

Arbres généraux pour la représentation de fonctions booléennes aléatoires.

Journées Aléa 2012

B. Gittenberger, A. Genitrini, V. Kraus et *C. Mailler**.

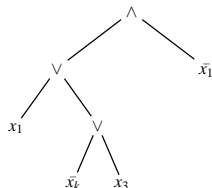
*Laboratoire de Mathématiques de Versailles
Université de Versailles-Saint-Quentin-en-Yvelines

7 mars 2012

Fonctions Booléennes et Arbres booléens de Catalan

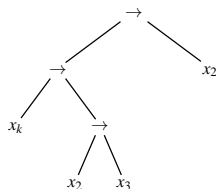
Fonctions Booléennes.

$$f : \begin{array}{l} \{0, 1\}^n \longrightarrow \{0, 1\} \\ (x_1, \dots, x_n) \longmapsto f(x_1, \dots, x_n) \end{array}$$

Modèle \wedge/\vee .connecteurs : \wedge ou \vee

littéraux positifs et négatifs

Modèle de l'implication

connecteur : \rightarrow

littéraux positifs uniquement

Un arbre (booléen) de Catalan calcule/représente une fonction booléenne, mais **plusieurs** arbres peuvent calculer une même fonction.

Quelques définitions

Taille d'un arbre booléen

taille = nombre total de littéraux (= nombre de feuilles dans le cas Catalan).

Si $\mathcal{T}_m = \{\text{arbres de Catalan de taille } m\}$ et $T_m = \#\mathcal{T}_m$, alors

$$T(z) = 2nz + 2T(z)^2 \Rightarrow T(z) = \frac{1 - \sqrt{1 - 16nz}}{4}$$

Singularité : $\rho_n = \frac{1}{16n}$.

Quelques définitions

Taille d'un arbre booléen

taille = nombre total de littéraux (= nombre de feuilles dans le cas Catalan).

Si $\mathcal{T}_m = \{\text{arbres de Catalan de taille } m\}$ et $T_m = \#\mathcal{T}_m$, alors

$$T(z) = 2nz + 2T(z)^2 \Rightarrow T(z) = \frac{1 - \sqrt{1 - 16nz}}{4}$$

Singularité : $\rho_n = \frac{1}{16n}$.

Complexité d'une fonction booléenne.

Complexité de f : $L(f)$ est la taille d'un arbre de taille minimale calculant f .

Exemple : la fonction XOR $x \oplus y = (x \wedge \bar{y}) \vee (\bar{x} \wedge y)$ est de complexité 4.

Loi de probabilité des arbres de Catalan.

Soit f une fonction booléenne à n variables.

$$\mathbb{P}_{m,n}(f) = \frac{T_m(f)}{T_m} = \frac{\# \text{ arbres de Catalan de taille } m \text{ calculant } f}{\# \text{ arbres de Catalan de taille } m}.$$

$$\xrightarrow{m \rightarrow \infty} \mathbb{P}_n(f) = \text{fraction limite des arbres calculant } f,$$

$\mathbb{P}_n(f)$ est appelée probabilité de f dans le modèle des arbres de Catalan.

Théorème [CHAUVIN ET AL. And/Or trees revisited. '04] [KOZIK. Subcritical pattern languages for And/Or trees. '09] :

La distribution de probabilité \mathbb{P}_n est bien définie et pour toute fonction f de n_0 variables,

$$\mathbb{P}_n(f) = \Theta\left(\frac{1}{n^{L(f)+1}}\right) \text{ quand } n \text{ tend vers } +\infty.$$

Ce modèle n'exhibe pas l'effet Shannon. (effet Shannon = “presque toutes les fonctions booléennes sont de complexité presque maximale”) [GENITRINI GITTENBERGER. '10]

Associativité et commutativité ?

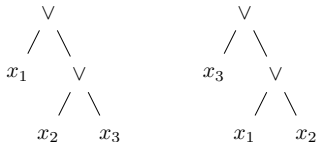
[GARDY. RANDOM BOOLEAN EXPRESSIONS. '06]

Dans le modèle des arbres de Catalan, la commutativité et l'associativité des connecteurs \wedge et \vee ne sont pas prises en compte...

Pourquoi distinguer les deux arbres suivants ?



Ou encore ces deux là ?



Associativité et commutativité ?

