Arbres pour la logique

Danièle GARDY

PRiSM, Université Versailles St-Quentin en Yvelines
(et DMG, T.U. Wien, Autriche)

daniele.gardy@prism.uvsq.fr

ADAMA, 17 Octobre 2012

Travaux réalisés avec
B. Chauvin, P. Flajolet, H. Fournier, A. Genitrini,
B. Gittenberger, J. Kozik, C. Mailler, A. Woods, M. Zaionc...

Pour une présentation du sujet et des techniques
Random boolean expressions, DMTCS proceedings AF, 2006, pages 1-36

Pour une partie des travaux sur le sujet
www.prism.uvsq.fr/index.php?id=46

Voir aussi les pages web d'Antoine Genitrini, Jakub Kozik, Marek Zaionc...

Un exemple pour commencer...

Quelques formules du calcul des propositions:

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

$$(x \to (y \to y)) \to (z \to x)$$

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

Quelles fonctions booléennes calculent-elles?

Un exemple pour commencer...

Quelques formules du calcul des propositions:

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

$$(x \to (y \to y)) \to (z \to x)$$

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

Quelles fonctions booléennes calculent-elles?

$$x,\ z \to x, \text{ et } x \vee y$$

Une autre question naturelle...

Probabilité qu'une formule écrite "au hasard", sur k variables boolénnes avec les connecteurs $\wedge$ et $\vee$, soit toujours vraie?

- $k = 1$: 4 fonctions booléennes; $\mathit{Proba}(\mathit{Vrai}) = 0.2886$
- $k = 2$: 16 fonctions boolénnes; $\mathit{Proba}(\mathit{Vrai}) = 0.209$
- $k = 3$: 256 fonctions booléennes; $\mathit{Proba}(\mathit{Vrai}) = 0.165$
- $k \geq 4$: 2^{2^k} fonctions booléennes; peut-on calculer numériquement $\mathit{Proba}(\mathit{Vrai})$?
- $k \to +\infty$: asymptotique de $\mathit{Proba}(\mathit{Vrai})$?

De quoi parlons-nous?

- Pouvons-nous compter les expressions (formules) de taille donnée?
- A quoi ressemble une tautologie tirée au hasard?
- Que veut dire “au hasard”?
- Pouvons-nous comparer le pouvoir expressif de differents systèmes propositionnels?
- Pouvons-nous utiliser les *expressions* booléennes pour définir des lois de probabilité sur les *fonctions* booléennes?
- A quoi ressembleraient de telles distributions?
- Quelle serait la forme typique d’un arbre représentant une fonction donnée?
- Serait-il possible de relier la probabilité d’une fonction à sa complexité?

Les arbres d'expressions booléennes

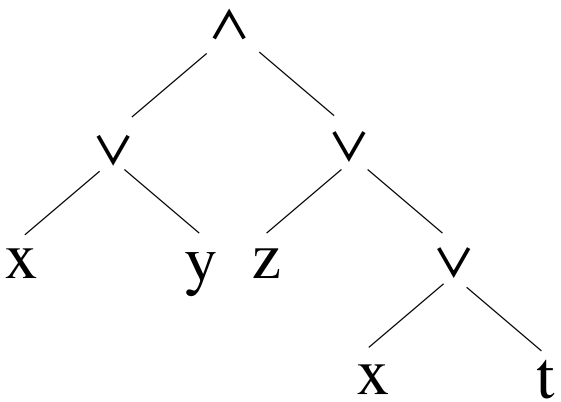
- On part des expressions booléennes...

$$(x \vee y) \wedge (z \vee (x \vee t))$$

- On part des expressions booléennes...

$$(x \vee y) \wedge (z \vee (x \vee t))$$

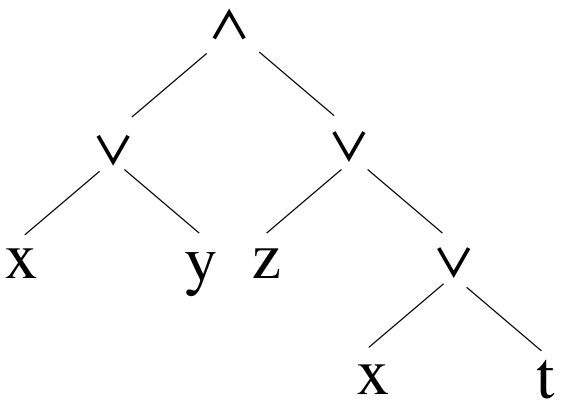
- ... elles peuvent être représentées par des arbres



- On part des expressions booléennes...

$$(x \vee y) \wedge (z \vee (x \vee t))$$

- ... elles peuvent être représentées par des arbres



- Une expression booléenne calcule (représente) une fonction booléenne

Cadre général

Formules booléennes sur un *nombre fixe k* de variables

$\mathcal{F}_k$ ensemble des fonctions booléennes à k variables

$$Card(\mathcal{F}_k) = 2^{2^k}$$

Cadre général

Formules booléennes sur un *nombre fixe k* de variables

$\mathcal{F}_k$ ensemble des fonctions booléennes à k variables

$$Card(\mathcal{F}_k) = 2^{2^k}$$

- Une formule sur k variables détermine une unique fonction de $\mathcal{F}_k$
- Une fonction de $\mathcal{F}_k$ est représentée par une infinité de formules booléennes sur (au moins) k variables

Cadre général

Système pour la logique propositionnelle: défini par des règles pour construire les formules:

- Ensemble de connecteurs logiques
- Chaque connecteur a une arité (fixe ou variable) et est, ou non, commutatif ou associatif
- On autorise, ou non, les littéraux négatifs
- On peut demander que tous les littéraux soient à la même profondeur

$\mathcal{E}_k$ ensemble des formules construites sur k variables, suivant un tel ensemble de règles
Notion de taille sur $\mathcal{E}_k$ (nombre de connecteurs, ou de littéraux, ou les deux)

Cadre général

On définit une application Φ

$$\begin{array}{rcl} \mathcal{E}_k & \rightarrow & \mathcal{F}_k; \\ \tau & \mapsto & f \text{ ssi } \tau \text{ calcule } f \end{array}$$

Cadre général

On définit une application Φ

$$\begin{array}{rcl} \mathcal{E}_k & \rightarrow & \mathcal{F}_k; \\ \tau & \mapsto & f \text{ ssi } \tau \text{ calcule } f \end{array}$$

Soit une loi de probabilité sur $\mathcal{E}_k$.

