

D.E.A. de Mathématiques Appliquées

<http://www.univ-lille1.fr/math/deaMA>

Mémoire de D.E.A. intitulé
Arbres aléatoires et analyse de singularités
présenté par
Christian Costermans
soutenu en juin 2004

Directeurs de mémoire : Hoang Ngoc Minh, Anne Philippe
Laboratoire de Mathématiques Appliquées CNRS FRE 2222
Université des Sciences et Technologies de Lille

Résumé : Dans ce mémoire, nous étudions des fonctionnelles additives sur des arbres aléatoires. Ces fonctionnelles sont à interpréter comme des coûts de complexité d'algorithmes récursifs. Notre but est de déterminer des développements asymptotiques du coût moyen de tels algorithmes, ainsi que la distribution limite de ce coût. Pour cela, nous allons nous ramener à l'étude de fonctions génératrices et faire appel à des techniques d'analyse de singularités. Celles-ci consistent à trouver des équivalents pour des fonctions complexes au voisinage de leur singularité, pour en déduire des équivalents asymptotiques de leurs coefficients de Taylor.

Mots-clés : fonctionnelles additives, modèle de permutation aléatoire, développements singuliers, analyse de singularités, théorèmes de transfert.

Table des matières

1	Analyse de singularités	5
1.1	Position du problème	5
1.2	Développement singulier	6
1.3	Opérations sur les développements	9
1.4	Produit de Hadamard	12
1.5	Le modèle des arbres de Catalan	15
2	Etude des arbres m-aires de recherche	18
2.1	Développement asymptotique de $f_n = \mathbb{E}[X_n]$	18
2.2	Théorèmes de transfert	19
2.3	Distribution limite de X_n	26
2.3.1	Normalité asymptotique pour des fonctions-test "faibles" . . .	26
2.3.2	Fonctions-test modérées et élevées	30
2.4	Simulations dans le cas $m = 2$	34
A	Propriétés du polynôme caractéristique $\psi(\lambda)$	39
B	Programmes utilisés pour les simulations	40

Introduction

Motivations

Nous allons dans ce mémoire étudier des fonctionnelles additives sur des arbres aléatoires, distribués selon différents modèles de probabilités. Typiquement, ces fonctionnelles peuvent être vues comme des coûts de complexité pour des algorithmes de type “Diviser pour régner” - par exemple, la recherche d’un motif -. Nous allons ainsi établir des résultats concernant le coût asymptotique de ce type d’algorithmes. Ce travail est basé sur les travaux de thèse de Kapur [5], eux-mêmes largement inspirés des techniques d’analyse de singularités, développées par Flajolet [3]. Il s’inscrit dans la prolongation des recherches actuelles de l’équipe de bio-informatique de Lille 2, relatives à la modélisation de puces à ADN par des arbres pondérés [7]. Ces recherches se concentrent actuellement sur la détermination d’une loi de probabilité associée à ces arbres. Cette loi connue, ce mémoire permettra d’étudier la complexité des algorithmes liés à ce modèle.

Vocabulaire

Nous définissons dans un premier temps les notions sur lesquelles se fonde ce mémoire, à savoir arbres m -aires de recherche, fonctionnelles additives. Nous définirons ensuite les modèles de probabilité sur lesquels nous allons travailler.

Un **arbre m -aire** est un arbre avec, au plus, m “descendants” pour chaque noeud. En termes récursifs, un arbre m -aire est, soit vide, soit un noeud principal (appelé racine) lié à un ensemble de m sous-arbres, chacun étant lui-même un arbre m -aire.

Un **arbre m -aire de recherche** est un arbre m -aire dans lequel chaque noeud peut contenir $m - 1$ éléments d’un ensemble ordonné (ensemble des “clés”). Ainsi, sans perte de généralité, on peut supposer l’ensemble des clés égal à $\{1, 2, \dots, n\}$. L’arbre m -aire de recherche correspondant à une suite de n clés distinctes se construit ainsi :

1. Si $n < m$, toutes les clés sont stockées dans la racine en ordre croissant.
2. Si $n \geq m$, alors les $m - 1$ premières clés sont stockées dans la racine en ordre croissant. Les $n - (m - 1)$ restantes sont stockées dans les sous-arbres avec la condition suivante : si $\sigma_1 < \sigma_2 < \dots < \sigma_{m-1}$ représente la suite ordonnée des clés contenues dans la racine, alors les clés contenues dans le j -ième sous-arbre sont toutes celles comprises entre σ_j et σ_{j+1} , avec la convention $\sigma_0 = 0$ et $\sigma_m = n + 1$.
3. Tous les sous-arbres sont des arbres m -aires de recherche répondant aux deux premières conditions.

Soit T un arbre m -aire de recherche, nous noterons $|T|$ le nombre de clés stockées dans T , et $L_j(T)$ le j -ième sous-arbre issu de la racine de T , pour $1 \leq j \leq m$.

Soit x un noeud de T , nous noterons T_x le sous-arbre de T composé de x et de ses descendants. Ces notations sont illustrées Figure 1.

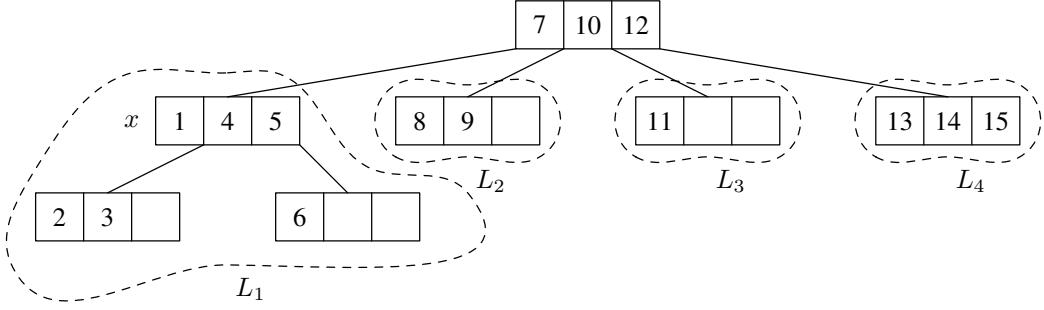


FIG. 1 – Exemple de notations pour l’arbre quaternaire T . Dans cet exemple, $|T| = 15$ and $T_x = L_1$. De plus, l’arbre T est construit à partir de la suite $(7, 10, 12, 1, 4, 5, 2, 6, 3, 8, 11, 14, 15, 13, 9)$

Nous pouvons maintenant définir une **fonctionnelle additive** :

Définition 1 Soit $m \geq 2$ et f une fonctionnelle définie sur l’ensemble des arbres m -aires de recherche, à valeurs réelles. Nous dirons que f est une fonctionnelle additive si elle vérifie la récurrence

$$f(T) = \sum_{i=1}^m f(L_i(T)) + t_{|T|} \quad (1)$$

pour tout arbre T tel que $|T| \geq m - 1$, et pour une suite de réels $(t_n)_{n \geq 0}$ fixée, que nous appelerons “suite-test”.

Comme nous l’avons déjà précisé, $f(T)$ doit être vu comme la complexité d’un algorithme récursif sur un arbre T , algorithme qui nécessite une complexité $t_{|T|}$ initiale (comparaisons, opérations arithmétiques...etc) puis effectue m appels récursifs.

Il nous reste maintenant à préciser les modèles de probabilité sous lesquels nous allons tirer nos arbres “au hasard”.

Modèles de probabilité

D’après la description des arbres de recherche effectuée précédemment, toute permutation de $\{1, 2, \dots, n\}$ engendre un arbre m -aire de recherche.

Nous parlerons de **modèle uniforme** sur l’ensemble des arbres m -aires de recherche, dans le cas où chaque arbre, à n clés, a la même probabilité. Dans ce cas, cette probabilité est simplement l’inverse du nombre d’arbres m -aires de recherche à n clés. Dans le cas $m = 2$, ce nombre est le n -ième nombre de Catalan, et c’est pourquoi nous parlerons plutôt du modèle de Catalan.

Si nous mettons maintenant une probabilité uniforme sur les $n!$ permutations de $\{1, 2, \dots, n\}$, alors les arbres ne sont plus distribués selon le modèle uniforme. Par exemple, si $m = 2$ et $n = 3$, le nombre d’arbre binaires à trois clés est $C_3 = 5$, mais les permutations (213) et (231) engendrent le même arbre, qui aura donc

pour probabilité $\frac{2}{3!} = \frac{1}{3}$. Nous parlerons dans ce cas de **modèle de permutation aléatoire**.

1 Analyse de singularités

1.1 Position du problème

Partons d'un exemple simple, que nous traiterons complètement dans la section suivante. On considère le cas $m = 2$ et on note notre fonctionnelle $f(T) = X_n$, pour $|T| = n > 0$. On peut alors reformuler l'équation (1) ainsi :

$$X_n = X_{K_n} + X_{n-1-K_n} + t_n$$

où K_n désigne une variable aléatoire indépendante de X_n et représentant le nombre de clés contenues dans le sous-arbre gauche. Dans le cas du modèle de permutation aléatoire, on voit facilement que $\mathbb{P}(K_n = k) = \frac{1}{n}$ pour $0 \leq k \leq n - 1$. En effet, si on note σ une variable uniformément distribuée sur l'ensemble des permutations de $\{1, 2, \dots, n\}$, on a

$$\mathbb{P}(K_n = k) = \mathbb{P}(\sigma(1) = k + 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

Donc, en conditionnant par K_n , et en notant $f_n = \mathbb{E}[X_n]$, on obtient

$$\begin{aligned} f_n &= t_n + \frac{1}{n} \sum_{k=0}^{n-1} f_k + f_{n-1-k} \\ &= t_n + \frac{2}{n} \sum_{k=0}^{n-1} f_k \end{aligned}$$

avec la condition initiale $f_0 = t_0$.

Le problème qui se pose alors est le suivant : la suite (t_n) étant connue, comment déterminer le comportement asymptotique de f_n ?

Pour cela, on considère les fonctions génératrices $f(z) = \sum_{n \geq 0} f_n z^n$ et $t(z) = \sum_{n \geq 0} t_n z^n$. L'équation reliant f_n à t_n se transfère alors, via des manipulations classiques, en une équation reliant $f(z)$ à $t(z)$:

$$f(z) = t(z) + 2 \int_0^z f(w) \frac{dw}{1-w}$$

équation qui se résout, par dérivation, et par méthode de variation de la constante, en :

$$f(z) = (1-z)^{-2} \int_0^z t'(w) (1-w)^2 dw$$

en supposant, sans perte de généralité, que $f_0 = t_0 = 0$.

Sous cette forme, on peut traiter un premier exemple directement.

Si $t_n = \binom{n+\alpha}{\alpha}$, i.e. $t(z) = (1-z)^{-\alpha-1}$, alors

– Si $\alpha \neq 1$

$$\begin{aligned} f(z) &= \frac{\alpha+1}{\alpha-1} \left[(1-z)^{-\alpha-1} - (1-z)^{-2} \right] \\ f_n &= \frac{\alpha+1}{\alpha-1} \left[\binom{n+\alpha}{\alpha} - (n+1) \right] \end{aligned}$$

– Si $\alpha = 1$, alors $f(z) = \frac{2}{(1-z)^2} \log \frac{1}{1-z}$ et $f_n = 2(n+1)(H_{n+1}-1)$, avec $H_n = \sum_{i=1}^n \frac{1}{i}$ le n -ième nombre harmonique.

Or, d'après la formule de Stirling, $\binom{n+\alpha}{\alpha} \sim \frac{n^\alpha}{\Gamma(\alpha+1)}$.

On peut donc résumer ces résultats ainsi :

$$\begin{aligned} t_n = \binom{n+\alpha}{\alpha}, \alpha > 1 &\implies f_n \sim \frac{\alpha+1}{\alpha-1} \frac{n^\alpha}{\Gamma(\alpha+1)} \\ t_n = n+1 &\implies f_n \sim 2n \log n \\ t_n = \binom{n+\alpha}{\alpha}, 0 < \alpha < 1 &\implies f_n \sim \frac{1+\alpha}{1-\alpha} n \end{aligned}$$

Dans cet exemple, le choix de t_n nous a conduit à une expression explicite de $t(z)$ et donc de $f(z)$. Mais si $t_n = \sqrt{n}$ ou $t_n = \ln n$, nous n'aurons plus de formule explicite, et nous devrons alors nous contenter d'équivalents asymptotiques, au voisinage de 1, pour $t(z)$.

L'exemple traité nous permet d'esquisser le raisonnement général, et sur lequel nous allons mettre des bases théoriques dans les paragraphes suivants :

- Trouver une relation entre f_n et t_n
- En déduire une relation entre $f(z)$ et $t(z)$
- Chercher un développement asymptotique pour $f(z)$
- Conclure sur le comportement asymptotique de f_n .

1.2 Développement singulier

Lorsqu'une fonction f possède une unique singularité, il est toujours possible, quitte à effectuer un changement de variable, de la ramener en $z = 1$. Partant de ce constat, une famille de fonctions assez naturelle pour effectuer nos développements asymptotiques est celle des fonctions $\{(1-z)^\alpha, \alpha \in \mathbb{R}\}$.

Définition 2 Une fonction définie par une série entière ayant un rayon de convergence égal à 1 est dite Δ -régulière si elle peut être prolongée analytiquement, à l'exception de $z = 1$, dans un domaine

$$\Delta(\phi, \eta) = \{z / |z| < 1 + \eta, |\arg(z-1)| > \phi\}$$

avec $\eta > 0$ et $0 < \phi < \pi/2$. On dit qu'une fonction f admet un développement Δ -singulier en $z=1$ si elle est Δ -régulière et s'il existe A tel que

$$f(z) = \sum_{n=0}^N c_n (1-z)^{\alpha_n} + O(|1-z|^A) \tag{2}$$

pour une suite de complexes $(c_n)_{0 \leq n \leq N}$ et une suite de réels $(\alpha_n)_{0 \leq n \leq N}$ telle que $\alpha_n < A$.

