Arbres et/ou: une approche par limite locale

- Aléa 2016 -

Nicolas Broutin et Cécile Mailler

UNIVERSITY OF BATH

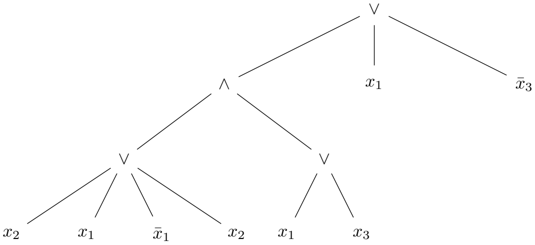
Arbres booléens et fonctions booléennes

Fonction booléenne = $f : \{0,1\}^{\mathbb{N}} \to \{0,1\}$

$\mathcal{F}_k$ = {Fonctions booléennes à k variables}

1 = Vrai
0 = Faux

Un arbre booléen représente une fonction booléenne.



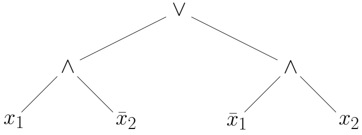
$$\begin{array}{ccc} \{0,1\}^k & \to & \{0,1\} \\ (x_1,\ldots,x_k) & \mapsto & 0 \end{array}$$

Taille d'un arbre = nombre de feuilles

Complexité

Complexité $L(f)$ = taille des plus petits arbres calculant f.

Fonctions constantes	$\texttt{Vrai} : (x_i)_{1..k} \mapsto 1$	$L(\texttt{Vrai}) = 0$
	$\texttt{Faux} : (x_i)_{1..k} \mapsto 0$	$L(\texttt{Faux}) = 0$
Littéraux	$x_m : (x_i)_{1..k} \mapsto x_m$	$L(x_m) = 1$
Negations	$\bar{x}_m : (x_i)_{1..k} \mapsto \bar{x}_m$	$L(\bar{x}_m) = 1$
XOR	$\text{XOR} : (x_i)_{1..k} \mapsto x_1 \text{ XOR } x_2$	$L(\text{XOR}) = 4$



Idée générale de l'exposé

- Tirer au hasard un arbre et/ou de taille n :
  - **tirer au hasard un arbre T_n non étiqueté de taille n** ;
  - l'étiqueter uniformément au hasard en arbre et/ou à k variables ;
- Noter $\mathbb{P}_{n,k}$ la loi de la fonction booléenne calculée par cet arbre et/ou ;
- Étudier $\mathbb{P}_{n,k}$, notamment quand $n \to +\infty$.

Comment $\mathbb{P}_{n,k}$ dépend de l'arbre aléatoire choisi ?

Arbre de Catalan

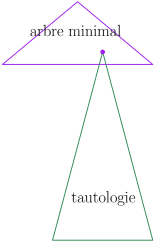
Si T_n est l'arbre de Catalan de taille n.

[Lefmann & Savický 1997] [Chauvin, Flajolet, Gardy & Gittenberger 2004] [Kozik 2008]

Théorème (prouvé par combinatoire analytique) :

- $\mathbb{P}_{n,k} \to \mathbb{P}_k$ quand $n \to \infty$;
- **Fixons** f telle que $L(f)$ ne dépend pas de k. Alors

$$\mathbb{P}_k(f) = \Theta\left(\frac{1}{k^{L(f)+1}}\right) \quad \text{quand } k \to \infty.$$



Idées de la preuve :

1. une tautologie typique est de la forme $x_m \vee \bar{x}_m \vee \cdots$ et donc $\mathbb{P}_k(\mathtt{Vrai}) = \Theta(1/k)$;
2. un arbre typique calculant f est de la forme arbre minimal + tautologie.

Arbre binaire de recherche aléatoire (ABR)

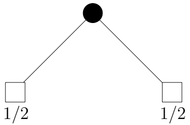
Si T_n est l'arbre binaire de recherche aléatoire à n feuilles.

[Chauvin, Gardy & Mailler 2011 & 2015]

Arbre binaire de recherche aléatoire (ABR)

Si T_n est l'arbre binaire de recherche aléatoire à n feuilles.

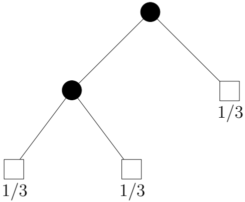
[Chauvin, Gardy & Mailler 2011 & 2015]



Arbre binaire de recherche aléatoire (ABR)

Si T_n est l'arbre binaire de recherche aléatoire à n feuilles.

