Distributions de valuations sur les arbres

Michel Nguyen-The, Jean-Marc Steyaert

$\{$**mnguyen,steyaert**$\}$**@lix.polytechnique.fr**

Plan

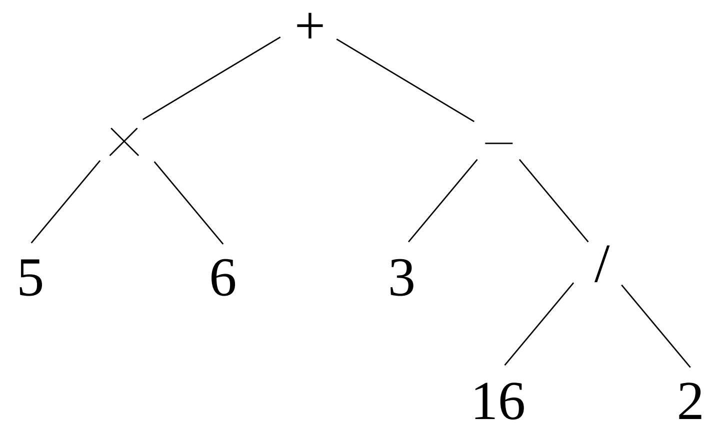
- Définitions
- Arbres $\{+, -\}$
- Arbres $\{-\}$ et feuille 1
- Bijections et nombres de Narayana
- Distributions limites
- Arbres $\{+, \times\}$ (ébauche)

Arbre d'expression arithmétique

Définition et notations : On se donne donne un jeu d'opérateurs Op, et un jeu de nombres entiers I. On considère un arbre planaire : on étiquette les feuilles par des éléments de I, et les nœuds internes par des opérateurs de même arité que le nœud correspondant. On obtient ainsi un *arbre d'expression arithmétique*.

Exemple :

$Op = \{+, -, \times, /\}$ et $I = \mathbb{N}_{20}$

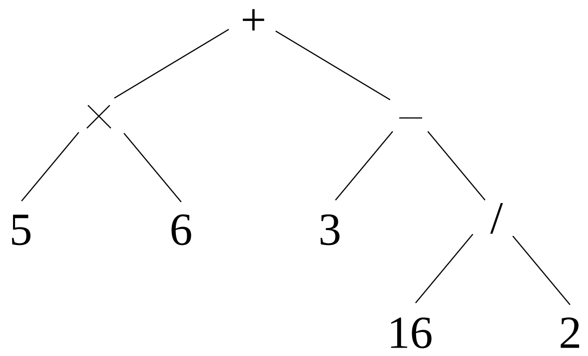


associé à l'expression $(5 \times 6) + (3 - (16/2))$.

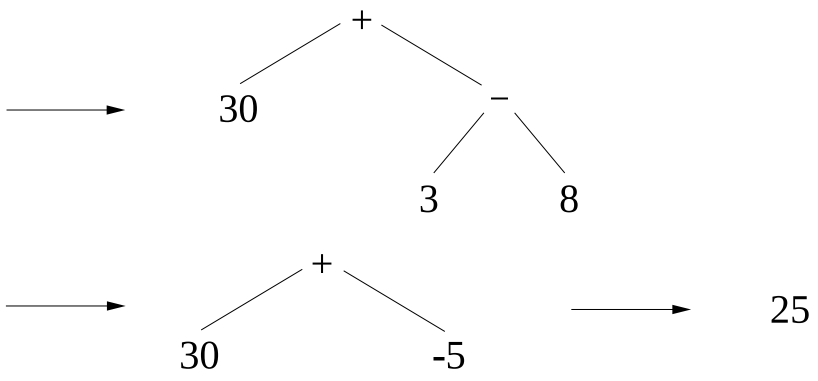
Définition d'une valuation

Définitions et notations : Soit un arbre d'expression arithmétique. La *valuation* de l'arbre est le résultat de cette expression.

Exemple :



associé à l'expression $(5 \times 6) + (3 - (16/2))$, se réécrit successivement en :



La valuation de l'arbre ci-dessus vaut donc 25.

Fins et moyens

But : Étudier la distribution (limite) des valuations, vues comme variables aléatoires X_n, dans l'espace des arbres binaires d'une taille donnée (nombres de nœuds internes).

