un type de formules/arbres:

- Elle s'obtient à partir de la fonction génératrice $T_f(z)$ des arbres calculant la fonction Booléenne donnée f
- Les T_f sont solution d'un système d'équations satisfaisant les conditions du théorème de Drmota-Lalley-Woods (en général).
- Alors: \exists une solution $(t_{f_1}, \dots, t_{f_{2^p}})$; les t_f ont une singularité dominante commune, unique et strictement positive $\rho < +\infty$, également rayon de convergence de la série globale $T(z)$

$$T(z) = \alpha - \beta\sqrt{1 - z/\rho} + O(1 - z/\rho);$$

$$T_f(z) = \alpha_f - \beta_f\sqrt{1 - z/\rho} + O(1 - z/\rho).$$

1. Calcul de $\pi(f)$

$$\pi(f) = \frac{T_f(\rho)}{T(\rho)} = \frac{\alpha_f}{\alpha}$$

2. Lemme de transfert (Flajolet-Odlyzko):

$$[z^m]\{a - b\sqrt{1 - z} + O(1 - z)\} = -b[z^m]\sqrt{1 - z} (1 + O(1/m)).$$

Pour tout f , $\lim_{n \rightarrow +\infty} P_m(f)$ existe, et $P(f)$ est donc bien défini.

$$P_m(f) = \frac{[z^m]T_f(z)}{[z^m]T(z)} \rightarrow \frac{\beta_f}{\beta}$$

Arbres non planaires:

- On peut définir $P(\cdot)$ de la même manière, comme limite de la proportion d'arbres de taille fixe calculant une fonction donnée
- La définition de $\pi(\cdot)$ peut s'étendre au cas non planaire
- Les équations définissant les T_f ne sont plus algébriques

On peut étendre le théorème de Drmota-Lalley-Woods à un système d'équations fonctionnelles (Drmota, Woods)

Un exemple simple: les arbres binaires non planaires

Si les opérateurs binaires \vee et \wedge sont commutatifs: les arbres de Catalan deviennent non planaires.

$$\mathcal{A} = \{\bullet\} \oplus \sum_{A \in \mathcal{A}} \{\bullet, A, A\} \oplus \sum_{\{A, B\} \in \mathcal{A}^2, A \neq B} \{\bullet, A, B\},$$

Sur la fonction génératrice:

$$A(z) = z + \sum_{A \in \mathcal{A}} z^{2|A|} + \sum_{\{A, B\} \in \mathcal{A}^2, A \neq B} z^{|A|+|B|},$$

D'où l'équation fonctionnelle:

$$A(z) = z + \frac{1}{2} (A(z^2) + A(z)^2).$$

Equation fonctionnelle:

$$A(z) = z + \frac{1}{2} (A(z^2) + A(z)^2).$$

Equation fonctionnelle \sim équation d'ordre 2 en $A(z)$ “perturbée” par $A(z^2)$.

$$A(z) = 1 - \sqrt{1 - 2z - A(z^2)}.$$

On remplace $A(z^2)$ par $1 - \sqrt{1 - 2z^2 - A(z^4)}$ et on obtient

$$A(z) = 1 - \sqrt{-2z + \sqrt{1 - 2z^2 - A(z^4)}}.$$

En itérant:

$$\begin{aligned} A(z) &= 1 - \sqrt{-2z + \sqrt{-2z^2 + \dots + \sqrt{1 - 2z^{2^p} - A(z^{2^{p+1}})}}} \\ &= z + z^2 + z^3 + 2z^4 + 3z^5 + 6z^6 + 11z^7 + 23z^8 \\ &\quad + 46z^9 + 98z^{10} + 207z^{11} + 451z^{12} + 983z^{13} \\ &\quad + 2179z^{14} + 4850z^{15} + 10905z^{16} + 24631z^{17} \\ &\quad + 56011z^{18} + 127912z^{19} + 293547z^{20} + O(z^{21}). \end{aligned}$$

Rayon de convergence ρ de $A(z)$?

$$A(\rho) = 1 \quad \Rightarrow \quad \rho = 0.4026975037\dots$$

Asymptotique des coefficients?