Le choix d'un tel domaine d'analyticité permet la mise en place de théorèmes reliant le comportement de $f(z)$ au voisinage de sa singularité $z = 1$ avec le comportement asymptotique de ses coefficients de Taylor. Nous faisons appel à la notation suivante : si $f(z) = \sum_{n \geq 0} f_n z^n$, on note $f_n = [z^n]f(z)$.

Théorème 1 *Si $f(z)$ admet un développement Δ -singulier de la forme (2) valide dans un domaine Δ , alors*

$$[z^k]f(z) = \sum_{n=0}^N c_n \binom{k - \alpha_n - 1}{-\alpha_n - 1} + O(k^{-A-1})$$

La seule partie à démontrer est en fait celle concernant le terme d'erreur :

Théorème 2 *Soit f une fonction analytique dans un domaine $\Delta(\phi_1, \eta_1)$, avec $\eta_1 > 0$ et $0 < \phi_1 < \frac{\pi}{2}$.*

Si $f(z) = O(|1 - z|^\alpha)$ pour un certain réel α , alors

$$f_k = [z^k]f(z) = O(k^{-\alpha-1})$$

Démonstration: La condition $f(z) = O(|1 - z|^\alpha)$ se traduit par l'existence d'une constante K vérifiant pour tout $z \neq 1$

$$|f(z)| < K|1 - z|^\alpha. \quad (3)$$

D'après la formule de Cauchy, nous avons

$$f_k = \frac{1}{2i\pi} \int_{O^+} f(z) \frac{dz}{z^{k+1}}$$

où O^+ désigne n'importe quel contour orienté positivement, entourant l'origine, et à l'intérieur de Δ : nous choisissons le contour suivant, orienté positivement : $C = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, avec :

$$\begin{aligned} \gamma_1 &= \{z / |z - 1| = \frac{1}{k}, |\arg(z - 1)| \geq \phi\} \\ \gamma_2 &= \{z / \frac{1}{k} \leq |z - 1|, |z| \leq 1 + \eta, |\arg(z - 1)| = \phi\} \\ \gamma_3 &= \{z / |z| = 1 + \eta, |\arg(z - 1)| \geq \phi\} \\ \gamma_4 &= \{z / \frac{1}{k} \leq |z - 1|, |z| \leq 1 + \eta, |\arg(z - 1)| = -\phi\} \end{aligned}$$

où $\frac{\pi}{2} > \phi > \phi_1$ et $0 < \eta < \eta_1$, de sorte que le contour C soit effectivement intérieur à Δ . Nous allons maintenant évaluer les contributions des intégrales sur les différents chemins, en posant

$$f_k^{(i)} = \frac{1}{2i\pi} \int_{\gamma_i} f(z) \frac{dz}{z^{k+1}}.$$

1. Petit cercle : d'après (3), on a, en utilisant des majorations immédiates, pour $k \geq 4$,

$$\begin{aligned} |f_k^{(1)}| &\leq \frac{1}{2\pi} \cdot K \left(\frac{1}{k}\right)^\alpha \cdot \left(1 - \frac{1}{k}\right)^{-k-1} \cdot \left(\frac{2\pi}{k}\right) \\ &\leq 5 (Kk^{-\alpha-1}) \end{aligned}$$

2. Partie rectiligne : on remarque d'abord que, pour des raisons de symétrie, la majoration obtenue pour $|f_k^{(2)}|$ sera également valable pour $|f_k^{(4)}|$. On pose $\omega = e^{i\phi}$, et on effectue le changement de variable $z = 1 + \frac{\omega t}{k}$. On obtient alors

$$\begin{aligned} |f_k^{(2)}| &\leq \frac{1}{2\pi} \int_1^{kE} K \left(\frac{t}{k}\right)^\alpha \left|1 + \frac{\omega t}{k}\right|^{-k-1} \frac{dt}{k} \\ &\leq (Kk^{-\alpha-1}) \cdot \frac{1}{2\pi} \int_1^\infty t^\alpha \left|1 + \frac{\omega t}{k}\right|^{-k-1} dt \end{aligned}$$

Dans cette majoration, E est défini de telle sorte que γ_2 et γ_3 se rejoignent : précisément E est la racine positive de l'équation $|1 + Ee^{i\phi}| = 1 + \eta$. Il reste maintenant à majorer l'intégrale, uniformément en k . Or,

$$\left|1 + \frac{\omega t}{k}\right| \geq 1 + \operatorname{Re} \left(\frac{\omega t}{k}\right) = 1 + \frac{t}{k} \cos \phi$$

Par conséquent,

$$|f_k^{(2)}| \leq \frac{J_k}{2\pi} (Kk^{-\alpha-1}) \text{ avec } J_k = \int_1^\infty t^\alpha \left(1 + \frac{t \cos \phi}{k}\right)^{-k} dt$$

A partir de là, on constate que

$$J_k \xrightarrow[k \rightarrow \infty]{} \int_1^\infty t^\alpha e^{-t \cos \phi} dt$$

et donc que la suite (J_k) est bornée par une constante ne dépendant que de α et ϕ . En résumé

$$|f_k^{(2)}| \leq \frac{J(\alpha, \phi)}{2\pi} (Kk^{-\alpha-1}) \text{ avec } J(\alpha, \phi) = \sup_{k \geq |\alpha|+4} \int_1^\infty t^\alpha \left(1 + \frac{t \cos \phi}{k}\right)^{-k} dt$$

3. Grand cercle : comme pour 1., en utilisant la majoration (3), on obtient :

$$\begin{aligned} |f_k^{(3)}| &\leq \frac{1}{2\pi} \cdot K(2 + \eta)^\alpha \cdot (1 + \eta)^{-k-1} \cdot (2\pi(1 + \eta)) \\ &\leq K \frac{(2 + \eta)^\alpha}{(1 + \eta)^k} \end{aligned}$$

4. En réunissant toutes les majorations trouvées, on obtient, pour $k \geq |\alpha| + 4$,

$$|f_k| \leq (Kk^{-\alpha-1}) \left[5 + \frac{J(\alpha, \phi)}{\pi} + \frac{(2 + \eta)^\alpha}{(1 + \eta)^k} k^{\alpha+1} \right]$$

Par ailleurs, il existe une constante k_1 (dépendant uniquement de α et η), telle que, pour $k \geq k_1$,

$$\frac{(2+\eta)^\alpha}{(1+\eta)^k} k^{\alpha+1} \leq 1$$

Ainsi, on obtient finalement, pour tout $k \geq k_0$, avec $k_0 = \max(k_1, |\alpha| + 4)$,

$$|f_k| \leq (Kk^{-\alpha-1}) \left[5 + \frac{J(\alpha, \phi)}{\pi} \right]$$

ce qui est le résultat cherché.

1.3 Opérations sur les développements

L'exemple de l'arbre binaire de recherche, sous le modèle de permutation aléatoire, nous a conduit à l'équation $f(z) = (1-z)^{-2} \int_0^z t'(w)(1-w)^2 dw$. Dans l'objectif de déterminer le comportement asymptotique de $f(z)$ au voisinage de $z = 1$, il convient donc de s'intéresser aux effets de la dérivation et de l'intégration sur des éléments singuliers du type $(1-z)^\alpha$.

Théorème 3 (Dérivation d'un développement singulier) *Soit f une fonction Δ -régulière admettant un développement Δ -singulier, alors pour tout entier $r > 0$, $\frac{d^r}{dz^r} f(z)$ est également Δ -régulière et admet un développement calculable par dérivation terme à terme :*

$$\frac{d^r}{dz^r} f(z) = (-1)^r \sum_{n=0}^N c_n \frac{\Gamma(\alpha_n + 1)}{\Gamma(\alpha_n + 1 - r)} (1-z)^{\alpha_n - r} + O(|1-z|^{A-r})$$

Démonstration: Là encore, la seule partie à démontrer est celle concernant le terme d'erreur. De plus, par une récurrence immédiate, on peut se ramener au cas $r = 1$. Soit donc $g(z) = \frac{f(z)}{(1-z)^A}$ où $f(z) = O(|1-z|^A)$. On a alors

$$g'(z) = \frac{f'(z)}{(1-z)^A} + A \frac{f(z)}{(1-z)^{A+1}}$$

qu'on peut reformuler ainsi :

$$f'(z) = g'(z)(1-z)^A - A \frac{f(z)}{1-z}$$

Par hypothèse, au voisinage de 1, on a $g(z) = O(1)$. Il nous faut donc démontrer que $g'(z) = O(|1-z|^{-1})$, au voisinage de 1, pour en déduire le théorème.

Soit $\Delta = \Delta(\phi_1, \eta_1)$ le domaine de régularité de f . Nous allons travailler dans un domaine $\Delta(\phi + \varepsilon, \eta)$, avec $\varepsilon > 0$, $\phi > \phi_1$ et $\eta < \eta_1$. On fixe $\varepsilon > 0$ "petit" et on restreint z au domaine $\Delta(\phi + \varepsilon, \eta)$. A partir de la formule de Cauchy

$$g(z) = \frac{1}{2i\pi} \int_{\gamma} g(\omega) \frac{d\omega}{\omega - z}$$

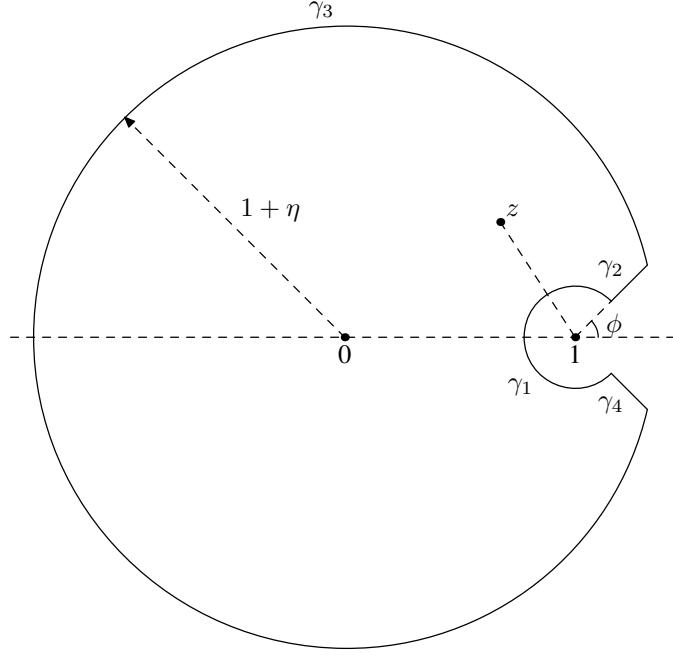


FIG. 2 – Le contour γ utilisé dans la démonstration du théorème de dérivation

on en déduit, par dérivation,

$$g'(z) = \frac{1}{2i\pi} \int_{\gamma} g(\omega) \frac{d\omega}{(\omega - z)^2}$$

De plus, on choisit comme contour γ celui dessiné Figure 2 : $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, avec

$$\begin{aligned} \gamma_1 &= \{\omega / |\omega - 1| = \frac{1}{2}|z - 1|, |\arg(\omega - 1)| \geq \phi\} \\ \gamma_2 &= \{\omega / \frac{1}{2}|z - 1| \leq |\omega - 1|, |\omega| \leq 1 + \eta, \arg(\omega - 1) = \phi\} \\ \gamma_3 &= \{\omega / |\omega| = 1 + \eta, |\arg(\omega - 1)| \geq \phi\} \\ \gamma_4 &= \{\omega / \frac{1}{2}|z - 1| \leq |\omega - 1|, |\omega| \leq 1 + \eta, \arg(\omega - 1) = -\phi\} \end{aligned}$$

Le choix $\phi > \phi_1$ et $\eta < \eta_1$ nous garantit que le contour γ est bien inclus dans le domaine d'analyticité de g . Nous allons maintenant évaluer les contributions des intégrales le long des différents chemins $\gamma_1, \gamma_2, \gamma_3, \gamma_4$. Pour le petit cercle,

$$\begin{aligned} \left| \int_{\gamma_1} \right| &= O(1) \cdot O\left(\frac{1}{|1-z|^2}\right) \cdot O(|1-z|) = O(|1-z|^{-1}) \\ \left| \int_{\gamma_3} \right| &= O(1) = O(|1-z|^{-1}) \end{aligned}$$

Si nous posons $|1-z| = \delta$, les intégrales le long de γ_2 et γ_4 sont majorées, via des considérations géométriques par :

$$\left| \int_{\gamma_2} \right| \leq \int_{\frac{\delta}{2}}^{+\infty} \frac{du}{\delta^2 + u^2} = O(\delta^{-1}) = O(|1-z|^{-1})$$

La combinaison des différentes majorations nous amène à $g'(z) = O(|1-z|^{-1})$, et donc à la conclusion du théorème.

Théorème 4 (Intégration d'un développement singulier) *Soit f une fonction Δ -régulière admettant un développement Δ -singulier, alors la fonction $z \mapsto \int_0^z f(t)dt$ est également Δ -régulière. Supposons que pour tout n , $\alpha_n \neq -1$ et que $A \neq -1$.*

– *Si $A < -1$, alors le développement singulier de $\int_0^z f(t)dt$ est donné par :*

$$\int_0^z f(t)dt = - \sum_{n=0}^N \frac{c_n}{\alpha_n + 1} (1-z)^{\alpha_n + 1} + O(|1-z|^{A+1})$$

– *Si $A > -1$, alors le développement singulier de $\int_0^z f(t)dt$ est donné par :*

$$\int_0^z f(t)dt = - \sum_{n=0}^N \frac{c_n}{\alpha_n + 1} (1-z)^{\alpha_n + 1} + L_0 + O(|1-z|^{A+1})$$

où la constante d'intégration L_0 vaut :

$$L_0 = \sum_{\alpha_n < -1} \frac{c_n}{\alpha_n + 1} + \int_0^1 (f(t) - \sum_{\alpha_n < -1} c_n (1-t)^{\alpha_n}) dt$$

Démonstration: Soit $r(z)$ le terme d'erreur dans le développement de $f(z)$:

$$r(z) = f(z) - \sum_{n=0}^N c_n (1-z)^{\alpha_n}$$

Par hypothèse, il existe une constante $K > 0$ telle que pour tout z appartenant à Δ , on a : $|r(z)| \leq K|1-z|^A$.

