# Classes d'équivalences d'arbres booléens Application à un problème de satisfaisabilité.

Aléa 2014

Cécile Mailler
Prob–L@B – University of Bath

LATIN'14 – en collaboration avec Antoine Genitrini (LIP6)

20 mars 2014

# Qu'est ce qu'un problème SAT ?

$$\mathcal{E}(x_1, x_2, x_3) = [(x_1 \wedge \bar{x}_2) \vee x_1 \vee x_2] \wedge x_3$$

Cette expression booléenne est-est satisfaisable ?
i.e. existe-t-il $(\alpha_1, \alpha_2, \alpha_3) \in \{0, 1\}^3$ tel que $\mathcal{E}(\alpha_1, \alpha_2, \alpha_3) = 1$ ?

# Qu'est ce qu'un problème SAT ?

$$\mathcal{E}(x_1, x_2, x_3) = [(x_1 \wedge \bar{x}_2) \vee x_1 \vee x_2] \wedge x_3$$

Cette expression booléenne est-est satisfaisable ?
i.e. existe-t-il $(\alpha_1, \alpha_2, \alpha_3) \in \{0, 1\}^3$ tel que $\mathcal{E}(\alpha_1, \alpha_2, \alpha_3) = 1$ ?

Oui, il suffit de prendre
$\alpha_2 = 1$ et $\alpha_3 = 1$.

$\boxed{\text{Problème NP–complet}}$

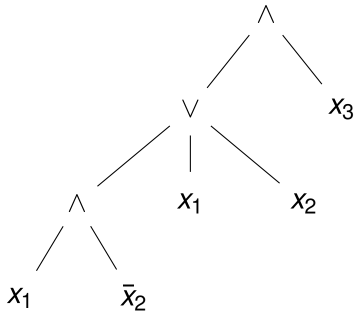
# Qu'est ce qu'un problème SAT ?

$$\mathcal{E}(x_1, x_2, x_3) = [(x_1 \wedge \bar{x}_2) \vee x_1 \vee x_2] \wedge x_3$$

Cette expression booléenne est-est satisfaisable ?
i.e. existe-t-il $(\alpha_1, \alpha_2, \alpha_3) \in \{0, 1\}^3$ tel que $\mathcal{E}(\alpha_1, \alpha_2, \alpha_3) = 1$ ?

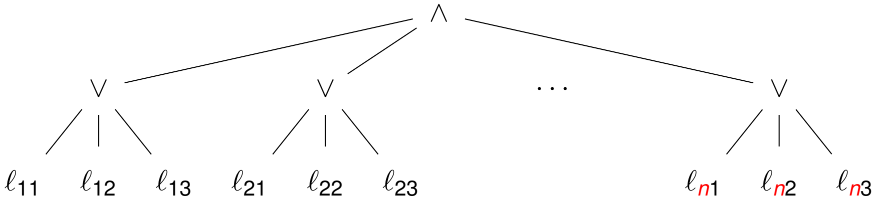
Oui, il suffit de prendre $\alpha_2 = 1$ et $\alpha_3 = 1$.

$\boxed{\text{Problème NP–complet}}$



# Problème 3–SAT

Dans le cadre 3–SAT, on se restreint aux expressions de la forme




où $\ell_{ij} \in \{x_1, \bar{x}_1 \ldots, x_k, \bar{x}_k\}$.

$\boxed{\text{Problème NP–complet}}$

# Cadre probabiliste

Tirons uniformément au hasard une expression de type 3–SAT à $n$ clauses et $k$ variables :

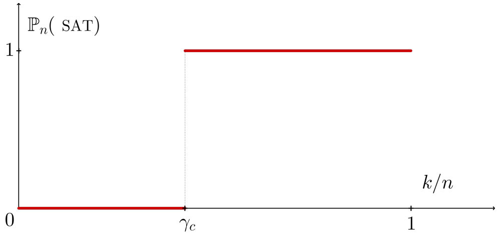
quelle est la probabilité qu'elle soit satisfaisable ?