Outils : Fonction génératrice

$W(z,u) = \sum_{n,k} w_{nk} z^n u^k$

w_{nk} : # arbres binaires étiquetés de taille n et de valuation k

$I \in \mathbb{Z}$ ensemble fini des entiers à placer aux feuilles

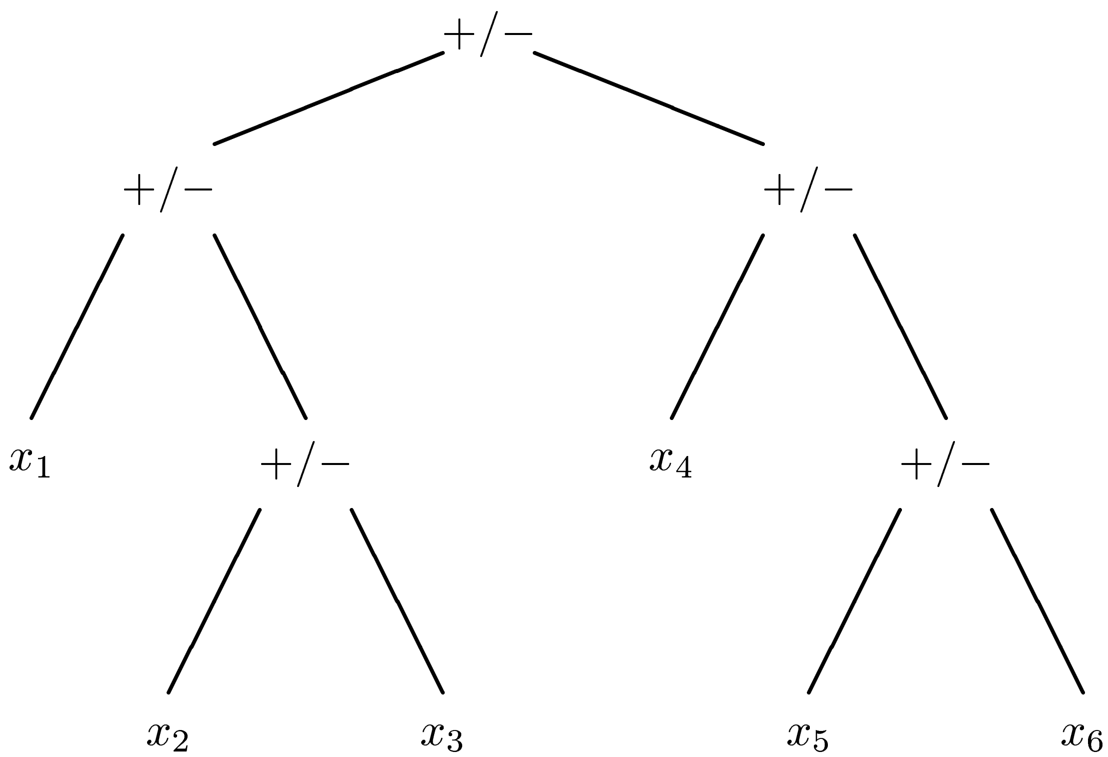
$F = \sum_{i \in I} u^i$: fonction génératrice des feuilles

X_n : variable aléatoire égale à la valuation d'un arbre de taille n

$$P(X_n = k) = \frac{w_{nk}}{C_n |Op|^n |I|^{n+1}}$$

(C_n $n^{\text{ème}}$ nombre de Catalan)

Arbres avec $Op = \{+, -\}$ et $I = \{1\}$

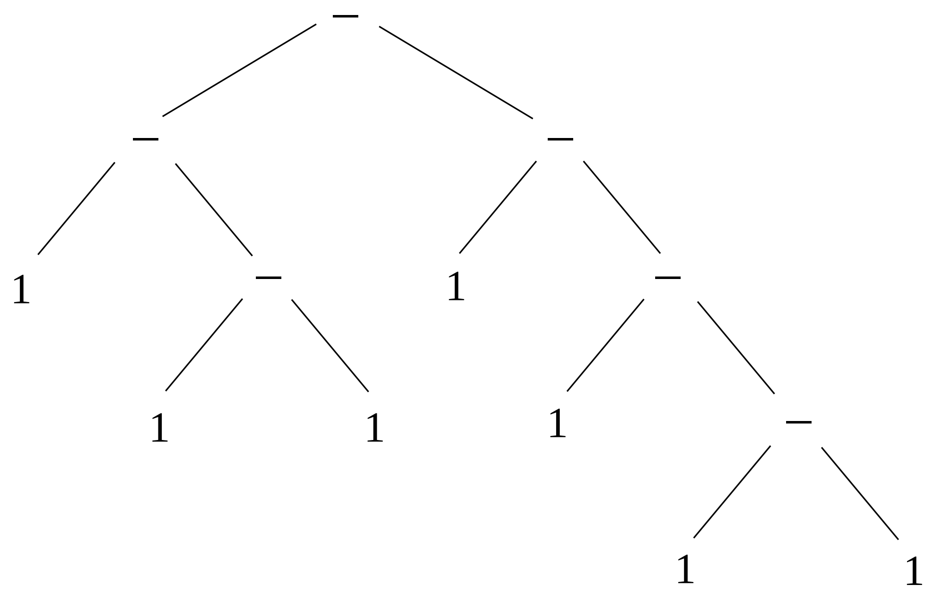


Définition et notations : $x_1 \pm x_2 \pm x_3 \pm x_4 \pm x_5 \pm x_6$

La valuation a même distribution que $F(F + \overline{F})^n$, $i.e.$ un jeu de pile ou face.

$\Rightarrow$ Limite gaussienne

Arbres avec $Op = \{-\}$ et $I = \{1\}$

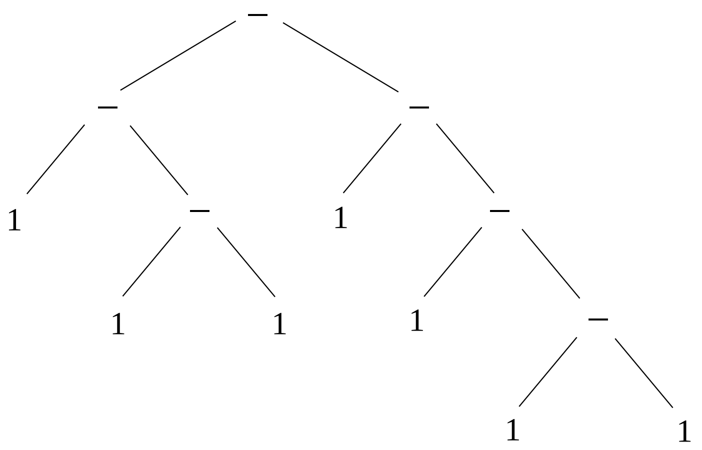


Fait : La valuation a même distribution que la différence entre le nombre de feuilles gauches et le nombre de feuilles droites.