1. Si $A < -1$, il suffit de démontrer que

$$\int_0^z r(t)dt = O(|1-z|^{A+1}),$$

où l'intégrale peut être calculée sur n'importe quel chemin reliant 0 à z , à l'intérieur du domaine d'analyticité de r . Nous choisissons le contour $\gamma = \gamma_1 \cup \gamma_2$, dessinée Figure 3 .

Alors

$$\begin{aligned} \left| \int_\gamma r(t)dt \right| &\leq \left| \int_{\gamma_1} r(t)dt \right| + \left| \int_{\gamma_2} r(t)dt \right| \\ &\leq K \int_{\gamma_1} |1-t|^A ds + K \int_{\gamma_2} |1-t|^A ds \\ &= O(|1-z|^{A+1}) \end{aligned}$$

Le résultat en $O(|1-z|^{A+1})$ provient pour l'intégrale le long de γ_1 d'un calcul explicite; pour l'intégrale le long de γ_2 , ce résultat résulte de la majoration par la longueur du contour γ_2 , la valeur $|1-t|$ étant fixe.

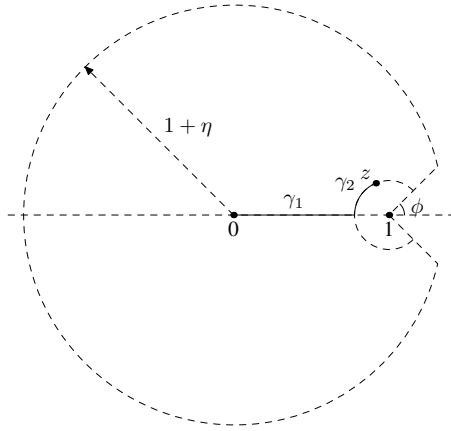


FIG. 3 – Le contour utilisé dans la démonstration du théorème d’intégration.

2. Si $A > -1$: nous posons $f_-(z)$ la “partie divergente” de f donnant lieu à des problèmes d’intégrabilité en 1 :

$$f_-(z) = \sum_{\alpha_j < -1} c_j (1-z)^{\alpha_j}.$$

Nous effectuons alors la décomposition $f = (f - f_-) + f_-$. On a :

$$\int_0^z f_-(t) dt = - \sum_{\alpha_j < -1} \frac{c_j}{\alpha_j + 1} (1-z)^{\alpha_j + 1} + \sum_{\alpha_j < -1} \frac{c_j}{\alpha_j + 1}$$

De plus, le choix de $f_-(z)$ nous garantit l’existence de $\int_0^1 (f - f_-)(t) dt$, donc

$$\int_0^z (f - f_-)(t) dt = \int_0^1 (f - f_-)(t) dt + \int_1^z (f - f_-)(t) dt$$

Le premier terme est une constante, que nous retrouvons dans l’expression de L_0 . Le deuxième terme se décompose ainsi :

$$\int_1^z (f - f_-)(t) dt = - \sum_{\alpha_j > -1} \frac{c_j}{\alpha_j + 1} (1-z)^{\alpha_j + 1} + \int_1^z r(t) dt.$$

Comme $A > -1$ le terme d’erreur est fini, et une intégration le long du segment reliant 1 à z fournit le résultat $\int_1^z r(t) dt = O(|1-z|^{A+1})$.

1.4 Produit de Hadamard

Dans certains cas, comme celui du modèle de Catalan, que nous traiterons en application, on peut être amené à considérer des fonctions génératrices à coefficients pondérés. Par exemple, plutôt que de poser $f(z) = \sum_{n \geq 0} f_n z^n$, nous pouvons poser $\tilde{f}(z) = \sum_{n \geq 0} C_n f_n z^n$, où les C_n sont les nombres de Catalan, des constantes connues. $\tilde{f}(z)$ se présente alors comme le produit de Hadamard des fonctions f et C définie par

$C(z) = \sum_{n \geq 0} C_n z^n$. Plus généralement, si $a(z) = \sum_{n \geq 0} a_n z^n$ et $b(z) = \sum_{n \geq 0} b_n z^n$, on définit le produit de Hadamard de a et b par :

$$a(z) \odot b(z) = \sum_{n \geq 0} a_n b_n z^n$$

Il convient donc de regarder comment se comporte l'analyse de singularités par passage au produit de Hadamard.

Proposition 1 *Si a , b et $a+b$ ne sont pas des entiers, alors le produit de Hadamard $(1-z)^a \odot (1-z)^b$ admet un développement infini dans l'échelle des puissances $\{0, 1, 2, \dots\} \cup \{a+b+1, a+b+2, \dots\}$ de $(1-z)$. Précisément, au voisinage de 1,*

$$(1-z)^a \odot (1-z)^b \sim \sum_{k \geq 0} \lambda_k^{(a,b)} \frac{(1-z)^k}{k!} + \sum_{k \geq 0} \mu_k^{(a,b)} \frac{(1-z)^{a+b+1+k}}{k!}$$

avec les coefficients λ et μ donnés par

$$\begin{aligned} \lambda_k^{(a,b)} &= \frac{\Gamma(1+a+b)}{\Gamma(1+a)\Gamma(1+b)} \frac{(-a)^{\bar{k}}(-b)^{\bar{k}}}{(-a-b)^{\bar{k}}} \\ \mu_k^{(a,b)} &= \frac{\Gamma(-a-b-1)}{\Gamma(-a)\Gamma(-b)} \frac{(1+a)^{\bar{k}}(1+b)^{\bar{k}}}{(2+a+b)^{\bar{k}}} \end{aligned}$$

avec la notation $x^{\bar{k}} = x(x+1) \cdots (x+k-1)$.

Démonstration: Ce résultat découle du développement :

$$(1-z)^a = 1 + \frac{-a}{1}z + \frac{(-a)(-a+1)}{2!}z^2 + \dots$$

Il vient alors, en effectuant une multiplication terme à terme :

$$(1-z)^a \odot (1-z)^b = {}_2F_1[-a, -b; 1; z]$$

où ${}_2F_1$ représente la fonction hypergéométrique de Gauss définie par

$${}_2F_1[\alpha, \beta; \gamma; z] = 1 + \frac{\alpha\beta}{\gamma} \frac{z}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)} \frac{z^2}{2!} + \dots$$

Mais alors, d'après la théorie des transformations de fonctions hypergéométriques [4], nous savons, en général, que celles-ci peuvent être développées au voisinage de $z = 1$ via la transformation $z \mapsto 1 - z$. Dans le cas $\gamma = 1$, nous avons en effet

$$\begin{aligned} {}_2F_1[\alpha, \beta; 1; z] &= \frac{\Gamma(1-\alpha-\beta)}{\Gamma(1-\alpha)\Gamma(1-\beta)} {}_2F_1[\alpha, \beta; \alpha+\beta; 1-z] \\ &+ \frac{\Gamma(\alpha+\beta-1)}{\Gamma(\alpha)\Gamma(\beta)} (1-z)^{-\alpha-\beta+1} {}_2F_1[1-\alpha, 1-\beta; 2-\alpha-\beta; 1-z]. \end{aligned}$$

Cette égalité nous fournit le développement donné dans la proposition, avec $\alpha = -a$ et $\beta = -b$.

Proposition 2 (Composition des termes d'erreurs) *Soit f et g deux fonctions Δ -régulières dans un domaine $\Delta(\phi, \eta)$ telles que, pour $z \in \Delta(\phi, \eta)$, $f(z) = O(|1 - z|^a)$ et $g(z) = O(|1 - z|^b)$, avec $a + b + 1 < 0$. Alors le produit de Hadamard $f \odot g$ est régulier dans un domaine Δ' , où il vérifie*

$$(f \odot g)(z) = O(|1 - z|^{a+b+1})$$

Démonstration: Admise [2]

Nous pouvons préciser le développement de $f \odot g$ dans le cas $a + b + 1 > -1$.

Proposition 3 *Soit f et g deux fonctions Δ -régulières, telles que pour $z \in \Delta$, $f(z) = O(|1 - z|^a)$ et $g(z) = O(|1 - z|^b)$. Supposons également que $k < a + b + 1 < k + 1$ pour un entier $k \geq -1$. Alors pour $z \in \Delta'$,*

$$(f \odot g)(z) = \sum_{j=0}^k \frac{(-1)^j}{j!} (f \odot g)^{(j)}(1) (1 - z)^j + O(|1 - z|^{a+b+1})$$

Démonstration: Soit $\partial = \partial_z$ l'opérateur $\frac{d}{dz}$ et soit ϑ l'opérateur d'Euler $z\partial$. On remarque que

$$\vartheta(f \odot g) = (\vartheta f) \odot g = f \odot (\vartheta g),$$

ce qui conduit à

$$\vartheta^{k+1}(f \odot g) = (\vartheta^{k+1} f) \odot g.$$

D'après le théorème de dérivation d'un développement singulier, on sait que $\vartheta^{k+1} f(z) = O(|1 - z|^{a-k-1})$. Donc, la proposition précédente s'applique :

$$(\vartheta^{k+1}(f \odot g))(z) = O(|1 - z|^{a+b-k}).$$

Par ailleurs, pour une fonction h dans l'image de ϑ , on a

$$(\vartheta^{-1}h)(z) = P_0 + \int_0^z h(t) \frac{dt}{t}$$

avec une certaine constante d'intégration P_0 . Il est alors possible de retrouver $h = f \odot g$ par intégrations successives, en utilisant le théorème d'intégration d'un développement singulier.

Par définition de k , on a $-1 < a + b - k < 0$. Des intégrations répétées montrent alors

$$(f \odot g)(z) = P(z) + O(|1 - z|^{a+b+1}),$$

où $P(z)$ est un polynôme de degré k englobant les constantes d'intégration. Ce polynôme est donc entièrement déterminé par les $(k + 1)$ premiers termes du développement de Taylor de $f \odot g$ en 1, ce qui revient précisément à l'énoncé de notre théorème.

Nous résumons maintenant les différents résultats obtenus concernant le produit de Hadamard.

Théorème 5 (Composition de singularités) Soit f et g deux fonctions Δ -régulières admettant les développements Δ -singuliers suivants :

$$f(z) = \sum_{m=0}^M c_m (1-z)^{\alpha_m} + O(|1-z|^A), \quad g(z) = \sum_{n=0}^N d_n (1-z)^{\beta_n} + O(|1-z|^B)$$

Alors le produit de Hadamard $f \odot g$ est également Δ -régulier. De plus, il admet le développement singulier suivant :

$$f \odot g(z) = \sum_{m,n} c_m d_n [(1-z)^{\alpha_m} \odot (1-z)^{\beta_n}] + P(1-z) + O(|1-z|^C),$$

avec $C = 1 + \min\{\alpha_0 + B, A + \beta_0\}$ et P un polynôme de degré inférieur à C .

1.5 Le modèle des arbres de Catalan

Nous allons traiter complètement le modèle des arbres de Catalan, c'est-à-dire le modèle uniforme pour $m = 2$, pour voir la mise en application des principes évoqués dans cette partie assez théorique.

Notre équation fondamentale (1) peut s'écrire ainsi :

$$X_n = X_{K_n} + X_{n-1-K_n} + t_n.$$

Or, dans le modèle de Catalan, $\mathbb{P}(K_n = k) = \frac{C_k C_{n-1-k}}{C_n}$ avec $C_n = \frac{1}{n+1} \binom{2n}{n}$, le n -ième nombre de Catalan représentant le nombre d'arbres binaires à n clés.

En effet, pour obtenir un arbre binaire à n clés (parmi les C_n possibles), avec un sous-arbre gauche à k clés, il faut choisir un arbre binaire (le sous-arbre gauche) à k clés (donc C_k choix possibles), et un arbre binaire (le sous-arbre droit) à $n-1-k$ clés (donc C_{n-1-k} choix possibles).

On obtient alors, en conditionnant par K_n :

$$f_n = \mathbb{E}[X_n] = t_n + \sum_{k=0}^{n-1} \frac{C_k C_{n-1-k}}{C_n} (f_k + f_{n-1-k})$$

d'où

$$\begin{aligned} C_n f_n &= C_n t_n + \sum_{k=0}^{n-1} C_k C_{n-1-k} (f_k + f_{n-1-k}) \\ &= C_n t_n + 2 \sum_{k=0}^{n-1} C_k C_{n-1-k} f_k \end{aligned}$$

On introduit alors des fonctions génératrices pondérées par des coefficients de normalisation :

$$\tilde{t}(z) = \sum_{n \geq 0} t_n C_n z^n \quad \tilde{f}(z) = \sum_{n \geq 0} f_n C_n z^n,$$

ce qui nous conduit à

$$\tilde{f}(z) = \tilde{t}(z) + 2zC(z)\tilde{f}(z), \quad \text{où} \quad C(z) = \sum_{n \geq 0} C_n z^n = \frac{1}{2z}(1 - \sqrt{1 - 4z}).$$

Donc

$$\begin{aligned} \tilde{f}(z) &= \frac{1}{\sqrt{1 - 4z}}\tilde{t}(z) \\ &= \frac{t(z) \odot C(z)}{\sqrt{1 - 4z}}, \quad \text{où} \quad t(z) = \sum_{n \geq 0} t_n z^n \end{aligned}$$

Nous pouvons maintenant établir le théorème suivant :

Théorème 6 *Sous le modèle de Catalan, les valeurs moyennes des coûts induits par des suites-test du type $t_n = n^\alpha (\alpha > 0)$ admettent des développements asymptotiques en puissances de n et de $\ln n$. Les termes principaux sont résumés dans le tableau suivant :*

Suite-test (t_n)	Coût (f_n)
n^α $(\frac{3}{2} < \alpha)$	$\frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} n^{\alpha + \frac{1}{2}} + O(n^{\alpha - \frac{1}{2}})$
$n^{3/2}$	$\frac{1}{\Gamma(3/2)} n^2 + O(n \ln n)$
n^α $(\frac{1}{2} < \alpha < \frac{3}{2})$	$\frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} n^{\alpha + \frac{1}{2}} + O(n)$
$n^{1/2}$	$\frac{1}{\sqrt{\pi}} n \ln n + O(n)$
n^α $(0 < \alpha < \frac{1}{2})$	$K_\alpha n + O(1)$

Démonstration: Pour des raisons de simplicité, nous allons ramener la singularité en $z = 1$. Ainsi :

$$t(z) \odot C\left(\frac{z}{4}\right) = \sum_{n=1}^{\infty} \frac{n^\alpha}{n+1} \binom{2n}{n} \left(\frac{z}{4}\right)^n$$

Par ailleurs, le développement asymptotique des nombres de Catalan est connu :

$$4^{-n} C_n \sim \frac{1}{\sqrt{\pi}} n^{-3/2} \left(1 - \frac{9}{8n} + \dots\right).$$

En multipliant par n^α , on trouve ainsi un équivalent asymptotique au n -ième coefficient de Taylor. De là, on cherche à en déduire un équivalent asymptotique, au voisinage de 1, de $t(z) \odot C\left(\frac{z}{4}\right)$. Or

$$[z^n](1 - z)^{-\alpha + \frac{1}{2}} \sim \frac{n^{\alpha - \frac{3}{2}}}{\Gamma(\alpha - \frac{1}{2})} \left(1 + \frac{(2\alpha - 1)(2\alpha - 3)}{8n} + \dots\right)$$

Nous pouvons donc trouver une fonction $H(z)$ dont les coefficients correspondent asymptotiquement à ceux de $t(z) \odot C(\frac{z}{4})$:

$$H(z) = \frac{\Gamma(\alpha - \frac{1}{2})}{\sqrt{\pi}} (1-z)^{-\alpha+\frac{1}{2}} (1 + c_1(1-z) + c_2(1-z)^2 + \dots),$$

pour une suite (c_n) calculable explicitement. Le développement singulier de $t(z) \odot C(\frac{z}{4})$ est alors la somme de $H(z)$ et d'une série $P(z) = \sum_{n=0}^{+\infty} p_n (1-z)^n$.