Quelle est la loi de probabilité induite sur $\mathcal{F}_k$?

Cadre général

Plusieurs modèles probabilistes "naturels" sur les arbres

1. **Modèle combinatoire**
   - Tous les arbres de même taille sont équiprobables
   - La taille tend vers l'infini
2. **Processus de branchement**
   - Arbres de Galton-Watson, processus de Boltzmann...
   - On construit un arbre, de taille aléatoire
   - On étiquette l'arbre obtenu suivant les règles du système logique
3. **Processus de croissance**
   - Croissance équilibrée (par niveaux)
   - Croissance des arbres binaires de recherche. Cf. le cours de Brigitte Chauvin

Loi image sur l'ensemble des fonctions booléennes?

Comment évaluer les probabilités sur les arbres et sur les fonctions?

Etablir des lois de probabilité sur des arbres revient souvent à énumérer les arbres calculant une fonction booléenne f, i.e. calculer des fonctions génératrices de dénombrement

1. Construction récursive des arbres $\sim$ établir un système d'equations (algébriques, différentielles, fonctionnelles...)
2. Résoudre le système?
   - Explicitement? mais 2^{2^k} fonctions booléennes $\Rightarrow$ système de **très** grande taille!
   - Définir des classes de fonctions équivalentes? Il y en a encore trop! [Polyà 40, Harrison 60]
   - Asymptotiquement? $\Rightarrow$ outils de combinatoire analytique Ex: système algébrique $\Rightarrow$ th. de Drmota-Lalley-Woods

Les outils de la combinatoire analytique

Enumeration et combinatoire analytique

Soit $\mathcal{C}$ une classe d'objets combinatoires (arbres, mots, ...) et soit c_n le nombre d'objets de *taille n*
Calcul de c_n? valeur asymptotique?

- $C(z) := \sum_n c_n z^n$ peut s'obtenir à partir d'une spécification de $\mathcal{C}$
- $C(z)$ vu comme une fonction analytique: on peut extraire $c_n = [z^n]C(z)$ (formule de Cauchy)
- Comportement asymptotique de c_n: à partir d'une expression explicite de c_n, ou par un lemme de transfert (Flajolet-Odlyzko), une analyse de point col, ...
- La clé: le comportement de $C(z)$ près de sa *singularité* dominante (réelle > 0)

Exemple: énumération des arbres de Catalan

On considère les arbres binaires complets (pas de noeud simple)

$$\mathcal{C} = \bullet + (\bullet, \mathcal{C}, \mathcal{C})$$

Taille: nombre n de noeuds internes (il y a $n+1$ feuilles)

Exemple: énumération des arbres de Catalan

On considère les arbres binaires complets (pas de noeud simple)

$$\mathcal{C} = \bullet + (\bullet, \mathcal{C}, \mathcal{C})$$

Taille: nombre n de noeuds internes (il y a $n+1$ feuilles)
C_n nombre d'arbres de taille n
Série génératrice $C(z) := \sum_{n \geq 0} C_n z^n$

$$C(z) = 1 + zC(z)^2$$

Exemple: énumération des arbres de Catalan

On considère les arbres binaires complets (pas de noeud simple)

$$\mathcal{C} = \bullet + (\bullet, \mathcal{C}, \mathcal{C})$$

Taille: nombre n de noeuds internes (il y a $n+1$ feuilles)
C_n nombre d'arbres de taille n
Série génératrice $C(z) := \sum_{n \geq 0} C_n z^n$

$$C(z) = 1 + zC(z)^2$$

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

Exemple: énumération des arbres de Catalan

Extraction de $C_n = [z^n]\frac{1-\sqrt{1-4z}}{2z}$?

$$[z^n](1+z)^\alpha = \frac{\alpha(\alpha-1)...(\alpha-n+1)}{n!}$$

d'où la valeur du nombre de Catalan C_n

$$C_n = \frac{(2n)!}{n!(n+1)!} \sim \frac{4^n}{\sqrt{\pi}\, n^{3/2}}$$

Exemple: énumération des arbres de Catalan

On peut aussi avoir l'asymptotique de C_n directement

- $C(z)$ a une singularité algébrique en $z = 1/4$ de type $\sqrt{1-z}$
- Lemme de transfert (Flajolet-Odlyzko)

Theorem

Sous de "bonnes" conditions, si $f(z)$ a pour singularité dominante 1 et $f(z) \sim (1-z)^{-a}$ avec $-a \notin N$, alors

$$[z^n]f \sim \frac{n^{a-1}}{\Gamma(a)}$$

- D'où l'asymptotique de C_n pour $n \to +\infty$

Les arbres Et/Ou

- Deux connecteurs $\wedge$, $\vee$ *binaires* et *non commutatifs*
- Littéraux x_i et $\bar{x}_i$ aux feuilles ($1 \leq i \leq k$)

Les arbres Et/Ou

- Deux connecteurs $\wedge$, $\vee$ *binaires* et *non commutatifs*
- Littéraux x_i et $\bar{x}_i$ aux feuilles ($1 \leq i \leq k$)

On aura des arbres de Catalan étiquetés

Les arbres Et/Ou

- Deux connecteurs $\wedge$, $\vee$ *binaires* et *non commutatifs*
- Littéraux x_i et $\bar{x}_i$ aux feuilles ($1 \leq i \leq k$)

On aura des arbres de Catalan étiquetés
Nombre d'expressions de taille n (nombre de noeuds internes)?