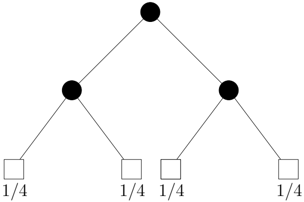
[Chauvin, Gardy & Mailler 2011 & 2015]



Arbre binaire de recherche aléatoire (ABR)

Si T_n est l'arbre binaire de recherche aléatoire à n feuilles.

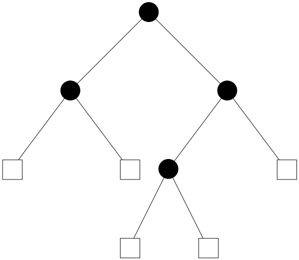
[Chauvin, Gardy & Mailler 2011 & 2015]



Arbre binaire de recherche aléatoire (ABR)

Si T_n est l'arbre binaire de recherche aléatoire à n feuilles.

[Chauvin, Gardy & Mailler 2011 & 2015]



Théorème :

Quand $n \to \infty$, $\mathbb{P}_{n,k}(\texttt{Vrai}) = \mathbb{P}_{n,k}(\texttt{Faux}) \to 1/2$.

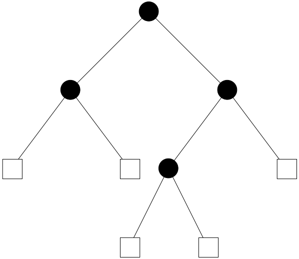
Prouvé par combinatoire analytique ou plongement en temps continu.

Comportement dégénéré.

Arbre binaire de recherche aléatoire (ABR)

Si T_n est l'arbre binaire de recherche aléatoire à n feuilles.

[Chauvin, Gardy & Mailler 2011 & 2015]



Théorème :

Quand $n \to \infty$, $\mathbb{P}_{n,k}(\texttt{Vrai}) = \mathbb{P}_{n,k}(\texttt{Faux}) \to {}^1\!/_2$.

Prouvé par combinatoire analytique ou plongement en temps continu.

Comportement dégénéré.

Quelle caractéristique de l'arbre de Catalan et de l'ABR explique cette différence de comportement ?

Catalan versus ABR

Catalan :

$$\mathbb{P}_k(f) = \Theta\left(\frac{1}{k^{L(f)+1}}\right)$$

ABR :

$$\lim_{n\to\infty} \mathbb{P}_{n,k}(\texttt{Vrai}) = \frac{1}{2}$$

	Catalan	ABR
Hauteur	$\sqrt{n}$	$\log n$
Saturation	$\texttt{cst}$	$\log n$

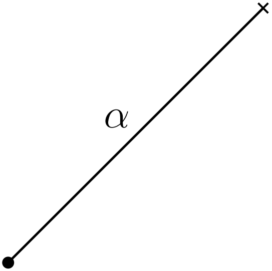
→ **propriété d'échelle**

→ **propriété locale**

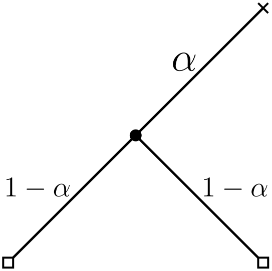
Objectifs :

- Comprendre laquelle des deux propriétés ci-dessus importe.
- Avoir un méta-théorème permettant de résoudre le cas d'autres arbres aléatoires.

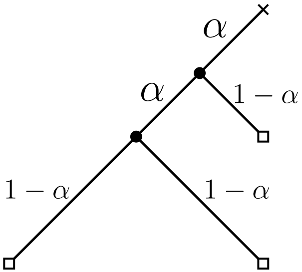
L’arbre de Ford



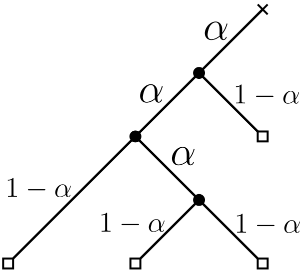
L'arbre de Ford



L'arbre de Ford



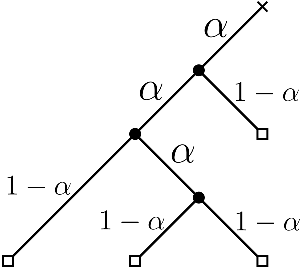
L'arbre de Ford



$$\alpha = \begin{cases} 0 & \text{ABR} \\ 1/2 & \text{Catalan} \\ 1 & \text{peigne} \end{cases}$$

	Catalan	ABR	$\alpha > 0$
Hauteur	$\sqrt{n}$	$\log n$	n^α
Saturation	cst	$\log n$	cst
$\lim_{n\to\infty} \mathbb{P}_{n,k}$	$1/k^{L(f)+1}$	dégénérée	? ?