Le développement singulier de $\tilde{f}(\frac{z}{4})$ s'obtient finalement en divisant par $\sqrt{1-z}$ et nous obtenons donc :

$$[z^n] \tilde{f}\left(\frac{z}{4}\right) \sim \frac{\Gamma(\alpha - \frac{1}{2})}{\sqrt{\pi} \Gamma(\alpha)} n^{\alpha-1} \left(1 + \frac{c'_1}{n} + \frac{c'_2}{n^2} + \dots\right) + [z^n] \frac{P(z)}{\sqrt{1-z}}$$

Après avoir divisé par $4^{-n} C_n$, on trouve enfin :

$$f_n \sim \frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma(\alpha)} n^{\alpha+\frac{1}{2}} \left(1 + \frac{c''_1}{n} + \dots\right) + R_n,$$

avec un terme R_n provenant de la série $P(z)$ de la forme

$$R_n \sim d_{-1} n + d_0 + \frac{d_1}{n} + \dots$$

Cette dernière estimation nous permet de remplir le tableau donné dans le théorème, sous réserve que α ne soit pas de la forme $\frac{k}{2}$, avec k entier. De plus, si $0 < \alpha < \frac{1}{2}$, la série définissant $\tilde{f}(\frac{z}{4})$ converge en $z = 1$; le terme dominant pour $\tilde{f}(\frac{z}{4})$ est alors $\frac{\tilde{t}(\frac{1}{4})}{\sqrt{1-z}}$, soit $\tilde{f}(\frac{z}{4}) \sim \frac{K_\alpha}{\sqrt{1-z}}$, avec $K_\alpha = \sum_{n=1}^{+\infty} \frac{n^\alpha}{n+1} \frac{1}{4^n} \binom{2n}{n}$.

Lorsque α est de la forme $\frac{k}{2}$, avec k entier, des termes logarithmiques vont apparaître, du fait de la présence de puissances entières négatives de n dans les coefficients de $\tilde{t}(\frac{z}{4})$, mais le raisonnement reste le même. Par exemple, si $\alpha = 0.5$, on a :

$$4^{-n} \sqrt{n} C_n \sim \frac{1}{\sqrt{\pi}} \frac{1}{n} + O\left(\frac{1}{n^2}\right),$$

et on a alors

$$H(z) = \frac{1}{\sqrt{\pi}} \log\left(\frac{1}{1-z}\right),$$

or

$$[z^n] (1-z)^{-1/2} \log\left(\frac{1}{1-z}\right) \sim \frac{\ln n}{\Gamma(1/2) \sqrt{n}}$$

d'où, en divisant par $4^{-n} C_n$:

$$f_n \sim \frac{1}{\sqrt{\pi}} n \ln n.$$

2 Etude des arbres m -aires de recherche

Dans cette section, nous allons nous concentrer sur le modèle de permutation aléatoire. Dans un premier temps, nous nous intéressons au cas des arbres binaires de recherche($m = 2$), pour déterminer le comportement asymptotique de $f_n = \mathbb{E}[X_n]$. Puis, nous considérerons m quelconque, ce qui nous conduira à des équivalents asymptotiques moins précis, mais nous permettra de trouver, le cas échéant, des distributions limites pour $f(T)$.

2.1 Développement asymptotique de $f_n = \mathbb{E}[X_n]$

Théorème 7 *Sous le modèle de permutation aléatoire, les valeurs moyennes des coûts induits par des suites-test du type $t_n = n^\alpha (\alpha > 0)$ admettent des développements asymptotiques en puissances de n et de $\ln n$. Les termes principaux sont résumés dans le tableau suivant :*

Suite-test (t_n)	Coût (f_n)
n^α ($2 < \alpha$)	$\frac{\alpha+1}{\alpha-1} n^\alpha + O(n^{\alpha-1})$
n^2	$3n^2 - 6n \ln n + (10 - 6\gamma)n + O(\ln n)$
n^α ($1 < \alpha < 2$)	$\frac{\alpha+1}{\alpha-1} n^\alpha + K_\alpha n + O(n^{\alpha-1})$
n	$2n \ln n + 2(\gamma - 1)n + 2 \ln n + 2\gamma + 1 + O\left(\frac{1}{n}\right)$
n^α ($0 < \alpha < 1$)	$K_\alpha n + \frac{\alpha+1}{\alpha-1} n^\alpha + K_\alpha + o(1)$

Démonstration: On rappelle l'équation établie au début de la première partie : en posant $f(z) = \sum_{n \geq 0} f_n z^n$ et $t(z) = \sum_{n \geq 0} t_n z^n$, on a la relation :

$$f(z) = (1 - z)^{-2} \int_0^z t'(w)(1 - w)^2 dw$$

On remarque d'abord que si $t_n = n^k$, avec $k \in \mathbb{N}$, l'intégration peut être menée explicitement, étant donné que $t(z)$ est alors une fraction rationnelle.

Sinon, il suffit d'examiner sur les éléments singuliers du type $(1 - z)^\beta$ les effets de la dérivation, de la multiplication par $(1 - z)^2$, de l'intégration, puis de la multiplication par $(1 - z)^{-2}$.

On a alors la chaîne suivante :

$$c(1 - z)^\beta \xrightarrow{\frac{d}{dz}} -c\beta(1 - z)^{\beta-1} \xrightarrow{\times(1-z)^{-2}} -c\beta(1 - z)^{\beta+1}$$

Pour continuer la chaîne, on va supposer que $\beta + 1 \neq -1$ (sinon, un terme logarithmique apparaît).

$$-c\beta(1 - z)^{\beta+1} \xrightarrow{f} c\frac{\beta}{\beta+2}(1 - z)^{\beta+2} \xrightarrow{\times(1-z)^{-2}} c\frac{\beta}{\beta+2}(1 - z)^\beta$$

Ainsi, un tel élément apporte à f_n une contribution $c \frac{\beta}{\beta+2} \binom{n-\beta-1}{-\beta-1}$, qui est équivalent à $c \frac{\beta}{\beta+2} \frac{n^{-\beta-1}}{\Gamma(-\beta)}$ (le traitement de termes logarithmiques est tout à fait similaire). En particulier si $t_n = n^\alpha$ ($\alpha > 1$), on sait que $t(z) \sim \Gamma(\alpha+1)(1-z)^{-\alpha-1}$, au voisinage de 1, et donc on peut reprendre la chaîne ci-dessus avec $c = \Gamma(\alpha+1)$ et $\beta = -\alpha - 1$. Ces étapes se résument alors ainsi :

$$t_n = n^\alpha \Rightarrow t(z) \sim \Gamma(\alpha+1)(1-z)^{-\alpha-1} \Rightarrow f(z) \sim \Gamma(\alpha+1) \frac{\alpha+1}{\alpha-1} (1-z)^{-\alpha-1} \Rightarrow f_n \sim \frac{\alpha+1}{\alpha-1} n^\alpha$$

Dans l'étape d'intégration, nous avons mis de côté le calcul des constantes. C'est le deuxième cas du Théorème 4 qui nous fournit leur forme :

$$\mathbf{K}[t] = \int_0^1 \left[t'(w)(1-w)^2 - (t'(w)(1-w)^2)_- \right] dw,$$

où f_- représente la somme des éléments singuliers de f avec un exposant strictement inférieur à -1 . Cette constante étant ensuite multipliée par $(1-z)^{-2}$, le terme résultant dans le développement de f_n est donc $\mathbf{K}[t] \cdot (n+1)$.

Dans le cas où t_n croît moins vite que n , alors la partie divergente est absente et alors :

$$\mathbf{K}[t] = \int_0^1 t'(w)(1-w)^2 dw = 2 \sum_{n=1}^{+\infty} \frac{t_n}{(n+1)(n+2)}.$$

Ces considérations nous incitent à définir les constantes suivantes, qui achèvent le remplissage du tableau :

$$K_\alpha = 2 \sum_{n=1}^{+\infty} \frac{n^\alpha}{(n+1)(n+2)}, \quad \text{si } \alpha < 1$$

et

$$K_\alpha = 2 \sum_{n=1}^{+\infty} \frac{n^\alpha - \Gamma(\alpha+1) \binom{n+\alpha}{\alpha}}{(n+1)(n+2)}, \quad \text{si } 1 < \alpha < 2.$$

2.2 Théorèmes de transfert

Afin de déterminer des distributions limites, nous allons utiliser la méthode des moments. Pour cela, nous devons donc chercher des équivalents asymptotiques pour les moments de notre fonctionnelle additive $f(T)$.

Sous le modèle de permutation aléatoire, nous admettrons que la loi jointe de la taille des sous-arbres $(|L_1|, \dots, |L_m|)$ est uniforme sur les $\binom{n}{m-1}$ m -uplets d'entiers positifs ayant pour somme $n - (m-1)$. Posons $\mu_n(k) = \mathbb{E} [X_n^k]$ (avec $X_n = f(T)$ pour $|T| = n$) et introduisons les notations $\sum_{\mathbf{j}}$ pour désigner la somme sur les m -uplets (j_1, \dots, j_m) ayant pour somme $n - (m-1)$ et $\sum_{\mathbf{k}}$ la somme sur les $(m+1)$ -uplets

(k_1, \dots, k_{m+1}) d'entiers positifs ayant pour somme k . Alors, en notant \oplus une somme de variables indépendantes, on a :

$$\begin{aligned}\mu_n(k) &= \mathbb{E}[X_n^k] = \mathbb{E}[\mathbb{E}[X_n^k | |L_1|, \dots, |L_m|]] = \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \mathbb{E}[X_{j_1} \oplus \dots \oplus X_{j_m} + t_n]^k \\ &= \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \sum_{\mathbf{k}} \binom{k}{k_1, \dots, k_m, k_{m+1}} \mu_{j_1}(k_1) \dots \mu_{j_m}(k_m) t_n^{k_{m+1}} \\ &= \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} \binom{n-1-j}{m-2} \mu_j(k) + r_n(k)\end{aligned}$$

avec

$$r_n(k) = \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_m, k_{m+1}} t_n^{k_{m+1}} \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \mu_{j_1}(k_1) \dots \mu_{j_m}(k_m)$$

en notant $\sum_{\mathbf{k}}^*$ la même somme que $\sum_{\mathbf{k}}$ avec la condition supplémentaire que $k_i < k$ pour $i = 1, \dots, m$. Nous avons ainsi établi une relation de récurrence linéaire de la même forme pour tous les moments de X_n , la seule différence provenant du terme $r_n(k)$. Notons d'ailleurs que $r_n(1) = t_n$, c'est-à-dire la suite-test.

Proposition 4 *Sous le modèle de permutation aléatoire, les moments (a_n) d'une fonctionnelle additive vérifient la récurrence suivante*

$$a_n = \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} a_j + b_n, \quad n \geq m-1 \quad (4)$$

avec les conditions initiales $a_j = b_j$, pour $0 \leq j \leq m-2$.

Notre démarche, présentée en introduction, consiste maintenant à considérer les fonctions génératrices $A(z) = \sum_{n \geq 0} a_n z^n$ et $B(z) = \sum_{n \geq 0} b_n z^n$, afin d'en déduire le comportement asymptotique de la suite (a_n) . En fait, la récurrence liant a_n et b_n se transfère en l'égalité suivante :

$$A^{(m-1)}(z) = B^{(m-1)}(z) + m!(1-z)^{-(m-1)} A(z)$$

Le théorème ci-dessous nous explicite la solution de cette équation.