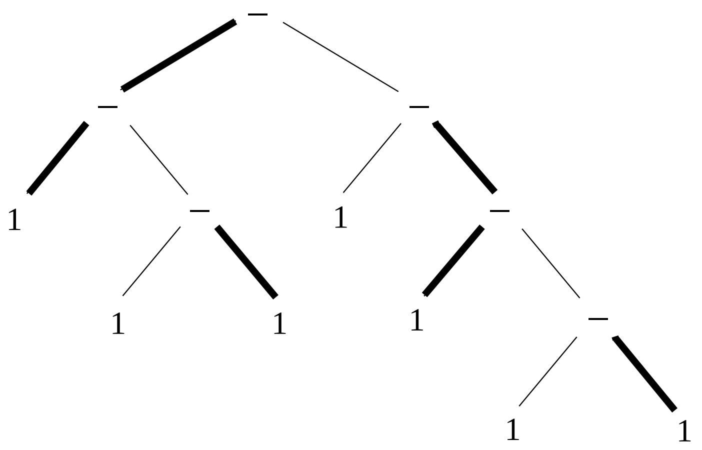
Arêtes positives et négatives

Définition : On appelle *arête positive* une arête telle que la valuation du sous-arbre remonte positivement à la racine. Les arêtes non positives sont dites *négatives*.

Exemple : Considérons l'arbre ci-dessous :



On fonce en noir les arêtes positives :

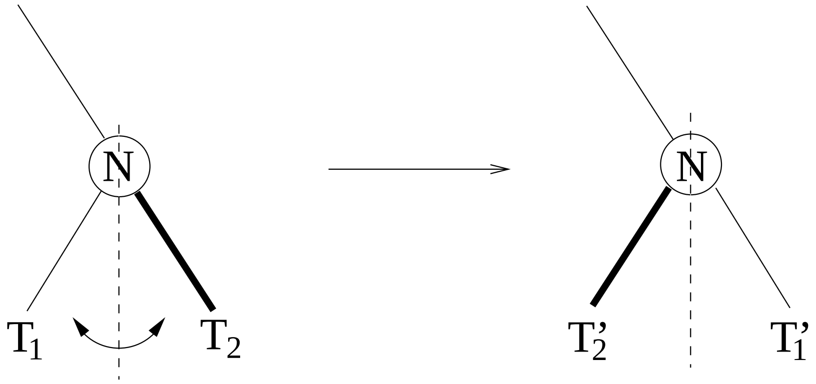


Algorithme de la bijection

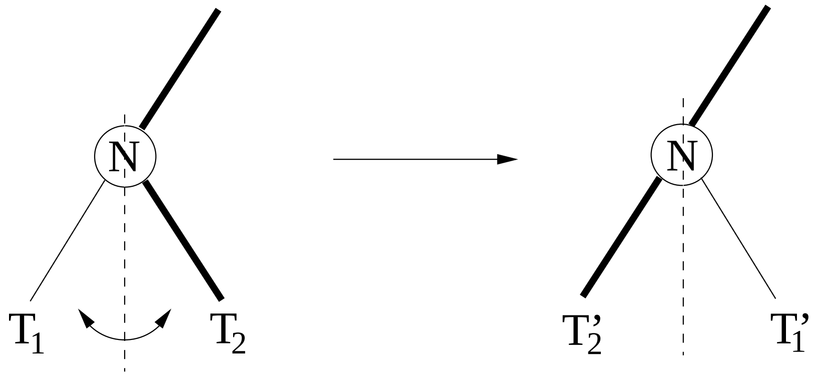
Pour passer d'un arbre étiqueté par des $-$ et des 1 de valuation k, à un arbre binaire tel que : (#feuilles gauches $-$ #feuilles droites)=k, on suit l'algorithme suivant :

Principe de l'algorithme : Lors de l'algorithme, on transforme les arêtes positives en arêtes gauches.

Algorithme : On parcourt l'arbre dans l'ordre préfixe. Chaque fois qu'on arrive à un nœud N dont l'arête descendante droite est positive, on effectue une symétrie axiale sur par rapport au sous-arbre de racine N :



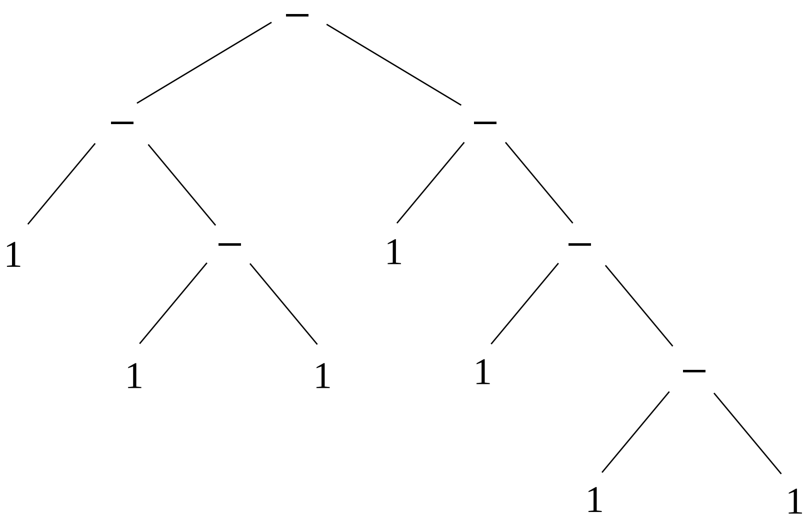
ou



avec T_1' symétrique de T_1 et T_2' symétrique de T_2.