# Cadre probabiliste

Tirons uniformément au hasard une expression de type 3–SAT à *n* clauses et *k* variables :

quelle est la probabilité qu'elle soit satisfaisable ?

## Conjecture




Asymptotiquement quand *n* tend vers $+\infty$.

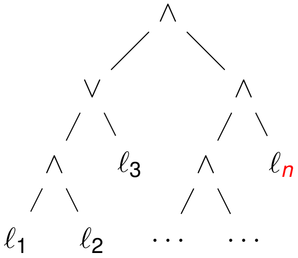
# Catalan–SAT

Notre idée : enlever la contrainte structurelle des formules 3–SAT.

Tirons uniformément au hasard un arbre (binaire) booléen à $n$ feuilles.

$\ell_1, \ldots, \ell_n \in \{x_1, \bar{x}_1, \ldots, x_k, \bar{x}_k\}$.

On prend $k = k_n$, fonction croissante de $n$ telle que $k_n \to +\infty$.




$$\boxed{\mathbb{P}_n(\mathrm{SAT}) = ?}$$

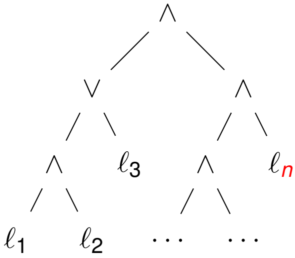
# Catalan–SAT

Notre idée : enlever la contrainte structurelle des formules 3–SAT.

Tirons uniformément au hasard un arbre (binaire) booléen à $n$ feuilles.

$\ell_1, \dots, \ell_n \in \{x_1, \bar{x}_1, \dots, x_k, \bar{x}_k\}$.

On prend $k = k_n$, fonction croissante de $n$ telle que $k_n \to +\infty$.




$\boxed{\mathbb{P}_n(\mathrm{SAT}) = ?}$

## Modèle ($\mathbb{F}$)

Pour toute $f : \{0,1\}^{k_n} \to \{0,1\}$,

$$\mathbb{P}_n(f) = \frac{\#\text{ arbres à } n \text{ feuilles et } k_n \text{ variables calculant } f}{\#\text{ arbres à } n \text{ feuilles et } k_n \text{ variables}} = ?$$

# Seconde introduction

Si $k$ est fixé, alors, pour toute $f : \{0,1\}^{k_0} \to \{0,1\}$,

$$\mathbb{P}_{n,k}(f) \xrightarrow{n\to+\infty} \mathbb{P}_k(f) = \underset{k\to+\infty}{\Theta}\left(\frac{1}{k^{L(f)+1}}\right),$$

où ***L(f)*** est la complexité d'une fonction booléenne, i.e. la taille des plus petits arbres la représentant. [LEFMANN-SAVICKÝ '97, KOZIK '09]

## Complexité

- fonctions constantes `Vrai` et `Faux` : $L(\texttt{Vrai}) = L(\texttt{Faux}) = 0$
- fonction XOR : $L(\text{XOR}) = 4$ $(x_1 \text{ XOR } x_2 = (x_1 \wedge \bar{x}_2) \vee (\bar{x}_1 \wedge x_2))$

# Seconde introduction

Si $k$ est fixé, alors, pour toute $f : \{0,1\}^{k_0} \to \{0,1\}$,

$$\mathbb{P}_{n,k}(f) \xrightarrow{n\to+\infty} \mathbb{P}_k(f) = \underset{k\to+\infty}{\Theta}\left(\frac{1}{k^{L(f)+1}}\right),$$

où ***L(f)*** est la complexité d'une fonction booléenne, i.e. la taille des plus petits arbres la représentant. [Lefmann-Savický '97, Kozik '09]

## Complexité

- fonctions constantes $\texttt{Vrai}$ et $\texttt{Faux}$ : $L(\texttt{Vrai}) = L(\texttt{Faux}) = 0$
- fonction XOR : $L(\text{XOR}) = 4$ $(x_1 \text{ XOR } x_2 = (x_1 \wedge \bar{x}_2) \vee (\bar{x}_1 \wedge x_2))$