Théorème 8 (Théorème de transfert exact) *Soit A et B les fonctions génératrices des suites (a_n) et (b_n) . Soit*

$$\widehat{B}(z) = B(z) - \sum_{j=0}^{m-2} b_j z^j = \sum_{n=m-1}^{\infty} b_n z^n \quad (5)$$

Alors

$$A(z) = \sum_{j=1}^{m-1} c_j (1-z)^{-\lambda_j} + \sum_{j=1}^{m-1} \frac{(1-z)^{-\lambda_j}}{\psi'(\lambda_j)} \int_0^z B^{(m-1)}(\xi) (1-\xi)^{\lambda_j+m-2} d\xi \quad (6)$$

$$= \sum_{j=1}^{m-1} c_j (1-z)^{-\lambda_j} + \widehat{B}(z) + m! \sum_{j=1}^{m-1} \frac{(1-z)^{-\lambda_j}}{\psi'(\lambda_j)} \int_0^z \widehat{B}(\xi) (1-\xi)^{\lambda_j-1} d\xi \quad (7)$$

où ψ désigne le polynôme caractéristique

$$\psi(\lambda) = \lambda^{\overline{m-1}} - m! = \lambda(\lambda + 1) \cdots (\lambda + m - 2) - m!$$

ayant pour racines $\lambda_1 = 2, \lambda_2, \dots, \lambda_{m-1}$, ordonnées suivant les parties réelles décroissantes. Dans l'équation (7), les coefficients c_1, c_2, \dots, c_{m-1} peuvent être écrits explicitement

$$c_j = \frac{m!}{\psi'(\lambda_j)} \sum_{k=0}^{m-2} b_k \frac{k!}{\lambda_j^{k+1}}$$

Démonstration: Nous ne démontrons pas ici l'égalité (6) : celle-ci fait appel à la résolution d'équations différentielles eulériennes et dépasse largement le cadre de notre propos. Par contre, pour démontrer l'égalité entre (6) et (7), nous remarquons que dans (6), B peut être remplacé par \widehat{B} . Nous utilisons alors plusieurs intégrations par parties et l'Identité 2, produite en annexe. En notant

$$\widehat{A}(z) = A(z) - \sum_{j=1}^{m-1} c_j (1-z)^{\lambda_j},$$

nous trouvons après $m-2$ intégrations par parties

$$\widehat{A}(z) = \sum_{j=1}^{m-1} \frac{(1-z)^{-\lambda_j}}{\psi'(\lambda_j)} (\lambda_j + m - 2) \cdots (\lambda_j + 1) \int_0^z \widehat{B}'(\xi) (1-\xi)^{\lambda_j} d\xi.$$

Mais

$$(\lambda_j + m - 2) \cdots (\lambda_j + 1) = \frac{\lambda_j^{\overline{m-1}}}{\lambda_j} = \frac{\psi(\lambda_j) + m!}{\lambda_j} = \frac{m!}{\lambda_j},$$

et donc

$$\widehat{A}(z) = m! \sum_{j=1}^{m-1} \frac{(1-z)^{-\lambda_j}}{\lambda_j \psi'(\lambda_j)} \int_0^z \widehat{B}'(\xi) (1-\xi)^{\lambda_j} d\xi.$$

Nous obtenons alors (7) en réalisant encore une intégration par parties, et en utilisant l'Identité 1 avec $\lambda = 0$.

Théorème 9 (Théorème de transfert asymptotique)

– Si

$$b_n = o(n) \text{ et } \sum_{n=0}^{\infty} \frac{b_n}{(n+1)(n+2)} \text{ converge,} \quad (8)$$

alors

$$a_n = \frac{K_1}{H_m - 1} n + o(n) \quad (9)$$

où

$$K_1 = \sum_{j=0}^{\infty} \frac{b_j}{(j+1)(j+2)} \quad \text{et} \quad H_m = \sum_{i=1}^m \frac{1}{i} \quad (10)$$

– Si $b_n = K_2(n+1) + h_n$ où (h_n) vérifie (8) (en remplaçant b_n par h_n), alors

$$a_n = \frac{K_2}{H_m - 1} n H_n + \frac{K_3}{H_m - 1} n + o(n) \quad (11)$$

avec

$$K_3 = \sum_{j=0}^{\infty} \frac{h_j}{(j+1)(j+2)} + K_2 \left(\frac{H_m - 1}{2} - 1 + \frac{H_m^{(2)} - 1}{2(H_m - 1)} \right)$$

– Si $b_n = K_4 n^v + o(n^v)$ avec $v > 1$, alors

$$a_n = \frac{K_4}{1 - \frac{m! \Gamma(v+1)}{\Gamma(v+m)}} n^\mu + o(n^v) \quad (12)$$

Pour démontrer ce théorème, nous aurons besoin de plusieurs lemmes. Ceux-ci nous permettront en effet de contrôler les différents termes intervenant dans l'expression de $A(z)$ donnée par le théorème de transfert exact 8.

Lemme 1 Soit $Y(z) = \sum_{n=0}^{\infty} y_n z^n$ avec $y_0 = 0$. Alors,

$$\forall \lambda \in \mathbb{C} \quad [z^n] \left((1-z)^{-\lambda} \int_0^z (1-\xi)^{\lambda-1} Y(\xi) d\xi \right) = \sum_{k=0}^{n-1} \frac{y_k}{k+1} \prod_{j=k+2}^n \left(1 + \frac{\lambda-1}{j} \right). \quad (13)$$

Démonstration: La fonction $W(z) = (1-z)^{-\lambda} \int_0^z (1-\xi)^{\lambda-1} Y(\xi) d\xi$ est l'unique solution vérifiant $W(0) = 0$, à l'équation différentielle

$$W'(z) = \lambda(1-z)^{-1} W(z) + (1-z)^{-1} Y(z).$$

Par conséquent $w_n = [z^n] W(z)$, $n \geq 0$, vérifie $w_0 = 0$ et

$$n w_n = \lambda \sum_{k=0}^{n-1} w_k + \sum_{k=0}^{n-1} y_k, \quad n \geq 1.$$

D'où

$$w_n = \frac{\lambda}{n} \sum_{k=0}^{n-1} w_k + \frac{1}{n} \sum_{k=0}^{n-1} y_k, \quad n \geq 1.$$

Mais cette récurrence se résout aisément :

$$n w_n - (n-1) w_{n-1} = \lambda w_{n-1} + y_{n-1},$$

soit

$$w_n = w_{n-1} \left(1 + \frac{\lambda-1}{n} \right) + \frac{y_{n-1}}{n},$$

ce qui, en itérant, nous démontre le lemme.

Notons d'ores et déjà que le produit dans l'expression (13) peut être précisé , grâce à la formule de Stirling :

$$\prod_{j=k+2}^n \left(1 + \frac{\lambda-1}{j}\right) = \frac{\Gamma(\lambda+n)}{\Gamma(k+\lambda+1)} \frac{\Gamma(2+k)}{\Gamma(1+n)} = \frac{n^{\lambda-1}[1+O(n^{-1})]}{(k+1)^{\lambda-1}[1+O((k+1)^{-1})]} \quad (14)$$

Par ailleurs, dans le cas $\lambda = 2$, le Lemme 1 se résume à :

$$[z^n] \left((1-z)^{-2} \int_0^z (1-\xi) Y(\xi) d\xi \right) = (n+1) \sum_{k=0}^{n-1} \frac{y_k}{(k+1)(k+2)} \quad (15)$$

Lemme 2

– Si $Re(\lambda) < 2$ et $Y(z) = \sum_{n=0}^{\infty} y_n z^n$ vérifie $y_0 = 0$ et $y_n = o(n)$, alors

$$[z^n] \left((1-z)^{-\lambda} \int_0^z (1-\xi)^{\lambda-1} Y(\xi) d\xi \right) = o(n)$$

– Soit \widehat{B} défini en (5), si la condition (8) est vérifiée, alors

$$[z^n] \left((1-z)^{-2} \int_0^z \widehat{B}(\xi) (1-\xi) d\xi \right) = n \sum_{j=m-1}^{+\infty} \frac{b_j}{(j+1)(j+2)} + o(n)$$

Démonstration: Si $Re(\lambda) < 2$,

$$\begin{aligned} [z^n] \left((1-z)^{-\lambda} \int_0^z (1-\xi)^{\lambda-1} Y(\xi) d\xi \right) &= \sum_{k=0}^{n-1} \frac{y_k}{k+1} \prod_{j=k+2}^n \left(1 + \frac{\lambda-1}{j}\right) \\ &= O \left(\sum_{k=0}^{n-1} \frac{|y_k|}{k+1} \exp \left((\Re \lambda - 1) \ln \frac{n}{k+1} \right) \right) \\ &= o \left(n^{\Re \lambda - 1} \sum_{k=1}^n k^{1-\Re \lambda} \right) \\ &= o(n) \end{aligned}$$

Si $\lambda = 2$, reprenons l'égalité (15), en remplaçant Y par \widehat{B} , et en se souvenant que $\widehat{b}_k = 0$ si $0 \leq k \leq m-2$ et $\widehat{b}_k = b_k$ sinon.

Alors

$$\begin{aligned}
[z^n] \left((1-z)^{-2} \int_0^z (1-\xi) \widehat{B}(\xi) d\xi \right) &= (n+1) \sum_{k=m-1}^{n-1} \frac{\widehat{b}_k}{(k+1)(k+2)} \\
&= n \sum_{k=m-1}^{n-1} \frac{b_k}{(k+1)(k+2)} + o(n) \\
&= n \sum_{k=m-1}^{+\infty} \frac{b_k}{(k+1)(k+2)} - n \sum_{k=n}^{+\infty} \frac{b_k}{(k+1)(k+2)} + o(n) \\
&= n \sum_{k=m-1}^{+\infty} \frac{b_k}{(k+1)(k+2)} - n o(1) + o(n) \\
&= n \sum_{k=m-1}^{+\infty} \frac{b_k}{(k+1)(k+2)} + o(n)
\end{aligned}$$

Lemme 3 Soit (b_n) et (b'_n) deux suites telles que

$$|b_n| \leq b'_n \text{ pour tout } n \geq 0$$

alors les suites correspondantes (a_n) et (a'_n) dans (4) vérifient

$$|a_n| \leq a'_n \text{ pour tout } n \geq 0$$

Démonstration: Ce résultat se démontre par une récurrence immédiate.

Démonstration: [du Théorème 9]

Pour le cas $b_n = o(n)$, il suffit de repartir de l'équation (7), et de voir que le coefficient dominant dans le développement de a_n est celui apporté par les termes en $(1-z)^{-\lambda_1} = (1-z)^{-2}$. On a donc :

$$a_n = c_1 n + o(n) + o(n) + \frac{m!}{\psi'(2)} \times \left(n \sum_{k=m-1}^{+\infty} \frac{b_k}{(k+1)(k+2)} + o(n) \right) + o(n),$$

avec $c_1 = \frac{1}{H_m - 1} \sum_{k=0}^{m-2} \frac{b_k}{(k+1)(k+2)}$ et $\psi'(2) = m!(H_m - 1)$ (où $H_m = \sum_{i=1}^m \frac{1}{i}$). Le premier point est donc acquis.

Remarque 1 En fait, les termes $(1-z)^{-\lambda}$ engendrent des coefficients de Taylor de l'ordre de $n^{\Re \lambda - 1}$. Il est donc normal que l'ensemble soit contrôlé par le terme correspondant à $\lambda_1 = 2$. Cela étant, le terme d'erreur peut être précisé si on dispose d'informations sur λ_2 (la racine de ψ ayant la plus grande partie réelle après λ_1). Par exemple, si $2 \leq m \leq 26$, on sait qu'alors $\alpha_2 = \Re \lambda_2 < \frac{3}{2}$ et donc le résultat précédent peut être amélioré ainsi :

Si $2 \leq m \leq 26$ et $b_n = o(\sqrt{n})$ alors

$$a_n = \frac{K_1}{H_m - 1} n + o(\sqrt{n}) \tag{16}$$

Pour le cas $b_n = K_2(n+1) + h_n$, on considère d'abord le cas particulier $b_n = n+1$, soit $B(z) = (1-z)^{-2}$, et donc $B^{(m-1)}(z) = m!(1-z)^{-(m+1)}$. En reportant cette égalité dans l'équation (6), on obtient :

$$a_n = (n+1) \left[c_1 + m! \sum_{j=2}^{m-1} \frac{1}{(2-\lambda_j)\psi'(\lambda_j)} \right] + \frac{m!}{\psi'(2)} [z^n] \left[(1-z)^{-2} \log((1-z)^{-1}) \right] + o(n).$$

Mais alors, d'après les Identités 3 et 4 énoncées en annexe (avec dans ce cas précis $c_1 = 1$) :

$$\begin{aligned} a_n &= (n+1) \left(1 + \frac{1}{2} \left[\frac{H_m^{(2)} - 1}{(H_m - 1)^2} - 1 \right] \right) + \frac{1}{H_m - 1} [(n+1)H_n - n] + o(n) \\ &= \frac{1}{H_m - 1} n H_n + \left[\frac{1}{2} - \frac{1}{H_m - 1} + \frac{H_m^{(2)} - 1}{2(H_m - 1)^2} \right] n + o(n). \end{aligned}$$

Ceci complète la démonstration du deuxième point pour $b_n = n+1$, le cas général s'en déduit grâce au premier point, en utilisant le principe de superposition.

Pour le dernier point, on suppose d'abord que $b_n = (v+1)^{\bar{n}}/n! \sim n^v/\Gamma(v+1)$ c'est-à-dire $B(z) = (1-z)^{-(v+1)}$ et donc $B^{(m-1)}(z) = (v+1)^{\bar{m}-1}(1-z)^{-(v+m)}$. En reportant cette égalité dans l'équation (6), et en utilisant l'Identité 1 avec $\lambda = v+1$, ainsi que l'égalité $\frac{(v+1)^{\bar{m}-1}}{(v+1)^{\bar{m}-1} - m!} = \left[1 - \frac{m!\Gamma(v+1)}{\Gamma(v+m)} \right]^{-1}$, on obtient :

$$A(z) = \left[1 - \frac{m!\Gamma(v+1)}{\Gamma(v+m)} \right]^{-1} (1-z)^{-(v+1)} + O(|1-z|^{-2}).$$

Mais nous savons par ailleurs que $[z^n](1-z)^{-(v+1)} \sim n^v/\Gamma(v+1)$, et ceci achève donc l'étude du cas particulier.