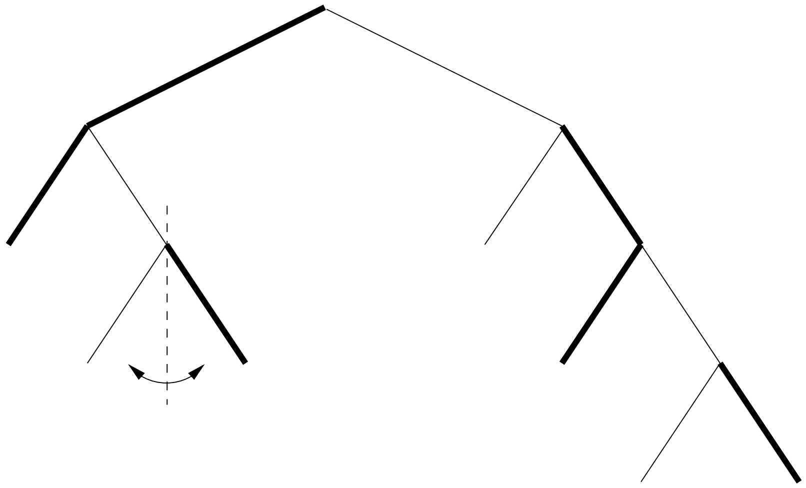
Exemple d’exécution de l’algorithme (1)

On part de l’exemple d’arbre précédent :



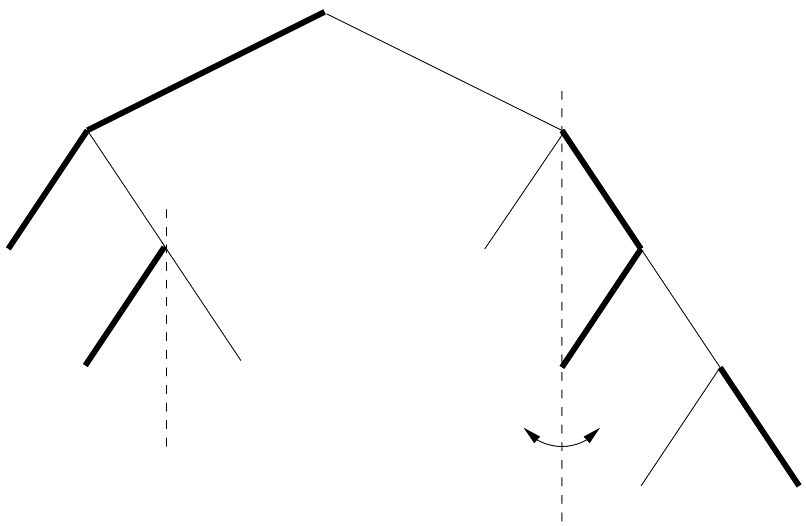
colorié en :

Étape 0 de l’algorithme :

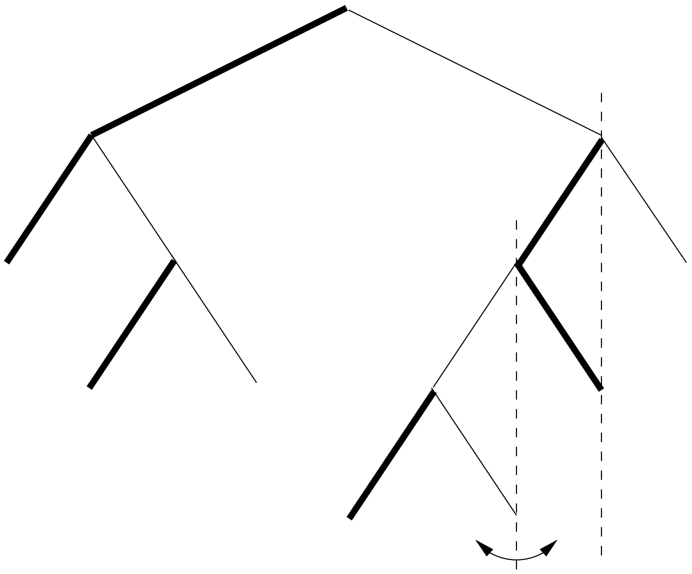


Exemple d'exécution de l'algorithme (2)

Étape 1 :

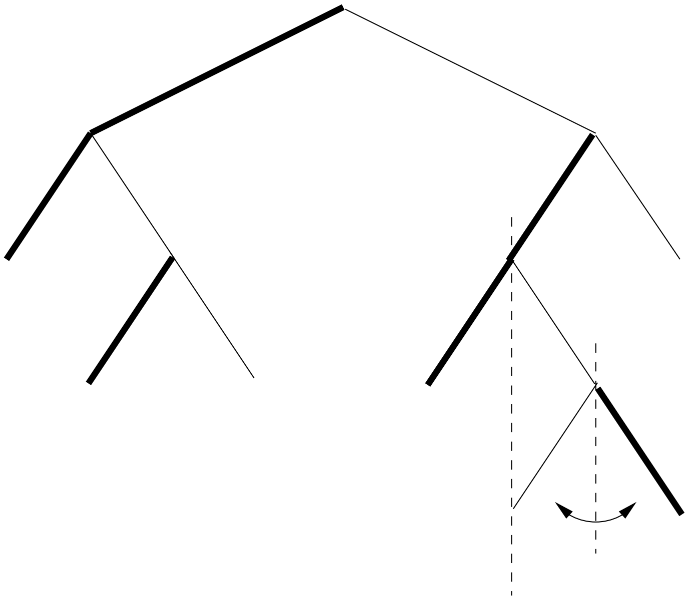


Étape 2 :