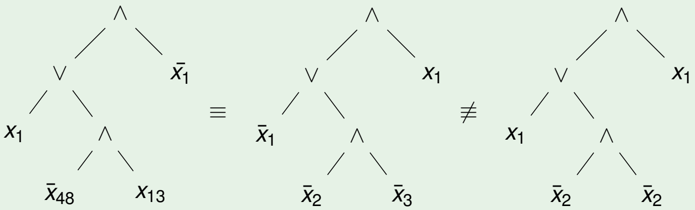
Faisons tendre $n$ et $k$ vers l'infini simultanément !

## Modèle ($\mathbb{F}$)

$$\mathbb{P}_n(f) = \frac{\#\text{ arbres à } n \text{ feuilles et } k_n \text{ variables calculant } f}{\#\text{ arbres à } n \text{ feuilles et } k_n \text{ variables}} \underset{n\to+\infty}{\sim} \quad ?$$

# Classes d'équivalence

## Exemples




Même relation d'équivalence pour les fonctions :
$\langle \boldsymbol{f} \rangle$ = classe d'équivalence de $f$.

$$\mathbb{P}_n\langle f\rangle = \frac{\begin{array}{c}\#\text{ de classes d'équivalence d'arbres de taille } n \text{ à } k_n \text{ variables}\\ \text{calculant une fonction de } \langle f\rangle\end{array}}{\#\text{ de classes d'équivalence d'arbres de taille } n \text{ à } k_n \text{ variables}}$$

# Nos deux modèles

## Modèle ($\mathbb{E}$)

$$\mathbb{P}_n\langle f\rangle \underset{n\to+\infty}{\sim} ?$$

## Modèle ($\mathbb{F}$)

$$\mathbb{P}_n(f) = \frac{\#\text{ arbres à } n \text{ feuilles et } k_n \text{ variables calculant } f}{\#\text{ arbres à } n \text{ feuilles et } k_n \text{ variables}} \underset{n\to+\infty}{\sim} ?$$

# Définitions

Quelques définitions avant le résultat :

- Une variable $x_i$ est **essentielle** pour $f$ si $f_{|x_i=1} \not\equiv f_{x_i=0}$.
  $\texttt{Ess}(f) = \#$ variables essentielles de $f$.
- **multiplicité** de $f$ = $\#$ répétitions dans un arbre minimal de $f$,

$$R(f) = L(f) - \texttt{Ess}(\texttt{f}).$$

## Exemples

- $R(\texttt{Vrai}) = R(\texttt{Faux}) = L(\texttt{Vrai}) - \texttt{Ess}(\texttt{Vrai}) = 0 - 0 = 0$
- $R(x_1 \vee x_2) = 2 - 2 = 0$
- $R(x_1 \text{ XOR } x_2) = 4 - 2 = 2$

# Résultat

## Modèle ($\mathbb{F}$)

Pour toute fonction $f : \{0,1\}^{k_0} \to \{0,1\}$,

$$\mathbb{P}_n(f) = \Theta\left(\left(\frac{1}{k_n}\right)^{L(f)+1}\right).$$

## Modèle ($\mathbb{E}$)

Pour toute fonction $f : \{0,1\}^{k_0} \to \{0,1\}$,

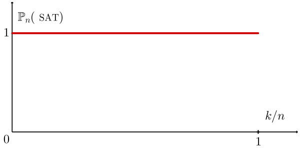
- Si $k_n \leq \frac{n}{\ln n}$, alors $\mathbb{P}_n\langle f\rangle = \Theta\left(\left(\frac{1}{k_n}\right)^{R(f)+1}\right)$ ;
- Si $k_n \geq \frac{n}{\ln n}$, alors $\mathbb{P}_n\langle f\rangle = \Theta\left(\left(\frac{\ln n}{n}\right)^{R(f)+1}\right)$.