Pour compléter le cas général, il suffit de montrer que si $b_n = o(n^v)$, pour $v > 1$, alors $a_n = o(n^v)$. Pour cela, soit $\varepsilon > 0$; alors il existe une suite (b'_n) telle que $|b_n| \leq b'_n$ pour tout n et $b'_n = \varepsilon(v+1)^{\bar{n}}/n!$ pour n suffisamment grand. On a alors, d'après l'étude du cas particulier :

$$a'_n = \varepsilon' n^v + o(n^v), \quad \text{où } \varepsilon' = \frac{\varepsilon}{\Gamma(v+1)} \left[1 - \frac{m!\Gamma(v+1)}{\Gamma(v+m)} \right]^{-1}.$$

Et donc, d'après le Lemme 3 :

$$\limsup_n |a_n| n^{-v} \leq \varepsilon';$$

et puisque ε (et donc ε') sont arbitraires, ceci achève le cas général.

2.3 Distribution limite de X_n

2.3.1 Normalité asymptotique pour des fonctions-test "faibles"

Théorème 10 (TCL1 pour des fonctions-test "faibles")

Si $2 \leq m \leq 26$ et la suite-test (t_n) vérifie

$$(a) t_n = o(\sqrt{n}) \quad \text{et} \quad (b) \sum_{n>0} n^{-1} \max_{n^\delta \leq k \leq n} \frac{t_k^2}{k} < \infty \quad (17)$$

pour une certaine valeur $\delta \in]0; 1[$, alors l'espérance μ_n et la variance σ_n de la fonctionnelle additive correspondante X_n vérifient respectivement

$$\mu_n = \frac{K_1}{H_m - 1} n + o(\sqrt{n}) = \mu n + o(\sqrt{n}) \quad (18)$$

avec K_1 défini en (10), et

$$\sigma_n^2 = \sigma^2 n + o(n), \quad \text{où } \sigma^2 = \frac{1}{H_m - 1} \sum_{j>0} \frac{r_j}{(j+1)(j+2)} \quad (19)$$

avec la suite (r_n) définie par $r_j = 0$ si $0 \leq j \leq m-2$ et

$$r_n = \frac{1}{\binom{n}{m-1}} \sum_j (t_n + \mu_{j_1} + \dots + \mu_{j_m} - \mu_n)^2 \quad (20)$$

De plus,

$$\frac{X_n - \mu n}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

Démonstration:

Nous allons démontrer la convergence en loi par la méthode des moments. Posons $\widetilde{X}_n = X_n - \mu(n+1)$, $\widetilde{\mu}_n(k) = \mathbb{E} \left[\widetilde{X}_n^k \right]$ et montrons par récurrence sur $k \geq 1$ que

$$\widetilde{\mu}_n(2k) \sim \frac{(2k)!}{2^k k!} \sigma^{2k} n^k \quad (21)$$

$$\widetilde{\mu}_n(2k-1) = o(n^{k-(1/2)}) \quad (22)$$

On observe d'abord que les équations (18) et (19) rendent les équations précédentes vraies pour $k = 1$.

1. Pour démontrer (18), il faut se rappeler que la suite des espérances (μ_n) des variables (X_n) vérifie la relation de récurrence (4) avec b_n remplacé par t_n , et donc que les fonctions génératrices $A(z)$ et $B(z)$ des suites (μ_n) et (t_n) vérifient le théorème de transfert exact. On a alors le résultat sur μ_n comme application directe de la Remarque 1.

2. (19) est une simple application du théorème de transfert asymptotique ; il suffit de vérifier que σ_n vérifie l'équation de récurrence (4), avec b_n remplacé par r_n (ceci s'obtient en passant par la variance conditionnelle) puis que la suite (r_n) vérifie les hypothèses (8).

Lemme 4 *Sous les hypothèses du Théorème Central Limite 1, la suite (r_n) vérifie les hypothèses (8) du théorème de transfert asymptotique.*

Démonstration: Comme $\tilde{\mu}_n = \mu_n - \mu(n+1)$ (avec la notation $\tilde{\mu}_n$ pour $\tilde{\mu}_n(1)$), on peut réécrire r_n sous la forme suivante :

$$r_n = \frac{1}{\binom{n}{m-1}} \sum_j [t_n + \tilde{\mu}_{j1} + \cdots + \tilde{\mu}_{jm} - \tilde{\mu}_n]^2, \quad n \geq m-1$$

avec $\tilde{\mu}_n = o(\sqrt{n})$, d'après (18). Nous allons maintenant utiliser le résultat classique suivant : pour tous réels ξ_1, \dots, ξ_k ,

$$\left[\sum_{i=1}^k \xi_i \right]^2 \leq k \sum_{i=1}^k \xi_i^2. \quad (23)$$

Appliqué à r_n , cela donne

$$\frac{r_n}{m+2} \leq t_n^2 + \tilde{\mu}_n^2 + \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} \binom{n-1-j}{m-2} \tilde{\mu}_j^2$$

et ceci établit la première des deux conditions (8).

Par ailleurs, au vu de cette inégalité, il suffit, pour établir la sommabilité de r_n/n^2 , d'établir celle de $\tilde{\mu}_n^2/n^2$. En effet, une conséquence immédiate de (17) (hypothèses du TCL1) est que $\sum_{n>0} \frac{t_n^2}{n^2} < +\infty$. De plus, nous disposons de l'estimation suivante :

$$\begin{aligned} n^{-2} \sum_{n=m-1}^{\infty} \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} \binom{n-1-j}{m-2} \tilde{\mu}_j^2 &= m(m-1) \sum_{j=0}^{+\infty} \tilde{\mu}_j^2 \sum_{n=j+m-1}^{+\infty} \frac{\binom{n-1-j}{m-2}}{n^2 \binom{n}{m-1}} \\ &\leq m(m-1) \sum_{j=0}^{+\infty} \tilde{\mu}_j^2 \sum_{n=j+m-1}^{+\infty} n^{-3} \\ &= O\left(\sum_{j>0} \frac{\tilde{\mu}_j^2}{j^2}\right) < +\infty. \end{aligned}$$

Pour établir la sommabilité de $\tilde{\mu}_n^2/n^2$, il faut remonter aux deux premiers lemmes intervenant dans la démonstration du théorème de transfert asymptotique, et notamment à l'équation (14), afin d'obtenir un développement de $\tilde{\mu}_n$:

$$\begin{aligned} \tilde{\mu}_n &= O(n^{\alpha_2-1}) + t_n - \frac{1}{H_m-1} (n+1) \sum_{k=n}^{\infty} \frac{\hat{t}_k}{(k+1)(k+2)} \\ &\quad + \sum_{j=2}^{m-1} O\left(n^{\alpha_j-1} \sum_{k=0}^{n-1} \frac{|\hat{t}_k|}{(k+1)^{\alpha_j}}\right), \end{aligned}$$

avec $\alpha_j = \Re \lambda_j$ (en particulier $\alpha_2 < 3/2$, puisque $m \leq 26$) et $\hat{t}_k = t_k$ si $k \geq m-1$ et 0 sinon. En se souvenant de l'inégalité classique (23), il suffit donc d'établir la

sommabilité du carré de chacun des quatre termes, divisé par n^2 . Pour le premier terme, le résultat est immédiat ; pour le deuxième, nous avons déjà énoncé le résultat comme conséquence des hypothèses (17) du TCL1. Pour le troisième, nous utilisons l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} \left[\sum_{k=n}^{+\infty} \frac{\widehat{t}_k}{(k+1)(k+2)} \right]^2 &\leq \frac{1}{n} \left[\sum_{k=n}^{+\infty} \sqrt{n} \frac{\sqrt{k}}{(k+1)(k+2)} \frac{|t_k|}{\sqrt{k}} \right]^2 \\ &= O \left(\frac{1}{n} \sum_{k=n}^{+\infty} \sqrt{n} k^{-3/2} \frac{t_k^2}{k} \right) = O \left(n^{-1/2} \sum_{k=n}^{+\infty} k^{-5/2} t_k^2 \right), \end{aligned}$$

et donc

$$\begin{aligned} \sum_{n>0} \left[\sum_{k=n}^{+\infty} \frac{\widehat{t}_k}{(k+1)(k+2)} \right]^2 &= O \left(\sum_{n>0} n^{-1/2} \sum_{k=n}^{+\infty} k^{-5/2} t_k^2 \right) \\ &= O \left(\sum_k k^{-5/2} t_k^2 k^{1/2} \right) = O \left(\sum_{k>0} \frac{t_k^2}{k^2} \right) < +\infty \end{aligned}$$

Remarquons ici que si $m = 2$ la démonstration est finie (le quatrième terme est nul).

Pour ce dernier, dans le cas général, il suffit (toujours d'après (23)) d'établir la sommabilité de

$$n^{2\rho-4} \left[\sum_{k=1}^{n-1} \frac{|t_k|}{k^\rho} \right]^2$$

pour tout réel $\rho < 3/2$. Pour ceci, nous séparons la somme en $\sum_{k < n^\delta}$ et $\sum_{n^\delta \leq k < n}$ et faisons une fois de plus appel à l'inégalité (23).

Pour la somme $\sum_{k < n^\delta}$ nous utilisons seulement $t_k = O(\sqrt{k})$ et nous remarquons que :

$$n^{2\rho-4} \left[\sum_{k < n^\delta} O(k^{(1/2)-\rho}) \right]^2 = O \left(n^{2\rho-4} (n^\delta)^{3-2\rho} \right) = O(n^\tau)$$

avec $\tau < -1$.

Pour la somme $\sum_{n^\delta \leq k < n}$ nous utilisons à nouveau l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} n^{2\rho-4} \left[\sum_{n^\delta \leq k < n} \frac{|t_k|}{k^\rho} \right]^2 &= n^{2\rho-4} n^{3-2\rho} \left[\sum_{n^\delta \leq k < n} \frac{k^{(1/2)-\rho}}{n^{(3/2)-\rho}} \frac{|t_k|}{k^{1/2}} \right]^2 \\ &= O \left(n^{-1} \sum_{n^\delta \leq k < n} \frac{k^{(1/2)-\rho}}{n^{(3/2)-\rho}} \frac{t_k^2}{k} \right) = O \left(n^{-1} \max_{n^\delta \leq k < n} \frac{t_k^2}{k} \right), \end{aligned}$$

qui est sommable d'après l'hypothèse (17). Ceci achève la démonstration de notre lemme.

Démonstration: [fin de la démonstration du Théorème 9]

3. De la même manière que nous avions établi l'équation de récurrence fondamentale (4), on établit que la suite $(\tilde{\mu}_n(k))$ vérifie :

$$\tilde{\mu}_n(k) = \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-(m-1)} \binom{n-1-j}{m-2} \tilde{\mu}_j(k) + r_n(k)$$

avec

$$r_n(k) = \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_m, k_{m+1}} t_n^{k_{m+1}} \times \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \tilde{\mu}_{j_1}(k_1) \cdots \tilde{\mu}_{j_m}(k_m) \quad (24)$$

où $\sum_{\mathbf{k}}^*$ désigne la même somme que $\sum_{\mathbf{k}}$ avec la condition supplémentaire que $k_i < k$ pour $i \in \{1, \dots, m\}$. On remarque d'ailleurs que l'équation vérifiée par $\tilde{\mu}_n(k)$ est de la forme (4). Nous pourrons donc appliquer le théorème de transfert asymptotique après avoir évalué $r_n(k)$ asymptotiquement. Soit donc $k \geq 2$ fixé et supposons (21) – (22) établis pour tous les entiers inférieurs strictement à k . On calcule alors $r_n(2k)$ en isolant la somme sur tous les m -uples (k_1, \dots, k_m) d'entiers positifs, tous inférieurs strictement à k , et de somme k , notée $\sum_{\mathbf{k}}^{**}$. Ainsi,

$$\begin{aligned} r_n(2k) &= o(n^k) + \sum_{\mathbf{k}}^{**} \binom{2k}{2k_1, \dots, 2k_m} \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \tilde{\mu}_{j_1}(2k_1) \cdots \tilde{\mu}_{j_m}(2k_m) \\ &= o(n^k) + \sum_{\mathbf{k}}^{**} \binom{2k}{2k_1, \dots, 2k_m} \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \frac{(2k_1)!}{2^{k_1} k_1!} \sigma^{2k_1} j_1^{k_1} \cdots \frac{(2k_m)!}{2^{k_m} k_m!} \sigma^{2k_m} j_m^{k_m} \\ &= o(n^k) + \frac{(2k)!}{2^k k!} \sigma^{2k} n^k \sum_{\mathbf{k}}^{**} \binom{k}{k_1, \dots, k_m} \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \left(\frac{j_1}{n}\right)^{k_1} \cdots \left(\frac{j_m}{n}\right)^{k_m} \end{aligned}$$

Mais

$$\begin{aligned} \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \left(\frac{j_1}{n}\right)^{k_1} \cdots \left(\frac{j_m}{n}\right)^{k_m} &\xrightarrow{n \rightarrow \infty} \int_U x_1^{k_1} \cdots x_{m-1}^{k_{m-1}} (1 - x_1 - \cdots - x_{m-1})^{k_m} dx_1 \cdots dx_{m-1} \\ &= (m-1)! \frac{\Gamma(k_1+1) \cdots \Gamma(k_m+1)}{\Gamma(k+m)} = \frac{1}{\binom{k}{k_1, \dots, k_m} \binom{k+m-1}{m-1}} \end{aligned}$$

avec $U = \{(x_1, \dots, x_{m-1}) \in [0, 1]^{m-1} / \sum_{i=1}^{m-1} x_i \leq 1\}$. Puisque le nombre de termes intervenant dans $\sum_{\mathbf{k}}^{**}$ est $\binom{k+m-1}{m-1} - m$, nous avons donc :

$$\begin{aligned} r_n(2k) &= \frac{(2k)!}{2^k k!} \sigma^{2k} n^k \frac{\binom{k+m-1}{m-1} - m}{\binom{k+m-1}{m-1}} + o(n^k) \\ &= \frac{(2k)!}{2^k k!} \sigma^{2k} n^k \left[1 - \frac{m! \Gamma(k+1)}{\Gamma(k+m)} \right] + o(n^k). \end{aligned}$$

On démontre de même que $r_n(2k-1) = o(n^{k-(1/2)})$, ce qui achève la démonstration.