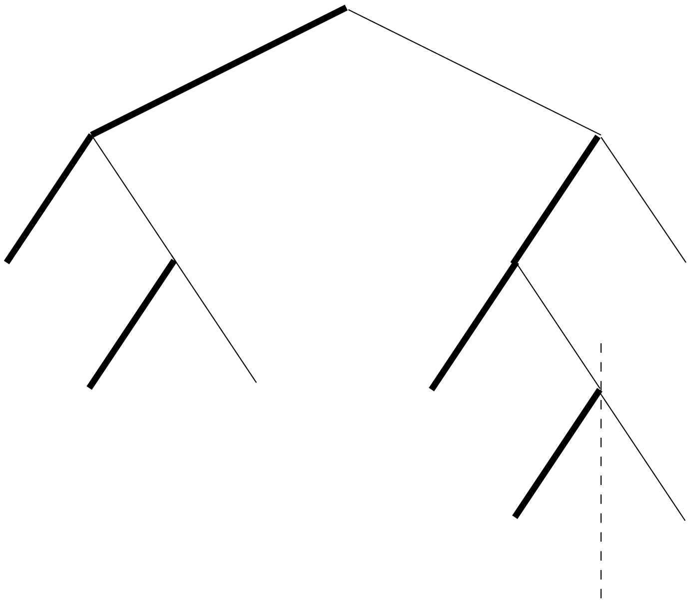


Exemple d'exécution de l'algorithme (3)

Étape 3 :



Étape 4 :



Changement de variables

Coefficients de $W(z,u)$:

n\k	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
0							1					
1						1						
2					1		1					
3				1		3		1				
4			1		6		6		1			
5		1		10		20		10		1		
6	1		15		50		50		15		1	
7		21		105		175		105		21		1

Coefficients de $\sqrt{u}W(\sqrt{u}z,\sqrt{u})-u$:

n\k	1	2	3	4	5	6	7	8	9	
1	1									
2	1	1								
3	1	3	1							
4	1	6	6	1						
5	1	10	20	10	1					
6	1	15	50	50	15	1				
7	1	21	105	175	105	21	1			
8	1	28	196	490	490	196	28	1		
9	1	36	336	1176	1764	1176	336	36	1	