- Chaque noeud interne a 2 étiquettes possibles
- Chaque feuille a $2k$ étiquettes possibles

$$T(z) = 2k + 2zT(z)^2 \Rightarrow T(z) = \frac{1 - \sqrt{1 - 16kz}}{4z} = 2k\, C(4kz)$$

D'où $T_n = (2k)^{n+1} 2^n C_n$

Arbres associatifs

Si $\wedge$ et $\vee$ ne sont plus binaires, mais d'arité au moins 2?

- Arbres généraux planaires, sans noeud d'arité 1
- Taille = nombre de *feuilles*

$$T(z) = z + \frac{T(z)^2}{1 - T(z)}$$

- Quand on étiquette les neuds, l'opérateur change à chaque niveau de l'arbre: deux classes d'arbres, selon que la racine est étiquetée par $\wedge$ ou par $\vee$
- Exercice: dénombrer ces arbres

Arbres commutatifs

Ici $\wedge$ et $\vee$ sont commutatifs; la taille est le nombre de *feuilles*

- Arbres binaires non planaires (arbres de Polya)

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

- Pas de forme close pour la fonction génératrice
- $F(z^2)$ a un R.C. plus grand que $F(z)$: on le traite comme un paramètre

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

- On itère...

$$F(z) = 1 - \sqrt{-2z + \sqrt{-2z^2 + \ldots + \sqrt{-2z^{2^{p-1}} + \sqrt{1 - 2z^{2^p} - F(z^{2^{p+1}})}}}}$$

Arbres commutatifs

- Pas de forme close pour $[z^n]F(z)$
- Comportement asymptotique?
  - Rayon de convergence $\rho < 1$, solution de $F(\rho) = 1$
  - Au voisinage de $\rho = 0.40269...$

$$F(z) = 1 - \sqrt{(z-\rho)F_1(\rho) + O((z-\rho)^2)}$$

  - Lemme de transfert:

$$[z^n]F(z) \sim \frac{\lambda}{n\sqrt{n}} \left(\frac{1}{\rho}\right)^n ; \quad \lambda = \sqrt{\frac{\rho + \rho^2 F'(\rho^2)}{2\pi}} = 0.31877...$$

- Exercice: adaptation quand on étiquette les noeuds internes par des connecteurs et les noeuds externes par des littéraux

Les tautologies

Pourquoi étudier les tautologies?

- pour comparer le pouvoir d'expression de deux systèmes logiques

$$A \vdash B \qquad \text{ssi} \qquad A \to B \text{ est une tautologie}$$

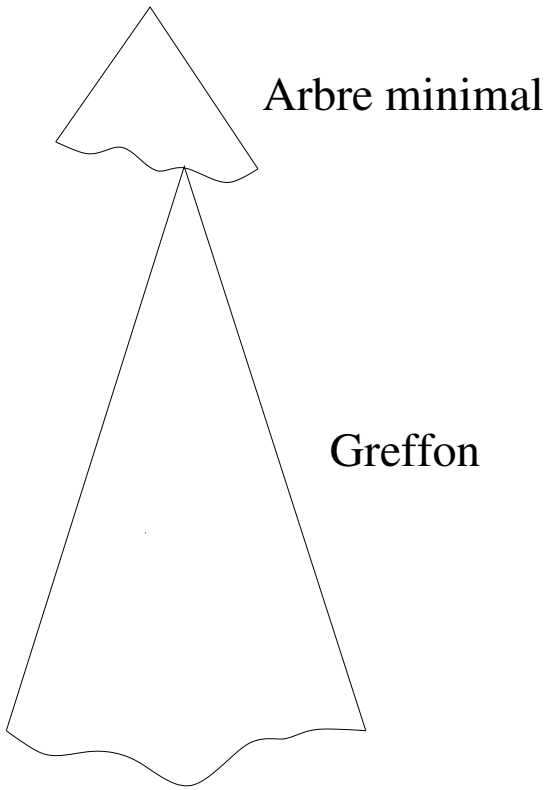
- parce que *Vrai* est une des fonctions les plus simples: on peut espérer voir à quoi ressemble une tautologie "typique"

Les tautologies dans différents systèmes logiques

Etude systématique initiée par Zaionc et al. vers 2000: probabilité asymptotique d'une tautologie dans divers systèmes logiques?

- Loi uniforme sur $\mathcal{F}_k$: $1/2^{2^k}$
- Connecteur $\leftrightarrow$: $1/2^k$ (taille de l'arbre paire) [Matecki 03]
- Implication: $1/k$; tautologies simples [Fournier et al. 07]
- Arbres Et/Ou: $1/k$; tautologies simples [Woods 05, Kozik 08]
- Même comportement en $1/k$ si les connecteurs sont associatifs ou commutatifs
- Pour des arbres "plus équilibrés": ?

Dans tous ces systèmes, les tautologies ont une forme semblable



Distributions en arbre

Ensemble $\mathcal{F}_k$ des fonctions booléennes sur k variables: quelles distributions de probabilité peut-on définir?

1. Loi uniforme sur $\mathcal{F}_k$
2. Dans un système logique pour lequel on définit un ensemble d'expressions, et on met une loi uniforme sur les expressions de taille donnée $\Rightarrow$ distribution de Catalan
3. Dans un système logique pour lequel on définit un ensemble d'expressions, et en faisant croître les arbres selon un processus de Galton-Watson
4. Autres possibilités?