[GARDY. RANDOM BOOLEAN EXPRESSIONS. '06]

Idée : créer des modèles d'arbres naturels prenant en compte les propriétés d'associativité et de commutativité des connecteurs utilisés, et comparer les distributions induites sur les fonctions booléennes.

- Modèle de l'implication : associativité et commutativité des prémices.
- Modèle et/ou :

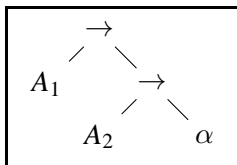
Propriété prise en compte	Modèle d'arbres utilisé
aucune	arbres de Catalan
associativité	arbres non binaires
commutativité	arbres non planaires
les deux	arbres non binaires, non planaires : arbres de Pólya

- Modèle et/ou associatif : une nouvelle notion de complexité.

Modèle de l'implication : commutativité des prémices

$$\frac{A_1, A_2 \vdash \alpha}{A_1 \vdash A_2 \rightarrow \alpha} \quad \begin{array}{l} \rightarrow i \\ \rightarrow i \end{array}$$

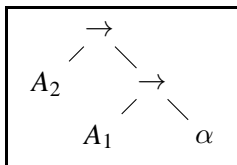
$$\vdash A_1 \rightarrow (A_2 \rightarrow \alpha)$$



$$A_1 \rightarrow (A_2 \rightarrow \alpha) \\ = \alpha \vee (\bar{A}_1 \wedge \bar{A}_2)$$

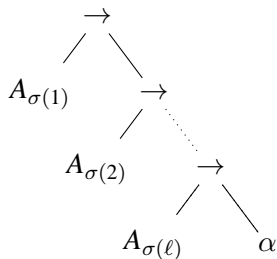
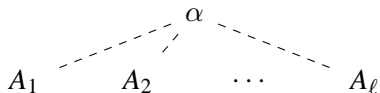
$$\frac{A_1, A_2 \vdash \alpha}{A_2 \vdash A_1 \rightarrow \alpha} \quad \begin{array}{l} \rightarrow i \\ \rightarrow i \end{array}$$

$$\vdash A_2 \rightarrow (A_1 \rightarrow \alpha)$$



$$A_2 \rightarrow (A_1 \rightarrow \alpha) \\ = \alpha \vee (\bar{A}_2 \wedge \bar{A}_1)$$

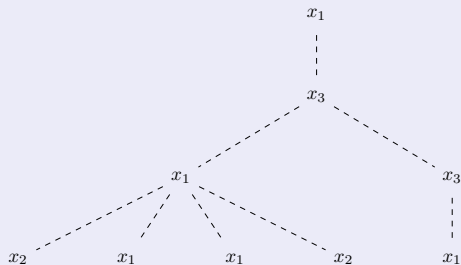
Modèle de l'implication : arbres généraux


 \Leftrightarrow


Arbre **général** : non planaire et d'arité quelconque.

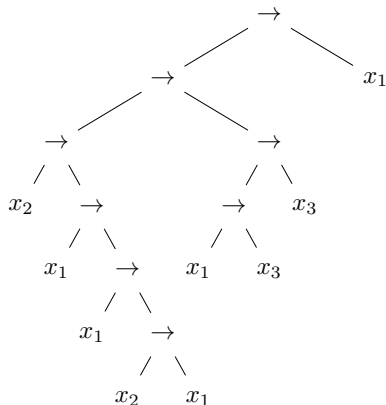
Un exemple

Exemple :



est de taille 9 et
calcule la fonction $x_3 \rightarrow x_1$.

Version binaire :



Résultats

- On définit $\mathbb{P}_m^i(f)$ la proportion d'arbres de taille m calculant la fonction f : la distribution limite $\mathbb{P}^i(f) = \lim_{m \rightarrow \infty} \mathbb{P}_m^i(f)$ existe.
(outils : fonctions génératrices, et théorème de Drmota-Lalley-Woods)
- Pour toute fonction f à n_0 variables,

$$\mathbb{P}^i(f) = \frac{\lambda_f^i}{n^{L(f)+1}} + \mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right) \text{ quand } n \rightarrow \infty.$$

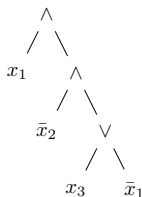
(idées clef :

- ▶ “un arbre typique contient le moins de répétitions de littéraux possible” et
 - ▶ “un arbre typique calculant f est un arbre minimal sur lequel est greffé un grand arbre”)
- Pas d'effet Shannon.