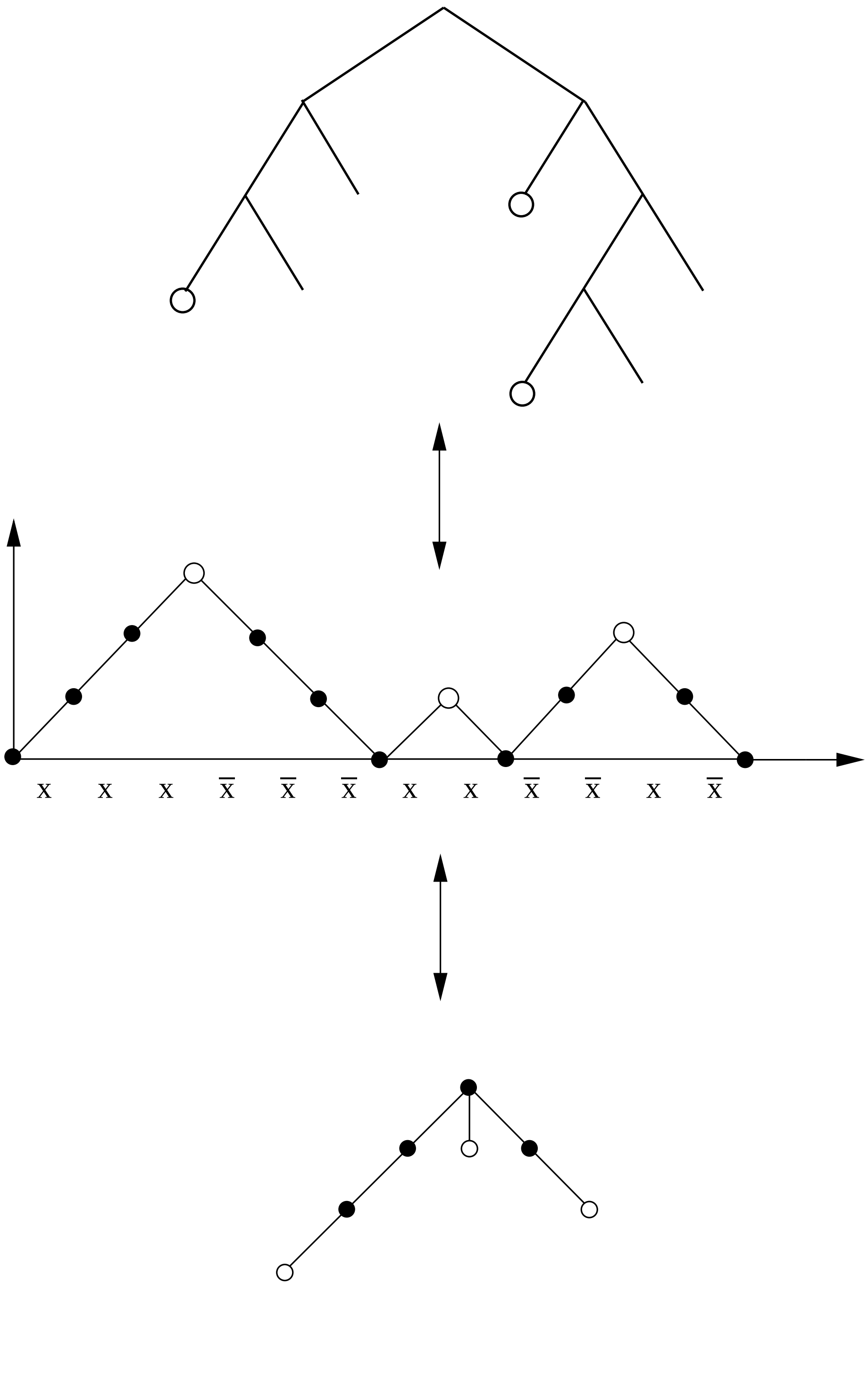
Nombres de Narayana

Les nouveaux coefficients sont égaux à $\frac{1}{n}\binom{n}{k}\binom{n}{k-1}$, nombres de Narayana apparaissant souvent dans la littérature combinatoire.

Ils sont connus pour dénombrer (entre autres) :

- le nombre d'arbres binaires de taille n à k feuilles gauches (respectivement droites) ;
- le nombre de chemins de Dyck de longueur $2n$ à k pics ;
- le nombre d'arbres généraux à $n+1$ nœuds dont k feuilles.

Bijections classiques



Distribution limite

Méthode des moments factoriels (Kreweras)

Notons la factorielle tombante :

$(n)_k = n \times (n-1) \times \cdots * (n-k+1)$. Le moment factoriel d'ordre m de la variable aléatoire S_n valant k avec la probabilité μ_{nk}/C_n s'écrit :

$$
\begin{aligned}
& \frac{1}{C_n} \sum_{k=1}^{n} (k)_m \frac{1}{n} \binom{n}{k} \binom{n}{k-1} \\
= \quad & \frac{m!}{nC_n} \sum_{k=1}^{n} \binom{k}{m} \binom{n}{k} \binom{n}{k-1} \\
= \quad & \frac{m!}{nC_n} \binom{n}{m} \sum_{k=1}^{n} \binom{n-m}{n-k} \binom{n}{k-1} \\
= \quad & \frac{m!}{nC_n} \binom{n}{m} \binom{2n-m}{n-1} \\
= \quad & \boxed{\frac{(n+1)_m n_m}{(2n)_m}}
\end{aligned}
$$

$$
\mu_n = \frac{n+1}{2} \quad \text{et} \quad \sigma_n^2 = \frac{(n+1)(n-1)}{4(2n-1)}
$$

$\Rightarrow$ Limite gaussienne

Distribution limite

On note :

$$\mu_n(u) = [z^n]T = \sum_{k=0}^{n} \frac{1}{n} \binom{n}{k} \binom{n}{k-1} u^k.$$

Théorème des quasi puissances (Hwang)

$$\mu_n(u) = \frac{-2}{\sqrt{\pi}} B(\rho(\sqrt{u}), \sqrt{u}) (\rho(\sqrt{u}))^n n^{-3/2} \left(1 + O\left(\frac{1}{n}\right)\right)$$

(calcul de Flajolet et Sedgewick)

où $B(z,u) = \frac{1}{2}\sqrt{1 - z(1-u)^2}$

et $\rho(u) = \dfrac{1}{(1+u)^2}$.

$\Rightarrow$ Limite gaussienne

Distribution limite

Généralisation de cette méthode pour I quelconque ?

Notant $\overline{F} = F(u^{-1})$ on a :

$$\begin{aligned} T &= W - F \\ &= z(F+T)(\overline{F}+\overline{T}) \end{aligned}$$

avec $T = \overline{T}$ d'où par la formule d'inversion de Lagrange :

$$\begin{aligned} [z^n]T &= \frac{1}{n}[t^{n-1}](F+t)^n(\overline{F}+t)^n \\ &= \cdots \\ &= \overline{F}^{n+1}\mu_n\left(\frac{F}{\overline{F}}\right) \end{aligned}$$

Sommes-nous dans les hypothèses du théorème des quasi puissances ?

Limite gaussienne ?

(attention à la période de $\frac{F}{\overline{F}}$)

Arbres $Op = \{+, \times\}$

Fonction génératrice :

On note $v(T)$ la valuation de l'arbre T. La fonction génératrice bivariée comptant les arbres par taille et valuation, vérifie :

$$\begin{aligned} W(z,u) &= \sum w_{n,k} u^k z^n = \sum_{T \in \mathcal{B}} u^{v(T)} z^{|T|} \\ W(z,u) &= \sum_{i \in I} u^i + zW(z,u)^2 \\ &\quad + z \sum_{n,k} w_{nk} z^n W(z, u^k) \end{aligned}$$

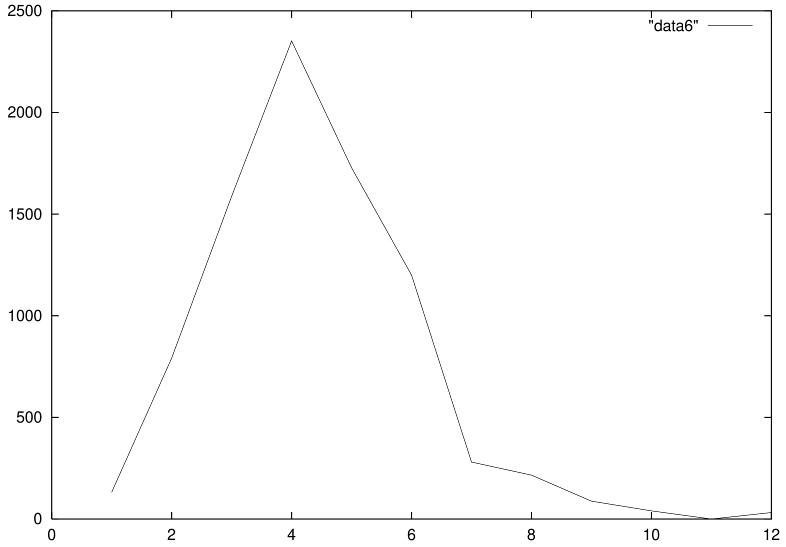
Calcul des coefficients ($I \subset \mathbb{N}$) :