Retour sur les arbres Et/Ou

Soient τ_g et τ_h deux arbres, calculant deux fonctions booléennes g et h

- l'arbre $(\wedge, \tau_g, \tau_h)$ calcule $g \wedge h$
- l'arbre $(\vee, \tau_g, \tau_h)$ calcule $g \vee h$

Retour sur les arbres Et/Ou

Soient τ_g et τ_h deux arbres, calculant deux fonctions booléennes g et h

- l'arbre $(\wedge, \tau_g, \tau_h)$ calcule $g \wedge h$
- l'arbre $(\vee, \tau_g, \tau_h)$ calcule $g \vee h$

Soit $\mathcal{E}_f$ l'ensemble des arbres calculant f

$$\mathcal{E}_f = \bigcup_{\ell \text{ litteral}} \{\ell\}.1_{\{f=\ell\}} \oplus \bigcup_{g,h:g\wedge h=f} (\wedge, \mathcal{E}_g, \mathcal{E}_h) \oplus \bigcup_{g,h:g\vee h=f} (\vee, \mathcal{E}_g, \mathcal{E}_h)$$

Retour sur les arbres Et/Ou

Soient τ_g et τ_h deux arbres, calculant deux fonctions booléennes g et h

- l'arbre $(\wedge, \tau_g, \tau_h)$ calcule $g \wedge h$
- l'arbre $(\vee, \tau_g, \tau_h)$ calcule $g \vee h$

Soit $\mathcal{E}_f$ l'ensemble des arbres calculant f

$$\mathcal{E}_f = \bigcup_{\ell \text{ litteral}} \{\ell\}.1_{\{f=\ell\}} \oplus \bigcup_{g,h:g\wedge h=f} (\wedge, \mathcal{E}_g, \mathcal{E}_h) \oplus \bigcup_{g,h:g\vee h=f} (\vee, \mathcal{E}_g, \mathcal{E}_h)$$

Sur les séries génératrices (T_f énumère $\mathcal{E}_f$)

$$T_f = 1_{f \text{ litteral}} + z \sum_{g,h:f=g\wedge h} T_g T_h + z \sum_{g,h:f=g\vee h} T_g T_h$$

Retour sur les arbres Et/Ou

On obtient ainsi un système de 2^{2^k} équations algébriques, de degré 2

On ne peut pas espérer résoudre explicitement...

Retour sur les arbres Et/Ou

On obtient ainsi un système de 2^{2^k} équations algébriques, de degré 2

On ne peut pas espérer résoudre explicitement...

... mais on peut connaître la singularité dominante commune à toutes les fonctions T_f, et son type

L'outil pour cela: théorème de Drmota-Lalley-Woods

Théorème de Drmota-Lalley-Woods

Famille d'arbres $\leftrightarrow$ langages algébriques $\leftrightarrow$ système d'équations algébriques.
Drmota 97, Lalley 93, Woods 97

Théorème de Drmota-Lalley-Woods

Système polynomial non linéaire

$$\{y_j = P_j(z, y_1, ..., y_m)\} \qquad 1 \leq j \leq m$$

- *a-positivité:* Termes des séries $P_j(\overrightarrow{y})$ tous ≥ 0
- *a-propre:* P est une contraction, i.e. satisfait une condition de Lipschitz ($K < 1$)

$$d(P(y_1, ..., y_m), P(y_1^{'}, ..., y_m^{'})) < K\, d((y_1, ..., y_m), (y_1^{'}, ..., y_m^{'}))$$

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

Théorème de Drmota-Lalley-Woods

Si les conditions précédentes sont vérifiées

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

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

- Toutes les autres singularités dominantes sont de la forme $\rho\,\omega$ avec ω racine de l'unité.
- 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

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

Les arbres Et/Ou et les fonctions de $\mathcal{F}_k$

- Fonction globale $T(z)$ énumérant les arbres
- Système algébrique sur les fonctions $T_f(z)$ (ou plutôt sur les m fonctions correspondant aux m classes d'équivalence)
- Le théorème de Drmota-Lalley-Woods s'applique
- $\exists$ une solution $(T_1, ..., T_m)$; les T_i (et donc les T_f) ont une singularité dominante commune $\rho < +\infty$, unique et strictement positive, également rayon de convergence de la série globale $T(z)$

$$\begin{aligned} T(z) &= \alpha - \beta\sqrt{1 - z/\rho} + O(1 - z/\rho); \\ T_f(z) &= \alpha_f - \beta_f\sqrt{1 - z/\rho} + O(1 - z/\rho). \end{aligned}$$

Modèle de Catalan

$$\begin{aligned} T(z) &= \alpha - \beta\sqrt{1 - z/\rho} + O(1 - z/\rho); \\ T_f(z) &= \alpha_f - \beta_f\sqrt{1 - z/\rho} + O(1 - z/\rho). \end{aligned}$$

Lemme de transfert (Flajolet-Odlyzko):

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

Theorem

Dans le modèle des arbres Et/ou, pour tout $f \in \mathcal{F}_k$,
$\lim_{n\to+\infty} \frac{[z^n]T_f(z)}{[z^n]T(z)}$ *existe et vaut*

$$P(f) := \frac{\beta_f}{\beta}$$

P est une loi de probabilité: distribution de Catalan

Modèle de branchement

Autre manière de construire un arbre?