$A(z) = 1 - \sqrt{B(z)}$ avec $B(z) = 1 - 2z - A(z^2)$: la singularité dominante est algébrique et, près de ρ :

$$A(z) = 1 - \sqrt{(z - \rho)B'(\rho) + O((z - \rho)^2)}.$$

Un lemme de transfert donne

$$A_m \sim \frac{\lambda}{m\sqrt{m}} \left(\frac{1}{\rho}\right)^m,$$

avec

$$\lambda = \sqrt{\frac{-\rho B'(\rho)}{4\pi}} = \sqrt{\frac{\rho + \rho^2 A'(\rho^2)}{2\pi}} = 0.3187766259\dots$$

Probabilités induites sur \mathcal{B}_n par des arbres non planaires

- On calcule de même les fonctions génératrices $T_f(z)$, qui vérifient maintenant des équations fonctionnelles
- L'extension du théorème de Drmota-Lalley-Woods permet de montrer l'existence d'une singularité dominante algébrique
- Un lemme de transfert permet d'extraire le coefficient $[z^m]T_f(z)$
- On termine en montrant que la limite $[z^m]T_f(z)/[z^m]T(z)$ existe bien, et définit donc la probabilité $P(f)$ sur \mathcal{B}_n

La probabilité de *Vrai*

Premières tentatives sur la définition d'une loi de probabilité

“naturelle”: Paris-Vencovská-Wilmers'94

- Modèle des arbres et/ou
- Chaque variable Booléenne suit une loi de Bernoulli
- Le nombre de variables tend vers l'infini

Alors

- Existence d'une mesure limite
- La probabilité de la fonction *Vrai* est nulle
- Si on ajoute le connecteur \rightarrow , on obtient la même mesure
- Si on se restreint aux formules “read-once”, on obtient la même mesure

La probabilité de *Vrai* si n est fini

- Connecteur \rightarrow et variables sans négations (Zaionc et al. 2000)

$$\frac{1}{n}(1 + o(1)) \leq \text{Proba}(\text{Vrai}) \leq \frac{3}{n}(1 + o(1))$$

- Connecteur \rightarrow et négations sur les formules (Zaionc 2004)

$\text{Proba}(\text{Vrai})$ calculé pour $n = 1$

- Arbres et/ou (Gardy-Woods 05)

$$\text{Proba}(\text{Vrai}) \geq \frac{1}{16n}$$

- Connecteur unique \leftrightarrow (Matecki 2003)

$$\text{Proba}(\text{Vrai}) \sim \frac{1}{2^n}$$

Liens entre complexité en arbre et probabilité

On a vu que, pour la loi uniforme, “presque toutes” les fonctions ont une complexité d’ordre $2^n / \log n$

Si on regarde d’autres lois de probabilité que la loi uniforme, peut-on relier la complexité d’une fonction à sa probabilité?

- Que devient la complexité d’une fonction “moyenne”?
- Quelle est la probabilité d’une fonction de faible complexité?

Arbres et/ou

- Borne générale (Lefman-Savicky + CFGG)

$$\frac{1}{4} \cdot \left(\frac{1}{8n}\right)^{L(f)+1} \leq P(f) \leq \exp\left(-c \frac{L(f)}{n^2}\right) (1 + o(1)).$$

- Amélioration (Gardy-Woods 05) et borne sur π

$$\pi(f) > \frac{2}{(8n)^{L(f)}} (1 + o(1));$$

$$P(f) > \frac{1}{(8n)^{L(f)+1}} (1 + o(1)).$$

Fonctions Booléennes avec connecteurs \wedge et \oplus .

(Savicky)

$$T(x_1, x_2, x_3, x_4) = (x_1 \wedge x_2) \oplus x_3 \oplus x_4.$$

Opérateur T itéré k fois: sous-ensemble \mathcal{E}_k de fonctions Booléennes, comprenant les fonctions de complexité $L(f) \leq (3/2)^k$. On définit une loi de probabilité p_k , induite par les formules construites sur l'opérateur T ; alors, pour les fonctions f de \mathcal{E}_k :

$$(4n)^{-4^k} \leq p_k(f) \leq c^{-4^{k/4}}.$$