$$\begin{aligned} w_{nk} &= (\sum_{i \in I} \delta_{ik}) \delta_{n0} \\ &\quad + \sum_{l=0}^{k} \sum_{m=0}^{n-1} w_{ml} w_{n-1-m,k-l} \\ &\quad + \sum_{l|k} \sum_{m=0}^{n-1} w_{ml} w_{n-1-m,k/l} \end{aligned}$$

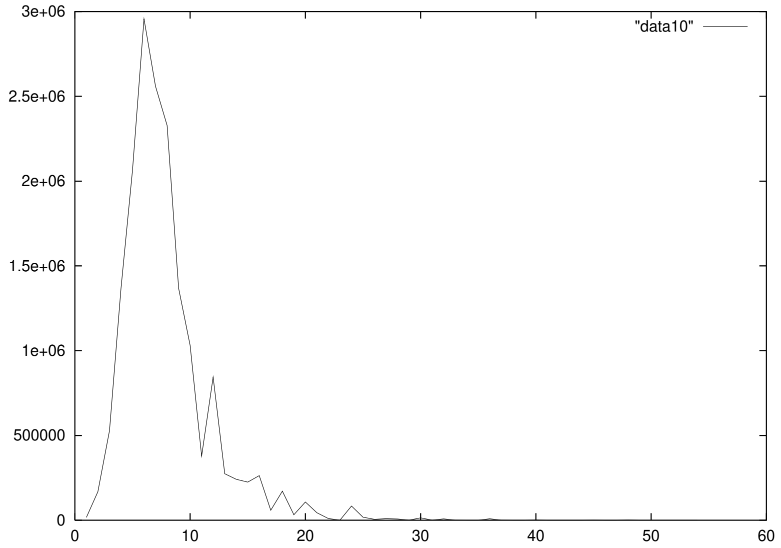
On peut donc calculer les coefficients par Maple.

Arbres $Op = \{+, \times\}$ $I = \{1\}$

Distribution des valuations pour $n = 6$:



Distribution des valuations pour $n = 10$:

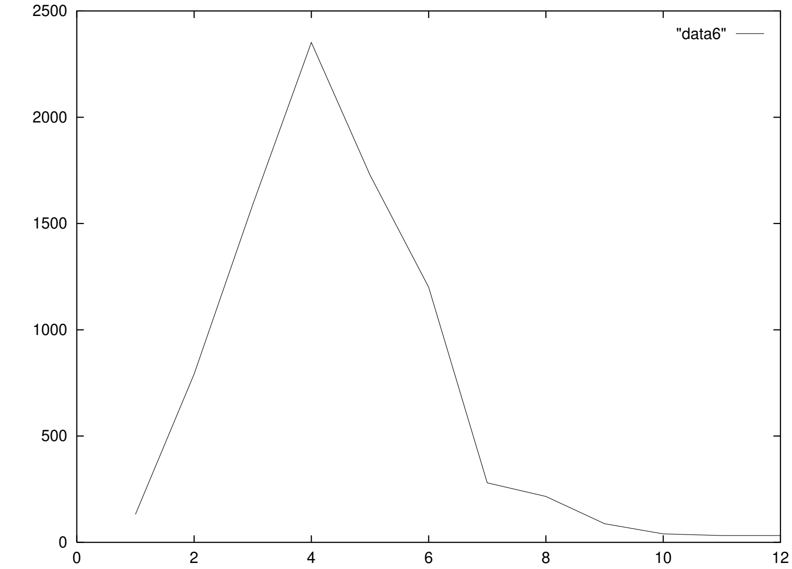


Limite log-normale ?

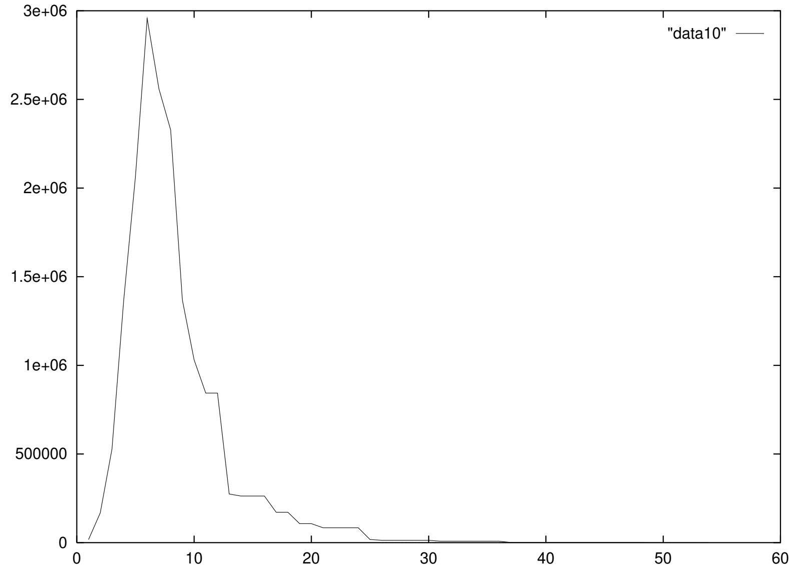
Arbres $Op = \{+, \times\}$ $I = \{1\}$

Courbes lissées

Distribution des valuations pour $n = 6$:



Distribution des valuations pour $n = 10$:

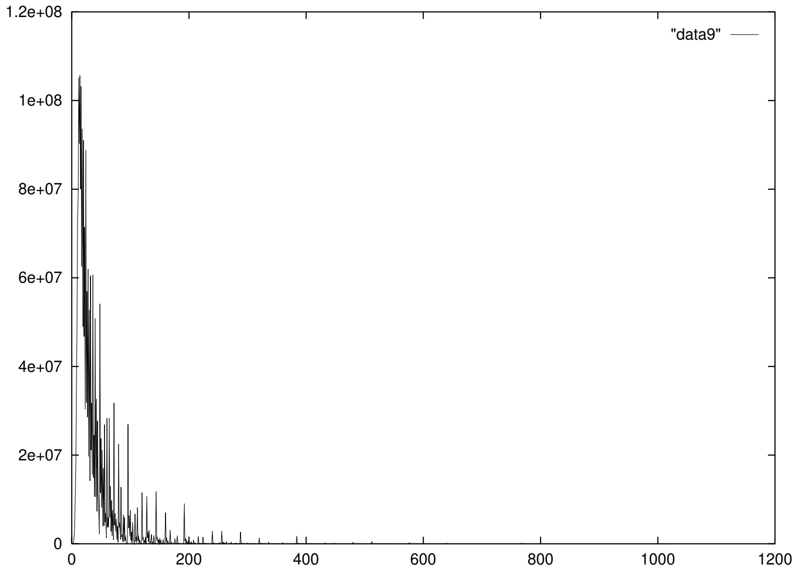


Limite log-normale ?

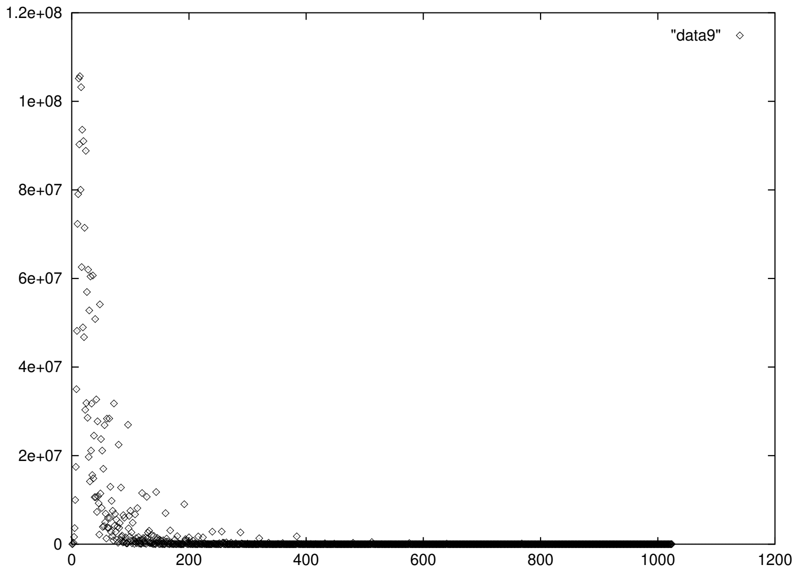
Arbres $Op = \{+, \times\}$ $I = \{1, 2\}$

Notons qu'on peut avoir des TROUS dans la courbe.

Distribution des valuations pour $n = 9$:



Même courbe tracée par points :



Arbres $Op = \{+, \times\}$

Moments :

La fonction génératrice $M_m(z) = \sum_{n\geq 0} \sum_{\mathcal{T},|\mathcal{T}|=n} v(T)^m z^n$ des moments d'ordre m vérifie :

$$M_m(z) = \sum_{k\in I} k^m + z \sum_{p=0}^{m} \binom{m}{p} M_p M_{m-p} + z M_m^2$$

$\Rightarrow$ Calcul possible par Maple.

Asymptotique des premiers moments pour $I = \{1\}$:

- $[z^n]M_0(z) \sim \frac{8^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_1(z) \sim 3\sqrt{3}\frac{9^n}{\sqrt{\pi} n^{3/2}} \sim 5,196152420\frac{9^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_2(z) \sim 7,163798465\frac{(10{,}74648525)^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_3(z) \sim 9,711750970\frac{(13{,}33806639)^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_4(z) \sim 13,15596181\frac{(17{,}07712111)^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_5(z) \sim 17,91555400\frac{(22{,}42213467)^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_6(z) \sim 24,57004515\frac{(30{,}03715145)^n}{\sqrt{\pi} n^{3/2}}$
- $[z^n]M_7(z) \sim 33,934488649\frac{(40{,}87467346)^n}{\sqrt{\pi} n^{3/2}}$

$\Rightarrow$ Pas de schéma de récurrence sympathique *a priori*.

Conclusion

- On obtient des limites (vraisemblablement) gaussiennes pour les jeux d'opérateurs $\{+, -\}$ et $\{-\}$.
- L'opérateur $-$ conduit à des bijections intéressantes faisant intervenir les nombres de Narayana.
- L'étude des jeux d'opérateurs comme $\{+, \times\}$ mène à des fonctions génératrices moyennement sympathiques *a priori*.
- À l'avenir on projette d'étudier des opérateurs extrémaux, dans le cadre de l'algèbre $\{\max, +\}$ par exemple, ainsi que des expressions rationnelles (opérateurs concaténation, union, étoile).