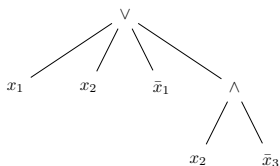
[GGKM. PROBABILITIES OF BOOLEAN FUNCTIONS GIVEN BY RANDOM IMPLICATIONAL FORMULAS. '11, SOUMIS À EJC.]

Arbres et/ou

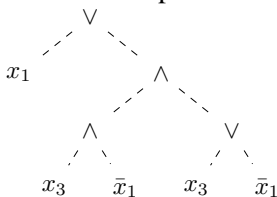
Arbres de Catalan



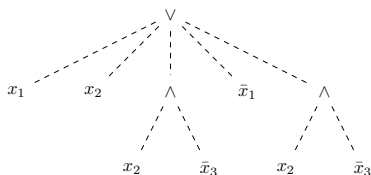
Arbres non binaires

**stratifiés**

Arbres non planaires



Arbres ni binaires ni planaires

**stratifiés**

Comment dénombrer des arbres non planaires ?

Arbres de Catalan : $T(z) = 2nz + 2T(z)^2$

Question : Comment obtenir une telle équation fonctionnelle pour des arbres binaires non planaires ?

$$C(z) = 2nz + 2 \left(\frac{C(z)^2}{2} + \frac{C(z^2)}{2} \right)$$

car $C(z)^2$ compte $2 \times \{A_1 \neq A_2\} + 1 \times \{A_1 = A_2\}$,

et $C(z^2) = \sum_{A \in \mathcal{C}} z^{2|A|}$ compte $1 \times \{A_1 = A_2\}$.



Cette méthode se généralise en une méthode symbolique grâce au **comptage de Pólya**.

L'indice de cycle

[PÓLYA. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und Chemische]

Définition :

- Soit $\sigma \in \mathcal{S}_n$. L'indice de cycle de σ est :

$$ZT(\sigma)(s_1, \dots, s_n) = s_1^{\lambda_1(\sigma)} s_2^{\lambda_2(\sigma)} \dots s_n^{\lambda_n(\sigma)},$$

où $\lambda_i(\sigma)$ est le nombre de cycle de longueur i de σ .

- Soit G un sous-groupe de \mathcal{S}_n , alors son indice de cycle est :

$$P(G)(s_1, \dots, s_n) = \frac{1}{|G|} \sum_{\sigma \in G} ZT(\sigma).$$

- Soit \mathcal{F} une famille d'objets. Pour tout $t \in \mathcal{F}$, on note $\mathcal{S}(t)$ le groupe des symétrie de t . L'**indice de cycle** de \mathcal{F} est :

$$Z_{\mathcal{F}}(s_1, s_2, \dots) = \sum_{t \in \mathcal{F}} P(\mathcal{S}(t))(s_1, s_2, \dots, s_{|t|}).$$

L'indice de cycle

[PÓLYA. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und Chemische]

Définition :

- Soit $\sigma \in \mathcal{S}_n$. L'indice de cycle de σ est :

$$ZT(\sigma)(s_1, \dots, s_n) = s_1^{\lambda_1(\sigma)} s_2^{\lambda_2(\sigma)} \dots s_n^{\lambda_n(\sigma)},$$

où $\lambda_i(\sigma)$ est le nombre de cycle de longueur i de σ .

- Soit G un sous-groupe de \mathcal{S}_n , alors son indice de cycle est :

$$P(G)(s_1, \dots, s_n) = \frac{1}{|G|} \sum_{\sigma \in G} ZT(\sigma).$$

- Soit \mathcal{F} une famille d'objets. Pour tout $t \in \mathcal{F}$, on note $\mathcal{S}(t)$ le groupe des symétrie de t . La **fonction génératrice** de \mathcal{F} est :

$$Z_{\mathcal{F}}(z, z^2, \dots) = \sum_{t \in \mathcal{F}} P(\mathcal{S}(t))(z, z^2, \dots, z^{|t|}).$$