# Résultat

## Modèle ($\mathtt{F}$)

Pour toute fonction $f : \{0,1\}^{k_0} \to \{0,1\}$,

$$\mathbb{P}_n(f) = \Theta\left(\left(\frac{1}{k_n}\right)^{L(f)+1}\right).$$




## Modèle ($\mathtt{E}$)

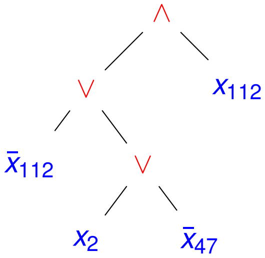
Pour toute fonction $f : \{0,1\}^{k_0} \to \{0,1\}$,

- Si $k_n \leq \frac{n}{\ln n}$, alors $\mathbb{P}_n\langle f\rangle = \Theta\left(\left(\frac{1}{k_n}\right)^{R(f)+1}\right)$ ;
- Si $k_n \geq \frac{n}{\ln n}$, alors $\mathbb{P}_n\langle f\rangle = \Theta\left(\left(\frac{\ln n}{n}\right)^{R(f)+1}\right)$.

# Idées de la preuve : Modèle ($\mathbb{E}$).

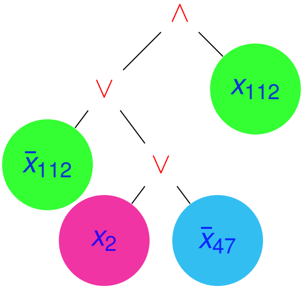
$\boldsymbol{A_n} = \#$ classes d'équivalence d'arbres à $n$ feuilles et $k_n$ variables

$\boldsymbol{A_n} = 2^{n-1} C_{n-1} B_{n,k_n}$



# Idées de la preuve : Modèle ($\mathbb{E}$).

$A_n = \#$ classes d'équivalence d'arbres à $n$ feuilles et $k_n$ variables
$A_n = 2^{n-1} C_{n-1} B_{n,k_n}$



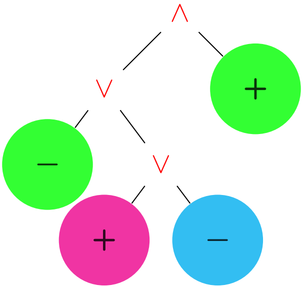

étiquetage : $B_{n,k_n} = \sum_{p=1}^{k_n} \left\{ {n \atop p} \right\} 2^{n-p}$

- $p$ est le nombre de parts de la partition

# Idées de la preuve : Modèle ($\mathbb{E}$).

$\boldsymbol{A_n} = \#$ classes d'équivalence d'arbres à $n$ feuilles et $k_n$ variables

$\boldsymbol{A_n} = 2^{n-1} C_{n-1} B_{n,k_n}$




étiquetage : $B_{n,k_n} = \sum_{p=1}^{k_n} \left\{ {n \atop p} \right\} 2^{n-p}$

- $p$ est le nombre de parts de la partition
- ensuite, chaque part est partitionnée en deux parts (dont une peut être vide)

# Idées de la preuve : Modèle $(\mathbb{E})$.

Donc $A_n = 2^{2n-1} C_{n-1} B_{n,k_n}$, où

$$B_{n,k_n} = \sum_{p=1}^{k_n} \left\{ {n \atop p} \right\} 2^{n-p}.$$

## Lemme technique

- Si $k_n \leq \frac{n}{\ln n}$, $\dfrac{B_{n-1,k_n}}{B_{n,k_n}} = \Theta\left(\dfrac{1}{k_n}\right)$ ;
- Si $k_n \geq \frac{n}{\ln n}$, $\dfrac{B_{n-1,k_n}}{B_{n,k_n}} = \Theta\left(\dfrac{\ln n}{n}\right)$.