L'arbre de Ford



$$\alpha = \begin{cases} 0 & \text{ABR} \\ 1/2 & \text{Catalan} \\ 1 & \text{peigne} \end{cases}$$

	Catalan	ABR	$\alpha > 0$
Hauteur	$\sqrt{n}$	$\log n$	n^α
Saturation	cst	$\log n$	cst
$\lim_{n\to\infty} \mathbb{P}_{n,k}$	$1/k^{L(f)+1}$	dégénérée	? ?

Comment se comporte $\mathbb{P}_{n,k}$ $(n \to \infty)$ dans ce modèle ?

Limite locale

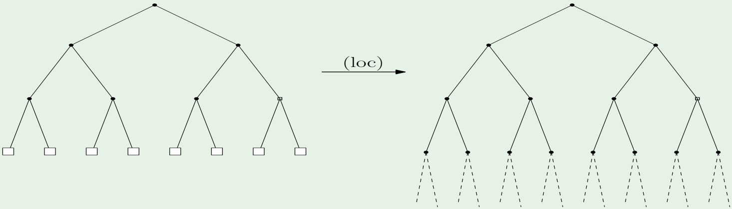
Notation : $t^{[h]}$ = l'arbre t tronqué à hauteur h.

Definition

Soit $(t_n)_{n\geq 1}$ une suite d'arbres, et t_∞ un arbre éventuellement infini.
On dit que t_n **converge localement** vers t_∞ si :

$$\forall h,\ \exists n(h)\ :\ \forall n \geq n(h),\ t_n^{[h]} = t_\infty^{[h]}.$$

Triangle à 2^n feuilles



Limite locale

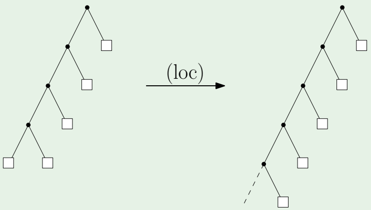
Notation : $t^{[h]}$ = l'arbre t tronqué à hauteur h.

Definition

Soit $(t_n)_{n\geq 1}$ une suite d'arbres, et t_∞ un arbre éventuellement infini.
On dit que t_n **converge localement** vers t_∞ si :

$$\forall h,\ \exists n(h)\ :\ \forall n \geq n(h),\ t_n^{[h]} = t_\infty^{[h]}.$$

Peigne à n feuilles



Le cas dégénéré

Théorème [BM++] :

Si $t_n \to t_\infty$ localement et si t_∞ n'a pas de feuilles, alors

$$\mathbb{P}_k[t_n](\texttt{Vrai}) = \mathbb{P}_k[t_n](\texttt{Faux}) \to 1/2 \quad (n \to \infty).$$

Nb : t_n et t_∞ sont déterministes.
Preuve : par récurrence sur le niveau de saturation.

Cas de l'ABR :

Si T_n est l'ABR de taille n, alors T_n converge presque sûrement localement vers l'arbre binaire complet infini.

Autres exemples : Arbre m-aire de recherche, arbre quadrant, etc.

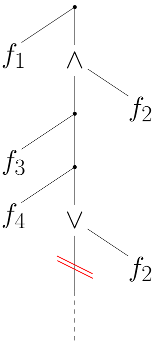
Cas non dégénéré

Théorème [BM++] :

Si $t_n \to t_\infty$ localement et si t_∞ a un nombre fini de branches infinies, alors $\mathbb{P}_k[t_n]$ converge vers une distribution limite $\mathbb{P}_k[t_\infty]$.

Idée de la preuve : principe des tiroirs
(ici : binaire et une branche infinie)

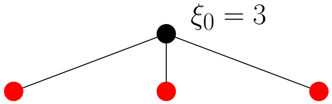
- $f^{[h]}$ = la fonction booléenne aléatoire calculée par $\hat{t}_\infty^{[h]}$
- presque sûrement, il existe $\sigma < +\infty$ (aléatoire) tel que $f^{[h]} = f^{[\sigma]}$ $(\forall h \geq \sigma)$
  (une fonction est répétée dans les $2^{2^k}+1$ premiers niveaux et avec proba 1/2 on peut couper)
- on définit $\mathbb{P}_k[t_\infty]$ comme la loi de $f^{[\sigma]}$



Catalan = Galton-Watson

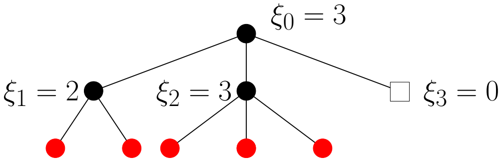
Les ξ's sont des variables aléatoires i.i.d.