Comptage de Pólya

Composition : si $\mathcal{C} = \mathcal{B} \circ \mathcal{A}$, alors,

$$C(z) = Z_{\mathcal{B}}(A(z), A(z^2), \dots).$$

Exemple : arbres binaires commutatifs

On note K l'arbre composé d'une racine et de deux feuilles. Son groupe de symétrie est $\mathcal{S}_K = \{Id, (12)\}$, et son indice de cycle : $Z_K(s_1, s_2) = \frac{1}{2}(s_1^2 + s_2)$.
Donc, comme $\mathcal{C} = \mathcal{Z} + 2K \circ \mathcal{C}$, on obtient :

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

Set : si $\mathcal{C} = \text{Set}(\mathcal{X})$, alors,

$$Z_{\mathcal{C}}(s_1, s_2, \dots) = \exp \left(\sum_{i \geq 1} \frac{s_i}{i} \right).$$

Résultats

Quelle que soit la famille d'arbres étudiée :

- On définit $\mathbb{P}_m(f)$ la proportion d'arbres de taille m calculant la fonction f : la distribution limite $\mathbb{P}(f) = \lim_{m \rightarrow \infty} \mathbb{P}_m(f)$ existe.
(outils : fonctions génératrices, et théorème de Drmota-Lalley-Woods)

- Pour toute fonction f à n_0 variables,

$$\mathbb{P}(f) = \frac{\lambda_f^*}{n^{L(f)+1}} + \mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right) \text{ quand } n \rightarrow \infty.$$

On note λ_f cette constante dans le cas Catalan, λ_f^c dans le cas commutatif, λ_f^a dans le cas associatif, et $\lambda_f^{a,c}$ dans le cas général.
(idées clef :

- ▶ “un arbre typique contient le moins de répétitions de littéraux possible” et
 - ▶ “un arbre typique calculant f est un arbre minimal sur lequel est greffé un grand arbre”)
- Pas d'effet Shannon.

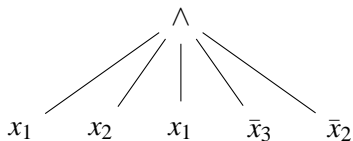
Les distributions ont le même comportement mais sont différentes

Pour toute fonction f à n_0 variables,

$$\mathbb{P}(f) = \frac{\lambda_f^*}{n^{L(f)+1}} + \mathcal{O}\left(\frac{1}{n^{L(f)+2}}\right) \text{ quand } n \rightarrow \infty.$$

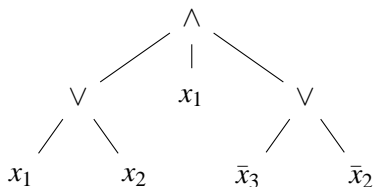
	Arbres de Catalan	Arbres associatifs	Arbres commutatifs	Arbres de Pòlya
<i>Vrai</i>	$\frac{3}{4} = 0.75$	$51 - 36\sqrt{2} \approx 0.0883$	$\frac{641}{1024} \approx 0.626$	$\frac{(2 \ln 2 - 1)^2}{4} \approx 0.0373$
x	$\frac{5}{16} \approx 0.313$	$546 - 386\sqrt{2} \approx 0.114$	$\frac{1153}{4096} \approx 0.281$	$\frac{(2 \ln 2 - 1)^2(2 \ln 2 + 1)}{4} \approx 0.0890$

Taille des arbres associatifs



nombre de littéraux = 5

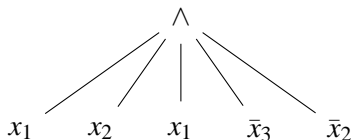
nombre de connecteurs = 1



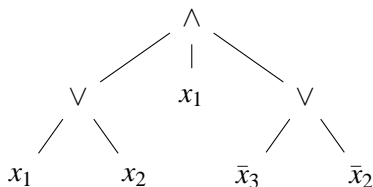
nombre de littéraux = 5

nombre de connecteurs = 3

Taille des arbres associatifs



nombre de littéraux = 5
nombre de connecteurs = 1



nombre de littéraux = 5
nombre de connecteurs = 3

Nouvelle notion de taille :

nombre total de nœuds \rightsquigarrow nouvelle notion de complexité.