- On part d'un noeud unique
- Il a 0 ou 2 fils avec même probabilité $1/2$
- Processus de Galton-Watson critique $\Rightarrow$ l'arbre est p.s. fini
- On étiquette uniformément et indépendamment les noeuds internes ($\wedge$ ou $\vee$) et les feuilles ($2k$ littéraux)

D'où une nouvelle distribution de probabilité sur $\mathcal{E}_k$, qui induit une nouvelle distribution π sur $\mathcal{F}_k$

π est la distribution de Galton-Watson

- $\tilde{\tau}$ arbre de Galton-Watson; taille $|\tilde{\tau}|$= nombre de noeuds internes

$$Proba(\tilde{\tau}) = \frac{1}{2^{2|\tilde{\tau}|+1}}.$$

- τ un des arbres obtenus par étiquetage de $\tilde{\tau}$; même taille

$$Proba(\tau) = Proba(\tilde{\tau}).\frac{1}{2^{|\tau|}}.\frac{1}{(2k)^{|\tau|+1}}.$$

- f fonction booléenne

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

$$\begin{aligned}
\pi(f) &= \sum_{\tau \text{ calcule } f} \mathit{Proba}(\tau) \\
&= \frac{1}{4k} \sum_{\tau \text{ calcule } f} \left(\frac{1}{16k}\right)^{|\tau|} \\
&= \frac{1}{4k} \sum_{n} \left(\frac{1}{16k}\right)^{n} [z^n]\phi_f(z)
\end{aligned}$$

Theorem

Dans le modèle des arbres Et/Ou, pour tout $f \in \mathcal{F}_k$

$$\pi(f) = \frac{\phi_f(1/16k)}{T(1/16k)} = \frac{\alpha_f}{\alpha}$$

Opérateurs commutatifs

- Ils conduisent à des arbres de Polyà
- Les fonctions $\phi_f(z)$ vérifient un système d'équations fonctionnelles
- On peut étendre le théorème de Drmota-Lalley-Woods
- Forme similaire de la solution

$$\phi_f(z) = \alpha_f + \beta_f \sqrt{z - \rho} + O(z - \rho)$$

- Analogue de la distribution de Catalan: limite de la loi obtenue en supposant tous les arbres de taille n équiprobables, lorsque $n \to +\infty$
- Comment définir l'extension de π? Pas de processus de branchement pour un arbre non planaire... mais distribution *de Bolzmann*

Liens entre probabilité et complexité

Complexité d'une fonction : taille des plus petits arbres la calculant

Exemple:

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

- 5 noeuds internes, 6 feuilles
- Calcule la fonction x
- Le plus petit arbre calculant cette fonction est donc réduit à une feuille

Liens entre probabilité et complexité

Theorem

Soit f de complexité $C(f)$. Sous des hypothèses générales, avec λ dépendant du système propositionnel

$$P(f) \sim \frac{\gamma_f}{(\lambda k)^{C(f)+1}}; \qquad \pi(f) \sim \frac{\delta_f}{(\lambda k)^{C(f)}}.$$

- Loi uniforme: effet Shannon

  Presque toutes les fonctions sont de complexité presque maximale

- Loi de Catalan ou de Galton-Watson: pas d'effet Shannon

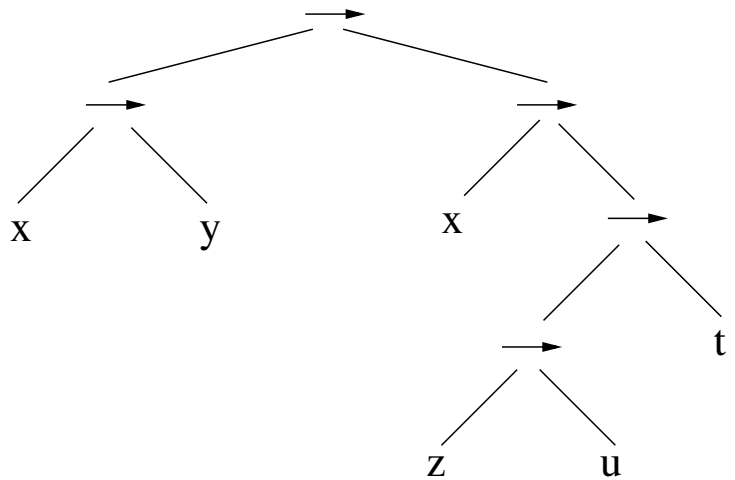
  Les fonctions de petite complexité sont les plus probables

Un système simple: l'implication

L'implication

- Connecteur unique $\rightarrow$
- Pas de littéraux négatifs

$$(x \rightarrow y) \rightarrow (x \rightarrow (z \rightarrow u) \rightarrow t)$$



L'implication

Pourquoi ce modèle?

- Simplicité: un seul connecteur, pas de négation, pas toutes les fonctions
  $\Rightarrow$ on peut espérer calculer la probabilité $P(f)$ d'une fonction f et étudier le lien avec sa complexité $C(f)$
- Logique intuitionniste
  Une tautologie $\sim$ une démonstration du but à partir des prémisses

L'implication

Quelles fonctions obtient-on?

Classe de Post $S_0 = \{f \in \mathcal{F}_k, f = x \vee g\}$

Combien de telles fonctions?

- Pour $k = 1$, 2 fonctions, *Vrai* et x
- Pour $k = 2$, 6 fonctions, *Vrai*, x, y, $x \to y$, $y \to x$, $x \vee y$
- Pour $k = 3$, 38 fonctions; pour $k = 4$, 942 fonctions
- Pour un alphabet de k variables, $Card(S_0) = \sum_{i=1}^{k} \binom{k}{i} (-1)^{i+1} 2^{2^k - 1}$
- E.I.S.: suite $A005530$

L'implication

Densité d'un sous-ensemble de formules

$\mathcal{I}_n$ ensemble des formules/arbres de taille n construites dans le système de l'implication

$\mathcal{I} = \cup_n \mathcal{I}_n$ ensemble de toutes les formules

Soit $E \subset \mathcal{I}$ et $E_n = E \cap \mathcal{I}_n$

$$\lim_{n \to +\infty} \frac{Card(E_n)}{Card(\mathcal{I}_n)} \ ?$$

Si cette limite exite, c'est la *densité* $\delta(E)$ de E dans $\mathcal{I}$

Logique intuitionniste (simplifiée)

Règles de calcul:

- Initial

$$\frac{}{A \vdash A}$$

- Introduction de $\rightarrow$

$$\frac{G, A \vdash B}{G \vdash (A \rightarrow B)}$$

- Elimination de $\rightarrow$ (Modus Ponens)

$$\frac{G \vdash A \qquad G \vdash (A \rightarrow B)}{G \vdash B}$$

Tautologies intuitionnistes

Une formule T est une tautologie intuitionniste
$\Leftrightarrow$ on peut trouver une preuve de T avec ces trois règles.