Catalan = Galton-Watson



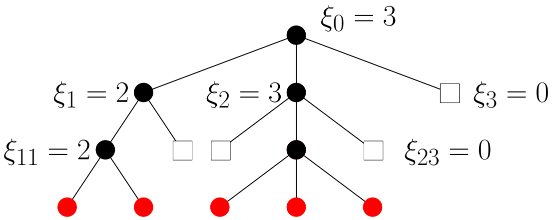
Les ξ's sont des variables aléatoires i.i.d.

Catalan = Galton-Watson



Les ξ's sont des variables aléatoires i.i.d.

Catalan = Galton-Watson



Les ξ's sont des variables aléatoires i.i.d.

Propriétés :

Si $\mathbb{E}\xi = 1$ on parle de GW critique et dans ce cas,

- $\mathtt{GW}_\xi$ est p.s. fini ;
- $\mathbb{E}\|\mathtt{GW}_\xi\| = +\infty$.

Théorème :

Si $\xi = 0$ ou 2 avec proba $1/2$, alors
$(\mathtt{GW}_\xi \,|\, \text{taille } n) = \mathtt{Cat}_n$ en loi.

Limite locale de l'arbre de Catalan

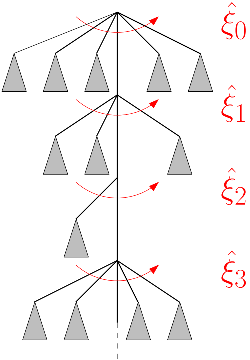
Théorème :

Si $\mathbb{E}\xi = 1$, alors $(\mathtt{GW}_\xi \,|\, \text{taille } n)$ converge localement en loi vers une branche infinie sur laquelle sont accrochées des copies i.i.d. de $\mathtt{GW}_\xi$.

$$\mathbb{P}(\hat{\xi} = i) = i\mathbb{P}(\xi = i)$$

Revenons aux fonctions booléennes

$\mathbb{P}_k[T_n]$ converge vers une distribution limite quand $n \to \infty$, dès que $T_n = \mathtt{Cat}_n$ ou n'importe quel GW critique conditionné à être de taille n.

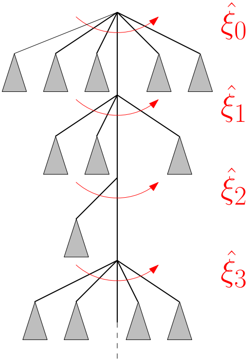


Limite locale de l'arbre de Catalan

Théorème :

Si $\mathbb{E}\xi = 1$, alors $(\mathtt{GW}_\xi \,|\, \text{taille } n)$ converge localement en loi vers une branche infinie sur laquelle sont accrochées des copies i.i.d. de $\mathtt{GW}_\xi$.

$\mathbb{P}(\hat{\xi} = i) = i\mathbb{P}(\xi = i)$



Revenons aux fonctions booléennes

$\mathbb{P}_k[T_n]$ converge vers une distribution limite quand $n \to \infty$, dès que $T_n = \mathtt{Cat}_n$ ou n'importe quel GW critique conditionné à être de taille n.

Que peut-on dire de plus sur $\mathbb{P}_k[T_\infty]$?

Cas *à la* Catalan

Théorème [BM++] :

Supposons que T_n converge localement en loi vers une limite T_∞ telle que T_∞ est une unique branche infinie sur laquelle sont attachées des forêts i.i.d. de loi A. Si, pour tout entier m,

$$\mathbb{E}\|A\|^m < +\infty \quad \text{alors} \quad \mathbb{P}_k[T_\infty] = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

Cas *à la* Catalan

Théorème [BM++] :

Supposons que T_n converge localement en loi vers une limite T_∞ telle que T_∞ est une unique branche infinie sur laquelle sont attachées des forêts i.i.d. de loi A. Si, pour tout entier m,

$$\mathbb{E}\|A\|^m < +\infty \quad \text{alors} \quad \mathbb{P}_k[T_\infty] = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

Rappel : Si $T_n = \mathtt{Cat}_n$, alors $A = \mathtt{GW}_\xi$ critique et $\mathbb{E}\|A\| = +\infty$...