Un modèle surprenant

- On définit $\mathbb{P}_m(f)$ la proportion d'arbres associatifs planaires de taille m calculant la fonction f : la distribution limite $\mathbb{P}(f) = \lim_{m \rightarrow \infty} \mathbb{P}_m(f)$ existe.
(outils : fonctions génératrices, et théorème de Drmota-Lalley-Woods)

Un modèle surprenant

- On définit $\mathbb{P}_m(f)$ la proportion d'arbres associatifs planaires de taille m calculant la fonction f : la distribution limite $\mathbb{P}(f) = \lim_{m \rightarrow \infty} \mathbb{P}_m(f)$ existe.

(outils : fonctions génératrices, et théorème de Drmota-Lalley-Woods)

- Il existe α et β deux constantes telles que

$$0 < \alpha < \mathbb{P}(\text{Vrai}) < \beta < \frac{1}{2}$$

(Rappel : avec la notion de taille précédente, $\mathbb{P}^a(\text{Vrai}) = \Theta\left(\frac{1}{n}\right)$)

Idées clef :

- ▶ ici les répétitions sont très probables
- ▶ un arbre a en moyenne \sqrt{n} feuilles au premier niveau

Un modèle surprenant

- On définit $\mathbb{P}_m(f)$ la proportion d'arbres associatifs planaires de taille m calculant la fonction f : la distribution limite $\mathbb{P}(f) = \lim_{m \rightarrow \infty} \mathbb{P}_m(f)$ existe.

(outils : fonctions génératrices, et théorème de Drmota-Lalley-Woods)

- Il existe α et β deux constantes telles que

$$0 < \alpha < \mathbb{P}(\text{Vrai}) < \beta < \frac{1}{2}$$

(Rappel : avec la notion de taille précédente, $\mathbb{P}^a(\text{Vrai}) = \Theta\left(\frac{1}{n}\right)$)

Idées clef :

- ▶ ici les répétitions sont très probables
 - ▶ un arbre a en moyenne \sqrt{n} feuilles au premier niveau
- Pas d'effet Shannon. Evident d'après le point précédent.

Un modèle surprenant

- Soit $\mathcal{D} = \{(x_1, \dots, x_n) \mapsto x_{i_1} \vee \dots \vee x_{i_k}\}_{\substack{k \in \{\sqrt{n}, \dots, 2\sqrt{n}\} \\ 1 \leq i_1 < \dots < i_k < n}}$.

Nous avons $\mathbb{P}(\mathcal{D}) > \gamma \geq 0$.

Ce type de résultat n'est pas démontré dans les autres modèles !

Conjecture :

Pour toute fonction booléenne f à n_0 variables,

$$\mathbb{P}(f) \sim \Theta\left(\frac{1}{n^{L(f)}}\right) \text{ quand } n \rightarrow \infty.$$

Vérifiée pour $f = \text{Vrai}$ et $f = (x_1 \dots, x_n) \mapsto x_i$.

Mais non démontré dans le cas général... On conjecture cependant qu'un arbre typique est aussi un arbre minimal sur lequel est greffé une grande tautologie.

Résumé des modèles étudiés

- **Modèle implication :**
 - ▶ arbres de Catalan
 - ▶ arbres généraux (prémices commutatifs et associatifs)
- **Modèle et/ou :**
 - ▶ arbres de Catalan
 - ▶ arbres plans et non binaires (associativité des connecteurs)
 - ▶ arbres non plans binaires (commutativité des connecteurs)
 - ▶ arbres généraux (commutativité et associativité)
- **Modèle et/ou associatif mais non commutatif : nouvelle complexité**
 - ▶ **mais forme des arbres typiques différente : plus de feuilles et plus de répétitions de littéraux.**
 - ▶ **comportement similaire de la distribution induite conjecturé,**

Perspectives

- Finir l'étude des arbres associatifs planaires avec nouvelle complexité, et l'étendre à des arbres associatifs et commutatifs.
- Mélanger différents types de connecteurs dans un même modèle (\wedge , \vee et \rightarrow).
- Généraliser à d'autres expressions : par exemple arithmétiques.

Perspectives

- Finir l'étude des arbres associatifs planaires avec nouvelle complexité, et l'étendre à des arbres associatifs et commutatifs.
- Mélanger différents types de connecteurs dans un même modèle (\wedge , \vee et \rightarrow).
- Généraliser à d'autres expressions : par exemple arithmétiques.

Merci de votre attention !