Tautologies intuitionnistes

- $A \to A$ est une tautologie intuitionniste
- $A \to (B \to A)$ est une tautologie intuitionniste
- $A_1 \to (A_2 \to (.... \to (A_p \to B)...))$ est une tautologie intuitionniste dès que $B \in \{A_1, ..., A_p\}$: tautologie *simple*
- $((A \to B) \to A) \to A$ est-elle une tautologie intuitionniste?

Tautologies intuitionnistes

- $\{$Taut. intuitionnistes (de $\mathcal{I}$)$\} \subset \{$Taut. classiques (de $\mathcal{I}$)$\}$
- Proportion des tautologies intuitionnistes qui sont "simples"?

Theorem

Asymptotiquement, lorsque le nombre k de variables booléennes tend vers l'infini, toute tautologie intuitionniste est simple

- Formule de Peirce: $((A \to B) \to A) \to A$
  - formule de $\mathcal{I}$ toujours vraie, i.e. tautologie
  - non démontrable en logique intuitionniste
- Proportion des taut. de $\mathcal{I}$, aussi taut. intuitionnistes?
  $\Leftrightarrow$ "densité" de la logique intuitionniste dans la logique classique?
  $\Leftrightarrow$ densité des formules "de Peirce"?

Tautologies intuitionnistes

On définit des classes de formules, incluses dans $\mathcal{I}$

- Tautologies simples
- Non-tautologies simples: les buts des prémisses diffèrent du but global
- Non-tautologies "fines"
- Tautologies "de Peirce"

Les densités des ensembles correspondants existent, et peuvent être calculées explicitement
[Fournier et al. 07, Genitrini et al. 08]

Logique intuitionniste vs. logique classique

Densité des tautologies intuitionnistes par rapport aux tautologies classiques?

$$\lambda := \frac{\delta(\mathit{Taut.\ intuit.})}{\delta(\mathit{Taut.})}$$

- Connecteur $\rightarrow$ seul: $\lambda = 1$
- On ajoute $\wedge$ ou 0: $\lambda = 1$
- On ajoute $\vee$: $\lambda = 5/8 < 1$

[Genitrini-Kozik]

L'implication

Theorem

Soit $f \in \mathcal{F}_k$, et soit $\mathcal{I}(f)$ l'ensemble des arbres qui calculent f. Alors la limite $\delta(\mathcal{I}(f))$ existe pour tout f; ceci définit une loi de probabilité P sur $\mathcal{F}_k$

Calcul de $P(f)$ pour $f \neq 1$

- Peut-on calculer la probabilité $P(f)$ de toute fonction f représentable par implication et (au plus) k variables?
- Peut-on relier $P(f)$ et $C(f)$?

L'implication

Quelques valeurs pour k petit

- $k = 1$: $P(1) = 0.72$, $P(x) = 0.28$.
- $k = 2$: $P(1) = 0.52$, $P(x) = 0.11$, $P(x \to y) = 0.10$, $P(x \vee y) = 0.06$.
- $k = 3$: $P(1) = 0.396$, $P(x) = 0.057$, $P(x \to y) = 0.033$, $P(x \vee y) = 0.013$, ...
- $k = 4$: $P(1) = 0.3$, $P(x) = 0.034$, $P(x \to y) = 0.014$, $P(x \vee y) = 0.004$, ...

L'implication

- Littéral x

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

- Fonction $x \to y$

$$\frac{9}{16k^3} + O\left(\frac{1}{k^4}\right)$$

- Fonction de complexité 2? 3? ...

L'implication

Theorem

Pour $f \in S_0 \setminus \{1\}$:

- *Il existe une constante γ_f (calculable, liée aux arbres minimaux de f) t.q.*

$$P(f) = \frac{\gamma(f)}{4^k C(f)^{k+1}} + O\left(\frac{1}{C(f)^{k+2}}\right)$$

- *Arbres calculant f : "simples" (obtenus p.s. par une expansion d'un arbre minimal)*

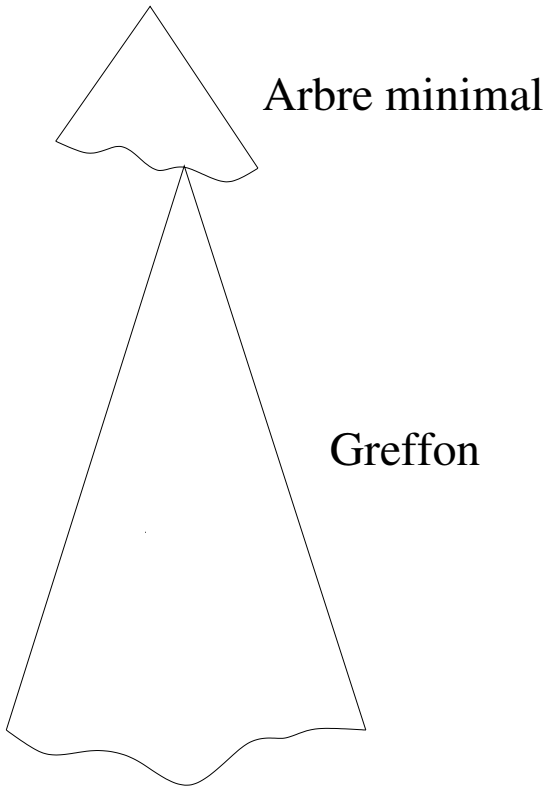
[Fournier et al. 08]

Idée de preuve

- On part d'un arbre minimal τ calculant f
- Expansion: dans τ, on remplace un sous-arbre $(\rightarrow, G, D)$ par $(\rightarrow, E, (\rightarrow, G, D))$ pour obtenir un arbre τ_E.