Le théorème suivant examine le cas extrême où $t_n \sim \sqrt{n}$, ou bien $t_n \sim \sqrt{n}L(n)$ avec L une fonction à variation lente.

Théorème 11 (TCL2 pour des fonctions-test "faibles")

Si $2 \leq m \leq 26$ et la suite-test (t_n) vérifie

$$t_n \sim \sqrt{n}L(n)$$

où L est une fonction à variation lente, alors l'espérance μ_n de la fonctionnelle additive correspondante X_n vérifie

$$\mu_n = \frac{K_1}{H_m - 1}n - \frac{(\frac{3}{2})^{\overline{m-1}}}{m! - (\frac{3}{2})^{\overline{m-1}}} \sqrt{n}L(n) + o(\sqrt{n}L(n)) \quad (25)$$

avec K_1 défini en (10).

Si $\sum_{k>0} \frac{L^2(k)}{k} < +\infty$, alors la variance σ_n^2 vérifie (19)-(20), et on pose alors $s^2(n) = \sigma^2 n$.

Si $\sum_{k>0} \frac{L^2(k)}{k} = +\infty$, alors $\sigma_n^2 \sim s^2(n) = \sigma^2 n \sum_{k \leq n} \frac{L^2(k)}{k}$, où nous définissons

$$\sigma^2 = \frac{\left(\left(\frac{3}{2}\right)^{\overline{m-1}}\right)^2 \left(\frac{\pi}{4}(m-1) + 1\right) - (m!)^2}{(H_m - 1) \left(m! - \left(\frac{3}{2}\right)^{\overline{m-1}}\right)^2}$$

De plus, dans les deux cas

$$\frac{X_n - \mu_n}{s(n)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Démonstration: Cette démonstration est assez similaire à la précédente, mais amène à énoncer diverses variantes des lemmes rencontrés jusque-ici, et donc à alourdir le propos. Nous l'admettrons donc.

Remarque 2 Le deuxième cas de ce théorème a le mérite de nous fournir explicitement la variance limite : dans le cas $m = 2$, celle-ci est de $\frac{9}{2}\pi - 14$.

Remarque 3 Dans le cas $m \geq 27$, la recherche des coefficients de Taylor, à partir du théorème de transfert, nous amène à la conclusion négative qu'en général, des phénomènes de périodicité apparaissent pour $\tilde{\mu}_n$ et donc qu'il n'y a pas de distribution limite naturelle pour une quelconque normalisation de X_n . Chern & Hwang [1] fournissent des exemples de telles situations.

2.3.2 Fonctions-test modérées et élevées

Ces deux derniers théorèmes nous affirment l'existence d'une distribution limite dans le cas $t_n \sim n^\beta$ avec $\beta > 0.5$. Plus précisément, nous allons voir que si cette distribution existe pour n'importe quelle valeur de m si $\beta > 1$, elle n'existe (en général) que pour m inférieur à un certain entier m_0 si $0.5 < \beta < 1$.

Théorème 12 Si la suite-test (t_n) vérifie

$$t_n \sim n^\beta L(n) \text{ avec } 0.5 < \beta < 1$$

avec L une fonction à variation lente et $\alpha_2 < 1 + \beta$, alors l'espérance de la fonctionnelle additive X_n sur les arbres m -aires de recherche, sous le modèle de permutation aléatoire, vérifie

$$\mu_n = \mu n - \frac{(1 + \beta)^{\overline{m-1}}}{m! - (1 + \beta)^{\overline{m-1}}} n^\beta + o(n^\beta L(n)), \quad \mu = \frac{K_1}{H_m - 1}$$

où K_1 est défini en (10). De plus,

$$\frac{X_n - \mu n}{n^\beta L(n)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y_\beta,$$

où Y_β est une variable aléatoire non gaussienne, et uniquement déterminée par β .

Remarque 4 Nous avons déjà vu que $\alpha_2 = \Re(\lambda_2) < \frac{3}{2}$ si $m \leq 26$. Mais, par ailleurs, nous savons que α_2 augmente avec m . Ainsi, pour une valeur $\beta \in]1/2; 1[$ fixée, la condition $\alpha_2 < 1 + \beta$ équivaut à $m \leq m_0$ pour un certain $m_0 \geq 26$.

Démonstration: Nous reprenons les notations $\tilde{X}_n = X_n - \mu(n + 1)$, $\tilde{\mu}_n = \mu_n - \mu(n + 1)$ et $\tilde{\mu}_n(k) = \mathbb{E} \left[\tilde{X}_n^k \right]$. Nous savons alors que :

$$\begin{aligned} \tilde{\mu}_n &= O(n^{\alpha_2-1}) + t_n - \frac{1}{H_m - 1}(n + 1) \sum_{k=n}^{\infty} \frac{\hat{t}_k}{(k + 1)(k + 2)} \\ &\quad + m! \sum_{j=2}^{m-1} \frac{1}{\psi'(\lambda_j)} [z^n] \left((1 - z)^{-\lambda_j} \int_0^z \hat{B}(\xi) (1 - \xi)^{\lambda_j - 1} d\xi \right) \end{aligned}$$

Nous obtenons alors le terme en $n^\beta L(n)$ dans le développement de $\tilde{\mu}_n$ grâce aux Lemmes 1 et 2 et aux Identités 1 et 3 (cf annexes) :

$$1 - \frac{1}{(1 - \beta)(H_m - 1)} + m! \sum_{j=2}^{m-1} \frac{1}{((1 + \beta) - \lambda_j) \psi'(\lambda_j)} = -\frac{(1 + \beta)^{\overline{m-1}}}{m! - (1 + \beta)^{\overline{m-1}}}.$$

Pour établir la convergence en loi, nous allons montrer par récurrence sur k que $\tilde{\mu}_n(k)$ vérifie, pour une certaine suite (g_k) ,

$$\tilde{\mu}_n(k) = g_k n^{k\beta} L^k(n) + o(n^{k\beta} L^k(n)). \quad (26)$$

L'affirmation est vraie pour $k = 1$ avec

$$g_1 = -\frac{(1 + \beta)^{\overline{m-1}}}{m! - (1 + \beta)^{\overline{m-1}}} = \left(1 - \frac{m! \Gamma(\beta + 1)}{\Gamma(\beta + m)} \right)^{-1}.$$

On suppose donc l'égalité vraie pour tout entier strictement inférieur à k , et on reprend l'expression (24) de $r_n(k)$:

$$\begin{aligned}
r_n(k) &= o(n^{k\beta} L^k(n)) \\
&+ \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_m, k_{m+1}} (n^\beta L(n))^{k_{m+1}} \\
&\times \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} g_{k_1}(j_1^\beta L(j_1))^{k_1} \cdots g_{k_m}(j_m^\beta L(j_m))^{k_m} \\
&= o(n^{k\beta} L^k(n)) \\
&+ \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_m, k_{m+1}} (n^\beta L(n))^k g_{k_1} \cdots g_{k_m} \\
&\times \frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \left(\frac{j_1}{n} \right)^{k_1\beta} \cdots \left(\frac{j_m}{n} \right)^{k_m\beta} \frac{L^{k_1}(j_1) \cdots L^{k_m}(j_m)}{L^{k_1+\cdots+k_m}(n)}.
\end{aligned}$$

Mais, en définissant $\overline{B}(x_1, \dots, x_m) = \frac{\prod_{i=1}^m \Gamma(x_i)}{\Gamma(\sum_{i=1}^m x_i)}$, on peut démontrer que

$$\frac{1}{\binom{n}{m-1}} \sum_{\mathbf{j}} \left(\frac{j_1}{n} \right)^{k_1\beta} \cdots \left(\frac{j_m}{n} \right)^{k_m\beta} \frac{L^{k_1}(j_1) \cdots L^{k_m}(j_m)}{L^{k_1+\cdots+k_m}(n)} \xrightarrow[n \rightarrow \infty]{} (m-1)! \overline{B}(k_1\beta+1, \dots, k_m\beta+1),$$

et donc

$$\begin{aligned}
r_n(k) &= o(n^{k\beta} L^k(n)) \\
&+ n^{k\beta} L^k(n) (m-1)! \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_m, k_{m+1}} g_{k_1} \cdots g_{k_m} \overline{B}(k_1\beta+1, \dots, k_m\beta+1).
\end{aligned}$$

En utilisant le troisième point du théorème de transfert asymptotique 9 (légèrement modifié par la présence de $L(n)$) avec $v = k\beta$, nous obtenons :

$$\begin{aligned}
\tilde{\mu}_n(k) &= o(n^{k\beta} L^k(n)) \\
&+ n^{k\beta} L^k(n) \frac{(m-1)!}{1 - \frac{m! \Gamma(k\beta+1)}{\Gamma(k\beta+m)}} \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_m, k_{m+1}} g_{k_1} \cdots g_{k_m} \overline{B}(k_1\beta+1, \dots, k_m\beta+1).
\end{aligned}$$

Donc, en définissant récursivement g_k par

$$g_k = \frac{(m-1)!}{1 - \frac{m! \Gamma(k\beta+1)}{\Gamma(k\beta+m)}} \sum_{\mathbf{k}}^* \binom{k}{k_1, \dots, k_{m+1}} g_{k_1} \cdots g_{k_m} B(k_1\beta+1, \dots, k_m\beta+1),$$

avec $g_0 = 1$, nous constatons que l'égalité (26) est encore vraie au rang k . Nous admettrons ce dernier lemme, très technique :

Lemme 5 *Les réels g_k sont les moments d'une unique distribution.*

Notons cette distribution Y_β , alors par la méthode des moments, on peut affirmer que

$$\frac{X_n - \mu n}{n^\beta L(n)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y_\beta.$$

Remarque 5 *Sans donner plus d'explications sur son origine, Kapur [5] introduit une variable Y définie par*

$$Y \stackrel{\mathcal{L}}{=} \sum_{j=1}^m S_j^\beta Y_j + 1, \quad \text{pour } \beta > 0.5, \beta \neq 1$$

où les (Y_j) sont des copies indépendantes de Y et où (S_1, \dots, S_m) est uniformément réparti sur le $(m-1)$ -simplexe : $\{(s_1, \dots, s_m) \geq 0 \text{ et } \sum_{j=1}^m s_j = 1\}$. Il démontre alors que cette variable Y est unique, puis que ses moments sont précisément les moments g_k du lemme précédent. Par conséquent, la distribution de cette variable Y est la distribution limite de notre théorème.

Remarque 6 *Comme dans le cas des suites-test faibles, si $\alpha_2 > 1 + \beta$ (donc si $m > m_0$), on arrive (en général) à la conclusion négative que des phénomènes de périodicité apparaissent pour $\tilde{\mu}_n$.*

On démontre de manière tout à fait semblable le théorème suivant, pour des suites-test “élevées”.

Théorème 13 *Si la suite-test (t_n) vérifie*

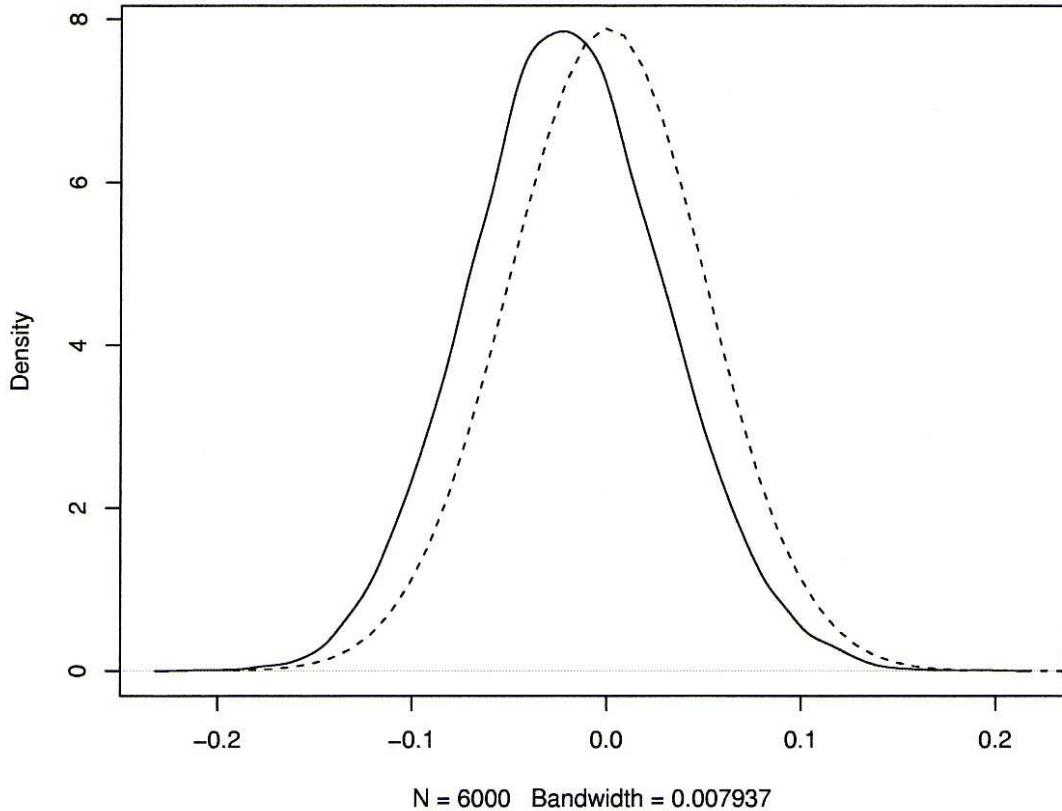
$$t_n \sim n^\beta L(n) \text{ avec } \beta > 1$$

avec L une fonction à variation lente, alors la fonctionnelle additive X_n sur les arbres m -aires de recherche, sous le modèle de permutation aléatoire, vérifie

$$\frac{X_n}{n^\beta L(n)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y_\beta$$

où Y_β est une variable aléatoire non gaussienne, et uniquement déterminée par β .