**intuition au tableau**

# Idées de la preuve : Modèle ($\mathtt{E}$).

Donc $A_n = 2^{2n-1} C_{n-1} B_{n,k_n}$, où

$$B_{n,k_n} = \sum_{p=1}^{k_n} \left\{ {n \atop p} \right\} 2^{n-p}.$$

## Lemme technique

- Si $k_n \leq \frac{n}{\ln n}$, $\dfrac{B_{n-1,k_n}}{B_{n,k_n}} = \Theta\left(\dfrac{1}{k_n}\right)$ ;
- Si $k_n \geq \frac{n}{\ln n}$, $\dfrac{B_{n-1,k_n}}{B_{n,k_n}} = \Theta\left(\dfrac{\ln n}{n}\right)$.

**intuition au tableau**

N.B. Dans le modèle ($\mathtt{F}$),

$$\begin{aligned} A_n &= 2^{n-1} C_{n-1} (2k_n)^n \\ &= 2^{n-1} C_{n-1} B_{n,k_n} \end{aligned}$$

avec $B_{n,k_n} = (2k_n)^n$, et

$$\boxed{\frac{B_{n-1,k_n}}{B_{n,k_n}} = \frac{1}{2k_n}.}$$

# Reformulation du résultat

**Théorème – Modèle ($\mathbb{E}$) :**

Pour toute fonction $f\ :\ \{0,1\}^{k_0} \to \{0,1\}$,

$$\mathbb{P}_n\langle f\rangle = \Theta\left(\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)^{R(f)+1}\right).$$

**Théorème – Modèle ($\mathbb{F}$) :**

Pour toute fonction $f\ :\ \{0,1\}^{k_0} \to \{0,1\}$,

$$\mathbb{P}_n(f) = \Theta\left(\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)^{L(f)+1}\right).$$

# Tautologies

$\boxed{\text{Comprendre comment le lemme technique implique le résultat}}$

On va regarder le cas particulier $f \equiv \texttt{Vrai}$ :

**Lemme :**

Asymptotiquement quand $n$ tend vers $+\infty$, $\mathbb{P}_n\langle \texttt{Vrai} \rangle = \Theta\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)$.

Preuve en deux étapes :

- asymptotiquement quand $n$ tend vers l'infini, presque toute tautologie est *simple*
- la proportion des tautologies simples à $n$ feuilles et $k_n$ variables se comporte en $\Theta\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)$

# Tautologies

Comprendre comment le lemme technique implique le résultat

On va regarder le cas particulier $f \equiv \texttt{Vrai}$ :

**Lemme :**

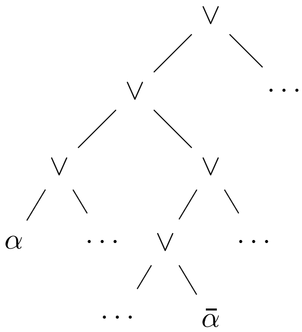
Asymptotiquement quand $n$ tend vers $+\infty$, $\mathbb{P}_n\langle\texttt{Vrai}\rangle = \Theta\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)$.

Preuve en deux étapes :

- asymptotiquement quand $n$ tend vers l'infini, presque toute tautologie est *simple* **[admis pour l'exposé]**
- la proportion des tautologies simples à $n$ feuilles et $k_n$ variables se comporte en $\Theta\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)$ **[on va montrer ça]**

# Tautologies

Comptons les tautologies simples :



# Tautologies simples

Comptons les tautologies simples :

$$S(z) = z^2 K''(A(z))$$

et

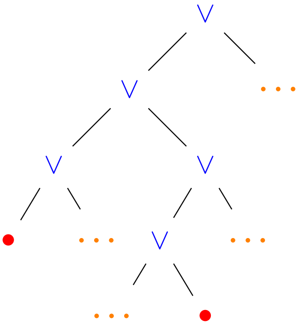
$$\boxed{TS_n = B_{n-1,k_n}\,[z^n]S(z)},$$