  L'expansion est *valide* ssi E est une tautologie, ou bien si (condition sur l'endroit de la greffe)
  - E a pour but α (littéral), τ_E calcule f, et tout arbre F de but α est tel que τ_F calcule f
  - *ou* E a une prémisse égale à β (littéral), τ_E calcule f, et tout arbre F avec une prémisse β est tel que τ_F calcule f
- On a donc à énumérer des familles d'arbres
- On montre que les arbres calculant f sont p.s.a. obtenus par une seule expansion valide d'un arbre minimal

Un arbre typique calculant f a la forme suivante



Autres systèmes propositionnels

- On part des arbres minimaux calculant une fonction donnée $f \in \mathcal{F}_k$
- Outil: langages de motifs [Kozik 2008]
- Un arbre typique calculant une fonction f (sous P ou π) est obtenu par une expansion d'un arbre minimal
- Le type précis d'expansion dépend du type d'arbres, et donc du système logique
- Exemple: Arbres avec connecteurs $\wedge$ et $\vee$ – ce qui importe est la structure des répétitions de variables

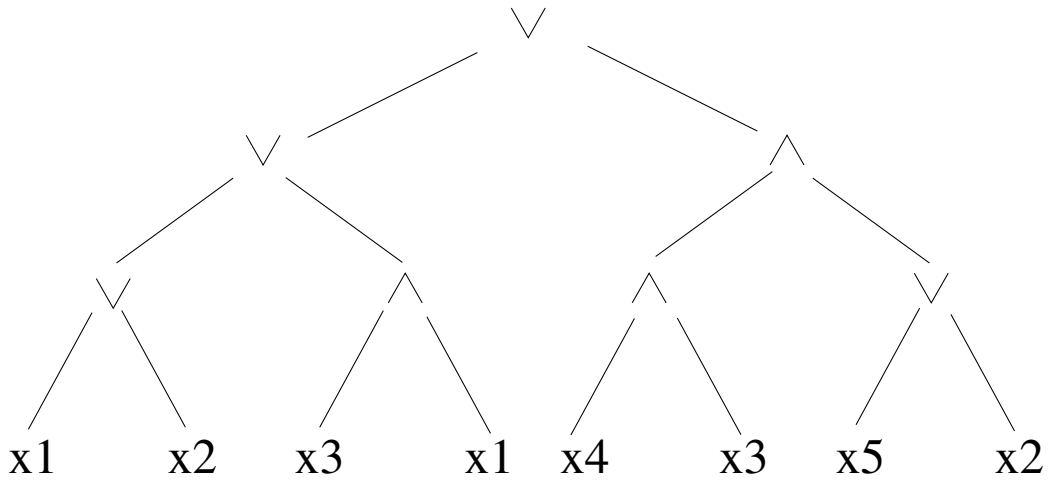
Arbres équilibrés

Et si on change le modèle d'arbre?

Peut-on obtenir une distribution dont le support n'est plus $\mathcal{F}_k$?

Arbres équilibrés

Toutes les feuilles sont au même niveau
Exemple: connecteurs $\vee$, $\wedge$ et littéraux



Processus de croissance

- Etape 0: on part d'une feuille
- A l'étape h, on prend deux arbres τ_g et τ_d obtenus à l'étape $h-1$ et on construit l'arbre $(\bullet, \tau_g, \tau_d)$
- A l'étape h, les arbres sont de hauteur h; toutes leurs feuilles sont au même niveau

Arbres équilibrés

Pour les définir à partir du processus de croissance précédent

- choisir un (ou des) connecteur(s) logique(s)
- choisir un ensemble H_0 d'étiquettes sur les feuilles (variables booléennes, littéraux, constantes) et une distribution de probabilité sur H_0

On obtient un ensemble H_h d'arbres de hauteur h
Loi uniforme sur H_h: induit une loi p_h sur $\mathcal{F}_k$

- Construction d'une (classe de) fonction(s) donnée?
- Etude de la probabilité limite sur $\mathcal{F}_k$?

Arbres équilibrés: connecteur aléatoire

Connecteur = $\vee$ ou $\wedge$ avec proba. $1/2$

$H_0 = \{x_1, ..., x_k\}$; distribution uniforme π_0 sur H_0

- Existence d'une distribution limite π sur $\mathcal{F}_k$
- Support $\{\phi_i, 1 \leq i \leq k\}$
  $\phi_i(x_1, ..., x_k) = 1$ ssi au moins i des variables booléennes sont vraies
- On peut caractériser π
- Vitesse de convergence accessible, et évaluation de la taille d'une expression calculant une des fonctions limites

Idée de preuve

- $\mathcal{E}_h$ ensemble des arbres de hauteur h, muni d'une loi de probabilité uniforme
- a affectation des k variables $x_1, \ldots x_k$
- Probabilité qu'une expression aléatoire e de $\mathcal{E}_h$ prenne la valeur 1 en a: $\omega(a)$ (poids de a)
- Deux affectations a et b; probabilité qu'un arbre de hauteur h donne la valeur 0 à a et 1 à b: $\omega(b) - \omega(a)$

Arbres équilibrés: connecteur aléatoire

On peut aussi avoir une distribution non uniforme sur H_0, ou des littéraux négatifs