Cas *à la* Catalan

Théorème [BM++] :

Supposons que T_n converge localement en loi vers une limite T_∞ telle que T_∞ est une unique branche infinie sur laquelle sont attachées des forêts i.i.d. de loi A. Si, pour tout entier m,

$$\mathbb{E}\|A\|^m < +\infty \quad \text{alors} \quad \mathbb{P}_k[T_\infty] = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

Rappel : Si $T_n = \mathtt{Cat}_n$, alors $A = \mathtt{GW}_\xi$ critique et $\mathbb{E}\|A\| = +\infty$... **>_<**

Cas *à la* Catalan

Théorème [BM++] :

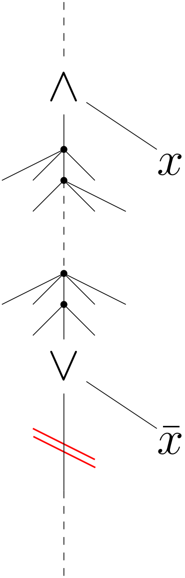
Supposons que T_n converge localement en loi vers une limite T_∞ telle que T_∞ est une unique branche infinie sur laquelle sont attachées des forêts i.i.d. de loi A. Si, pour tout entier m,

$$\mathbb{E}\|\mathbf{trim}(A)\|^m < +\infty \quad \text{alors} \quad \mathbb{P}_k[T_\infty] = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

Rappel : Si $T_n = \mathtt{Cat}_n$, alors $A = \mathtt{GW}_\xi$ critique et $\mathbb{E}\|A\| = +\infty$... **>_<**

Procédure de trim [≈ CFGG'04] :

Dès qu'on a le motif de droite, "on coupe".
Arbre booléen b trimé = $\mathtt{trim}(b)$.



Cas *à la* Catalan

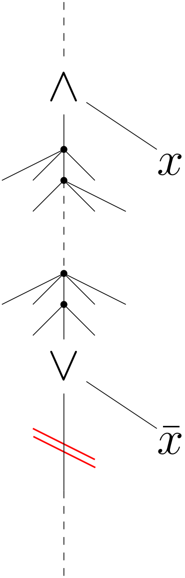
Théorème [BM++] :

Supposons que T_n converge localement en loi vers une limite T_∞ telle que T_∞ est une unique branche infinie sur laquelle sont attachées des forêts i.i.d. de loi A.

$$\text{Si } \mathbb{E}\|\mathtt{trim}(A)\| < +\infty \text{ alors } \mathbb{P}_k[T_\infty] = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

Applications :

- T_n = n'importe quel GW critique conditionné à être de taille n (dès que $\exists\varepsilon > 0 : \mathbb{E}e^{\varepsilon\xi} < +\infty$) ;
- T_n = arbre de Ford de taille n ($\forall\alpha > 0$)



Résumé des résultats

Cas dégénéré

Si $t_n \to t_\infty$ localement et si t_∞ n'a pas de feuilles, alors

$$\mathbb{P}_k[t_n](\texttt{Vrai}) = \mathbb{P}_k[t_n](\texttt{Faux}) \to 1/2 \quad (n \to \infty).$$

Cas non dégénéré – convergence

Si $t_n \to t_\infty$ localement et si t_∞ a un nombre fini de branches infinies, alors $\mathbb{P}_k[t_n]$ converge vers une distribution limite $\mathbb{P}_k[t_\infty]$.

Cas non dégénéré – comportement

Supposons que T_n converge localement en loi vers une limite T_∞ telle que T_∞ est une unique branche infinie sur laquelle sont attachées des forêts i.i.d. de loi A. Si, pour tout entier m,

$$\mathbb{E}\|\texttt{trim}(A)\|^m < +\infty \quad \text{alors} \quad \mathbb{P}_k[T_\infty] = \Theta\left(\frac{1}{k^{L(f)+1}}\right).$$

Remarques finales

Nouvelle méthode “universelle” par limite locale :

- résout des modèles déjà connus (ABR, Catalan, etc)
- mais aussi des modèles non étudiés (arbre m-aire de recherche, **arbre de Ford**).

Cette approche a des limites :

- échoue complètement pour des arbres booléens non plans (pourtant naturels puisque $\wedge$ et $\vee$ sont commutatifs) ;
- les hypothèses du théorème Θ sont *un peu* délicates à vérifier ;
- on obtient seulement un Θ alors que les preuves au cas pas cas par combinatoire analytique donnent un équivalent.

Remarques finales

Nouvelle méthode “universelle” par limite locale :

- résout des modèles déjà connus (ABR, Catalan, etc)
- mais aussi des modèles non étudiés (arbre *m*-aire de recherche, **arbre de Ford**).

Cette approche a des limites :

- échoue complètement pour des arbres booléens non plans (pourtant naturels puisque $\wedge$ et $\vee$ sont commutatifs) ;
- les hypothèses du théorème Θ sont *un peu* délicates à vérifier ;
- on obtient seulement un Θ alors que les preuves au cas pas cas par combinatoire analytique donnent un équivalent.

Merci ! !