où

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

et

$$A(z) = \frac{1 - \sqrt{1-8z}}{4} + \frac{z}{2}.$$



# Tautologies simples

Donc,

$$\frac{TS_n}{A_n} = \frac{B_{n-1,k_n}\ [z^n]z^2K''(A(z))}{B_{n,k_n}\ [z^n]2A(z)},$$

car

$$A_n = 2^{n-1}C_{n-1}B_{n,k_n} = B_{n,k_n}\ [z^n]2A(z).$$

Donc, quand $n \to +\infty$,

$$\mathbb{P}_n\langle \texttt{Vrai}\rangle \sim \frac{TS_n}{A_n} = \frac{B_{n-1,k_n}}{B_{n,k_n}}\Theta(1).$$

# Tautologies simples

Donc,

$$\frac{TS_n}{A_n} = \frac{B_{n-1,k_n}\ [z^n]z^2K''(A(z))}{B_{n,k_n}\ [z^n]2A(z)},$$

car

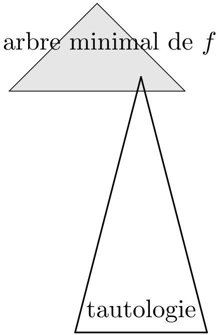
$$A_n = 2^{n-1}C_{n-1}B_{n,k_n} = B_{n,k_n}\ [z^n]2A(z).$$

Donc, quand $n \to +\infty$,

$$\mathbb{P}_n\langle \mathtt{Vrai} \rangle \sim \frac{TS_n}{A_n} = \frac{B_{n-1,k_n}}{B_{n,k_n}}\Theta(1).$$

*en vrai, on a multi-compté, mais on peut montrer que l'erreur est négligeable...*

# Cas général




On suit la stratégie de [KOZIK '09] :

- quand $n$ tend vers l'infini, presque tout arbre calculant $f$ est *simple*
- la proportion de *simple*-$f$ est donnée par

$$\Theta\left(\left(\frac{B_{n-1,k_n}}{B_{n,k_n}}\right)^{R(f)+1}\right),$$

dans le modèle ($\mathbb{E}$).

# Conclusion

On a réussi à comprendre :

- comment faire tendre $n$ et $k$ vers l'infini simultanément – Modèle ($\mathtt{F}$)

  $\rightarrow$ *pas de changement par rapport au modèle classique*

- comment le modèle change quand on introduit des classes d'équivalence – Modèle ($\mathtt{E}$)

  $\rightarrow$ *phénomène de saturation en* $k_n = \frac{n}{\ln n}$

# Conclusion

On a réussi à comprendre :

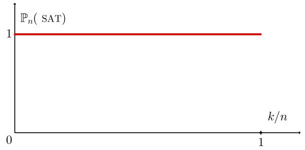
- comment faire tendre $n$ et $k$ vers l'infini simultanément – Modèle $(\mathtt{F})$

  $\rightarrow$ *pas de changement par rapport au modèle classique*

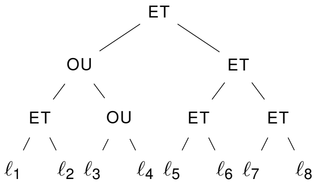
- comment le modèle change quand on introduit des classes d'équivalence – Modèle $(\mathtt{E})$

  $\rightarrow$ *phénomène de saturation en* $k_n = \frac{n}{\ln n}$

## Concernant le problème SAT :



# Ouvertures



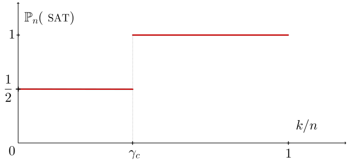

Changer la distribution sur les arbres : ABR aléatoire, ou arbres triangulaires (déterministes).

On conjecture une transition de phase dans ce modèle.

On sait que à $k$ fixé, les modèles d'arbres *saturés* vérifient

$$\mathbb{P}_{n,k}(\text{ SAT}) \xrightarrow{n\to\infty} \frac{1}{2}.$$

On espère



**Merci de votre attention !**