Le support de la distribution limite est une extension de l'ensemble des fonctions seuil

Exemples

- $H_0 = \{x_1, ..., x_k\}$; $\pi_0 = \mathcal{U}(H_0)$
  $\Rightarrow \pi$ uniforme sur les n fonctions seuil
- $H_0 = \{0, 1, x_1, ..., x_k\}$; $\pi_0 = \mathcal{U}(H_0)$
  $\Rightarrow \pi$ uniforme sur les $k + 2$ fonctions seuil ou constantes
- $H_0 = \{x_1, x_2, x_3\}$; $\pi_0(x_i) = i/6$ $(1 \leq i \leq 3)$
  $\Rightarrow \pi$ uniforme sur 6 (classes de) fonctions: 1-seuil, $x_2 \vee x_3$, $x_3 \vee (x_1 \wedge x_2)$, $x_3 \wedge (x_1 \vee x_2)$, $x_2 \wedge x_3$, 3-seuil

Et si on prend encore un autre modèle d'arbre?

- Les arbres totalement équilibrés: distribution à support sur un ensemble de fonctions seuil

Et si on prend encore un autre modèle d'arbre?

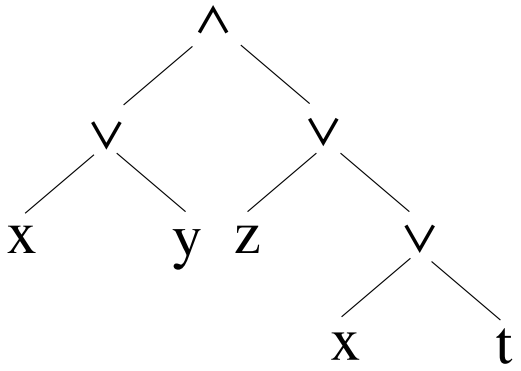
- Les arbres totalement équilibrés: distribution à support sur un ensemble de fonctions seuil
- Les arbres bourgeonnants

Et si on prend encore un autre modèle d'arbre?

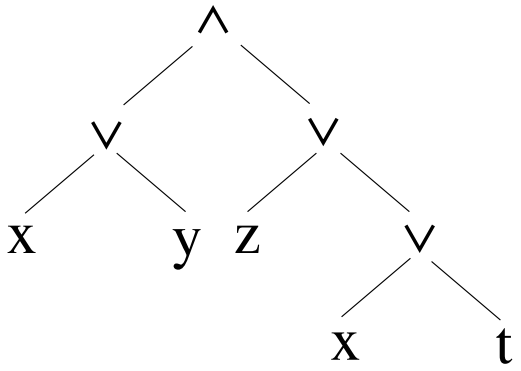
- Les arbres totalement équilibrés: distribution à support sur un ensemble de fonctions seuil
- Les arbres bourgeonnants
  Le support est encore plus restreint!
  Cf. le cours de Brigitte Chauvin

Des propositions aux prédicats...

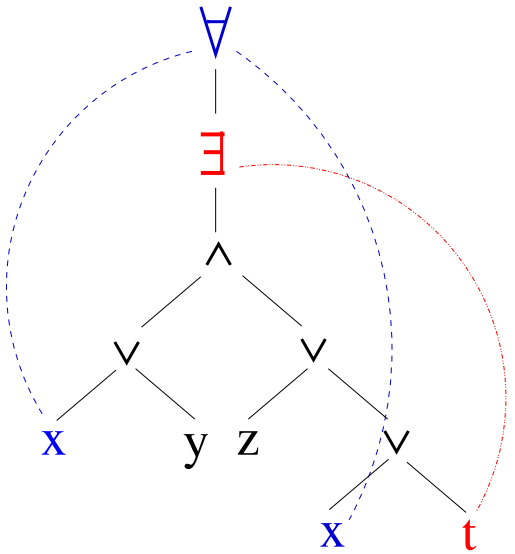
$(x \vee y) \wedge (z \vee (x \vee t))$



$(x \vee y) \wedge (z \vee (x \vee t))$



$\forall x \; \exists t \; (x \vee y) \wedge (z \vee (x \vee t))$



Enumeration de prédicats

On ne peut pas représenter les prédicats par des arbres!

La structure sous-jacente est bien un arbre...

Enumeration de prédicats

On ne peut pas représenter les prédicats par des arbres!

La structure sous-jacente est bien un arbre... mais on ajoute des liens entre quantificateurs et feuilles: on obtient des *arbres enrichis*

Enumeration de prédicats

On ne peut pas représenter les prédicats par des arbres!

La structure sous-jacente est bien un arbre... mais on ajoute des liens entre quantificateurs et feuilles: on obtient des *arbres enrichis*

On simplifie: *un seul* quantificateur, un seul type de noeud binaire. Ce sont les *lambda-termes*

Enumeration des lambda-termes sans variable libre et de taille donnée?

- Bornes (méthodes ad-hoc)
  - Si on ne compte que les noeuds internes [David et al. 00]

$$\left(\frac{(4-\epsilon)n}{\log n}\right)^{n(1-1/\log n)} \leq L_n \leq \left(\frac{(12+\epsilon)n}{\log n}\right)^{n(1-1/3\log n)}$$

  - Si on compte aussi les variables [Bodini-G-Gittenberger-Jacquot 13]

$$c_1 \left(\frac{4n}{e\log n}\right)^{n/2} \frac{\sqrt{\log n}}{n} \leq \lambda_n \leq c_2 \left(\frac{9(1+\epsilon)n}{e\log n}\right)^{n/2} \frac{(\log n)^{n/2\log n}}{n^{3/2}}$$

- Enumération asymptotique de quelques classes:
  - Nombre de quantificateurs borné sur un chemin
  - Chaque quantificateur lie le même nombre de variables