2.4 Simulations dans le cas $m = 2$

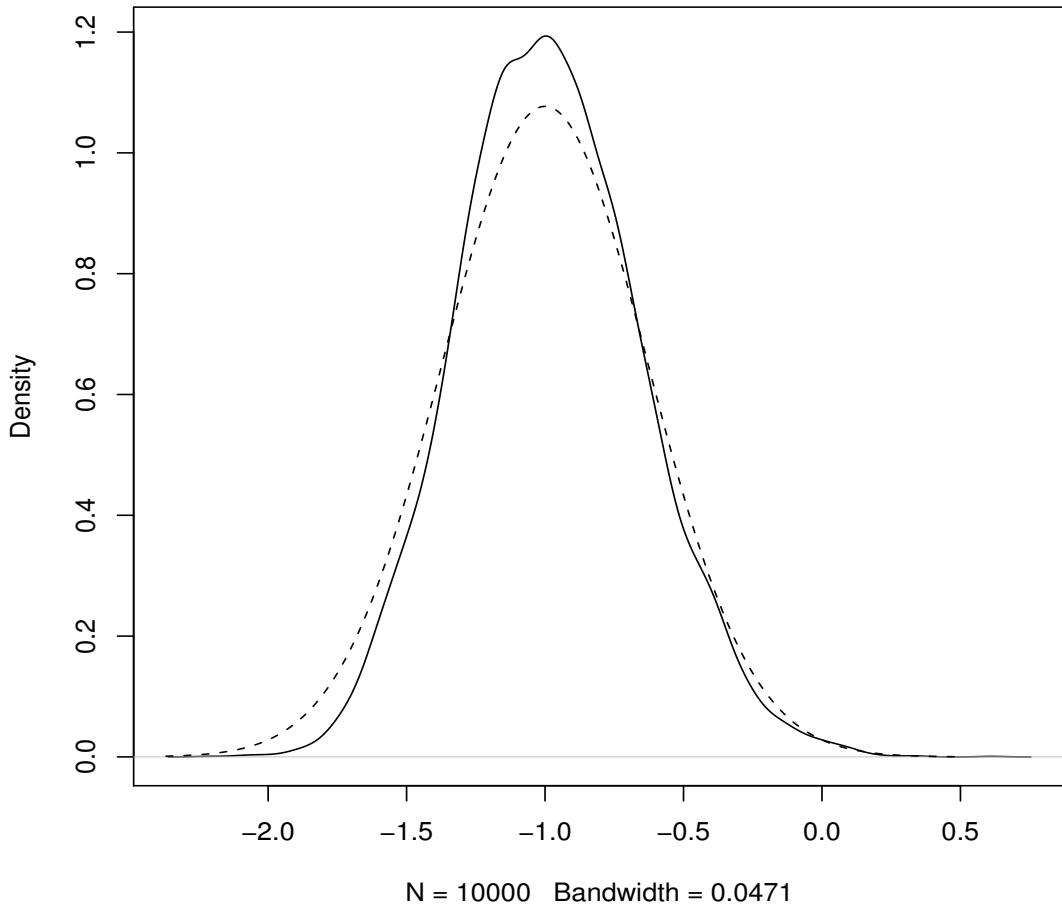


Densité de $\frac{X_n - \mu n}{\sqrt{n}}$ pour $n = 10000$

Dans cet exemple, nous avons cherché à illustrer le théorème central limite 10. Pour cela, nous avons simulé des arbres binaires de recherche à 10000 clés à partir de permutations aléatoires (suivant un modèle uniforme) de $\{1, 2, \dots, 10000\}$. Nous avons alors calculé explicitement, pour chaque arbre, la valeur de la fonctionnelle additive correspondante, dans le cas $t_n = n^{0.1}$. Sur un échantillon de 6000 tirages, nous pouvons constater la convergence de $\frac{X_n - \mu n}{\sqrt{n}}$ vers la loi normale, ici représentée en pointillés. Par ailleurs, l'écart constaté entre la moyenne de $\frac{X_n - \mu n}{\sqrt{n}}$ et la moyenne nulle de la normale centrée est de 0.02 précisément. Ce résultat est à rapprocher du développement asymptotique proposé dans le Théorème 7. En effet

$$\frac{1}{\sqrt{10000}} \left(\frac{0.1+1}{0.1-1} 10000^{0.1} + 2 \sum_{n \geq 0} \frac{n^{0.1}}{(n+1)(n+2)} \right) \approx -0.0193$$

densité de X_n normalisée pour $n=10000$ et $t_n=n^{0.5}$



Cet exemple illustre le théorème central limite 11. Il correspond au cas limite $t_n = \sqrt{n}$, pour lequel nous avons encore convergence vers une distribution gaussienne.

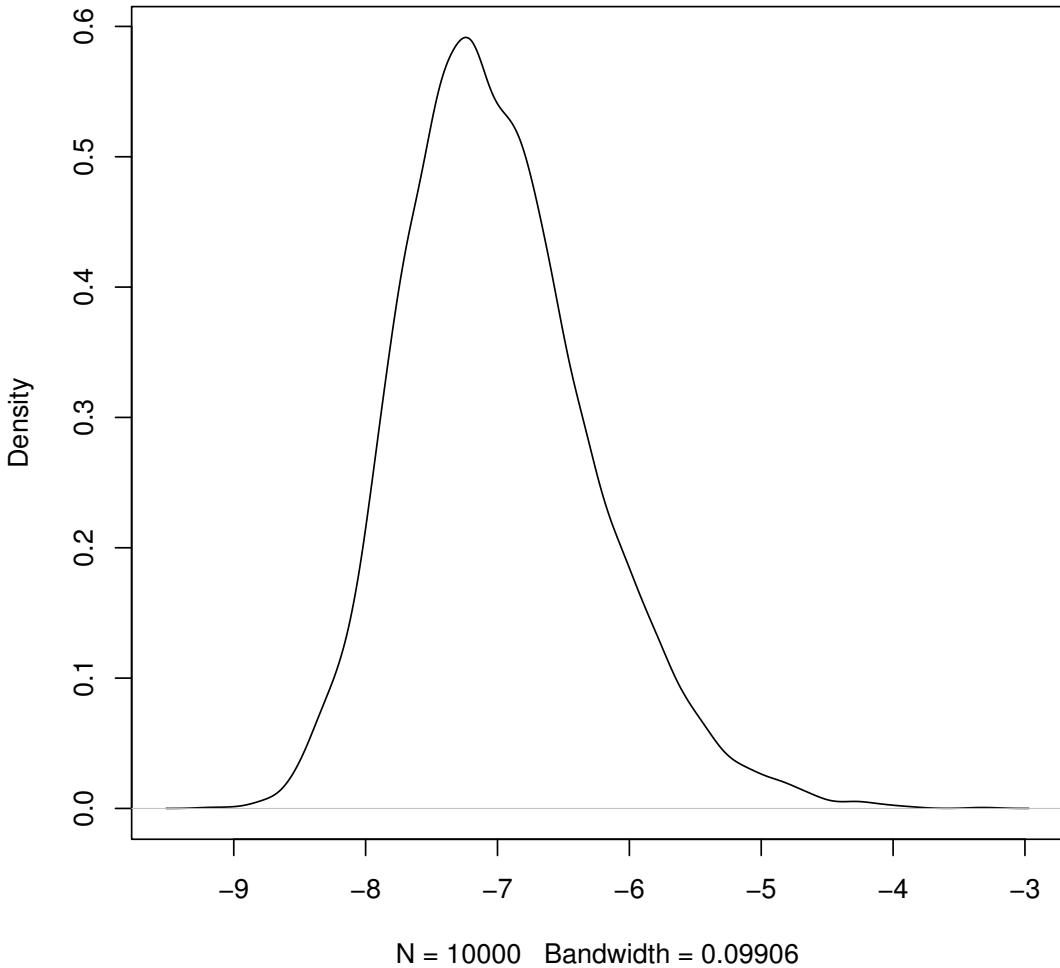
En ligne continue est représentée la densité de $\frac{X_n - \mu n}{\sqrt{n \sum_{i=1}^n \frac{1}{i}}}$, et en pointillés la densité

de la loi normale (non centrée, mais de moyenne -1) de variance $9\pi/2 - 14$. Nous avons préféré décaler la gaussienne pour pouvoir comparer le plus justement possible les courbes. L'échantillon simulé est de taille 10000 et est calculé pour des arbres de taille 10000. Comme dans l'exemple précédent, il est intéressant de regarder les écarts à la moyenne limite, théoriquement nulle. En effet, la moyenne pour l'échantillon simulé est de -0.942 (arrondi à 10^{-3}). Or, en appliquant le développement proposé par l'équation (25), nous avons le résultat :

$$\frac{1}{\sqrt{10000}} \left(\frac{0.5 + 1}{0.5 - 1} 10000^{0.5} \right) \approx -0.959,$$

Il est à noter que dans ce théorème limite, la vitesse de convergence n'est pas en \sqrt{n} , mais en $\sqrt{n \ln n}$.

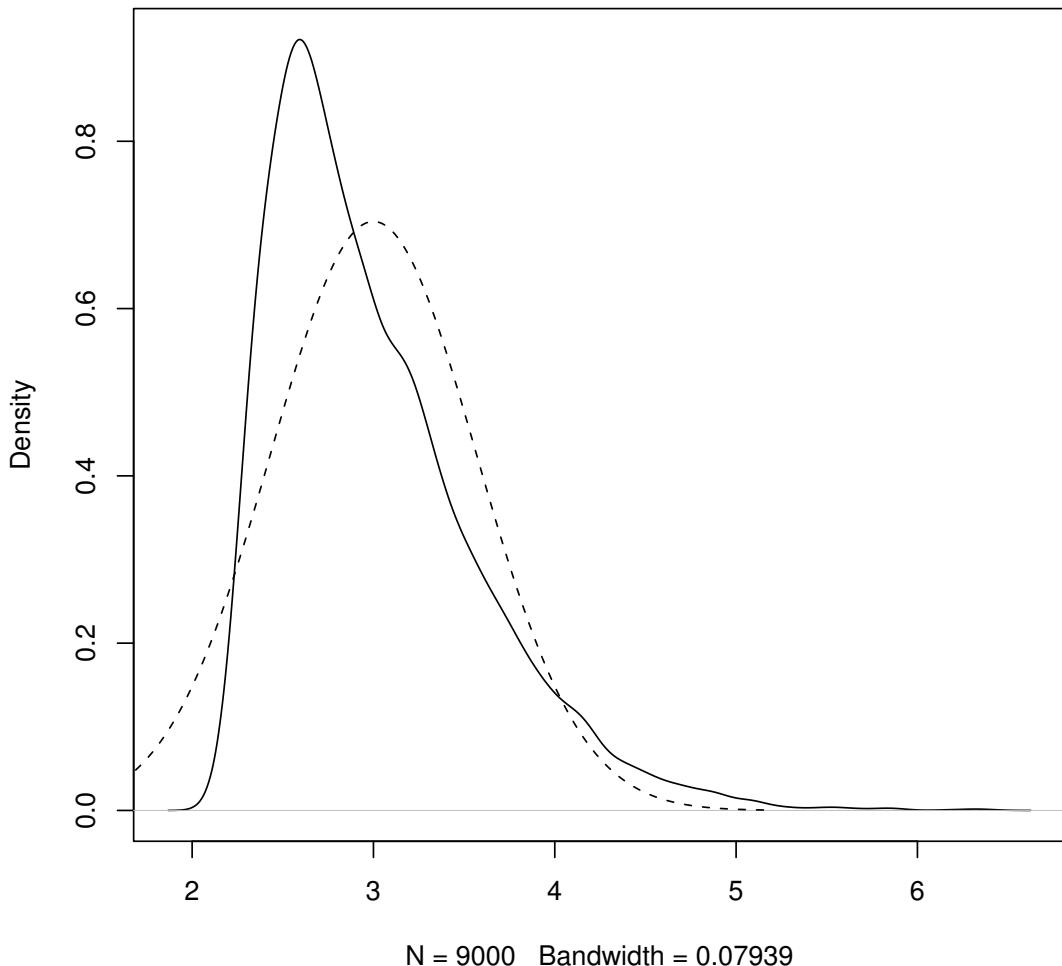
densité de X_n normalisée pour $n=10000$ et $t_n=n^{0.75}$



Ce troisième exemple illustre le Théorème 12, c'est-à-dire le cas où $t_n \sim n^\beta$, pour $0.5 < \beta < 1$ (ici $\beta = 0.75$). La distribution limite de $\frac{X_n - \mu n}{n^\beta}$, dans ce cas de figure, n'étant définie que par ses moments, il est très difficile d'obtenir une représentation correcte de sa densité théorique. La méthode de développement sur une base de polynômes de Legendre amène à des résultats très peu exploitables, et en tout cas bien plus mauvais que la densité obtenue pour un échantillon de taille 10000. On peut ainsi se satisfaire d'une moyenne estimée de -6.99 (à comparer à une moyenne théorique pour la variable $Y_{0.75}$ valant $g_1 = -7$) et d'une variance estimée de 0.519 (à comparer avec une variance théorique valant $g_2 - g_1^2 \approx 0.540$).

Par ailleurs, en posant $Z_n = \frac{X_n - \mu n}{n^\beta}$, et en regardant les moments centrés d'ordre impair de Z_n , on peut facilement constater que Z_n n'est pas gaussienne. En effet, les moments centrés d'ordre 3, 5 et 7 de Z_n valent 0.23, 1.28 et 8.33, alors qu'ils devraient être nuls si Z_n suivait une loi normale.

densité de X_n normalisée pour $n=10000$ et $t_n=n^2$



Ce dernier exemple illustre le Théorème 13, c'est-à-dire le cas où $t_n \sim n^\beta$, pour $\beta > 1$. Dans ce cas de figure, la densité de la distribution limite Y_β de $\frac{X_n}{n^\beta}$ pose le même problème de représentation que précédemment. Néanmoins, là encore, on peut se satisfaire de la moyenne de l'échantillon (dont la densité simulée est en ligne continue) valant 2.998 (à comparer à la valeur théorique $g_1 = 3$), et de sa variance 0.321 (à comparer avec une variance théorique de $g_2 - g_1^2 = \frac{1}{3}$). La loi normale de moyenne 3 et de variance $\frac{1}{3}$, ici dessinée en pointillés, est là pour se convaincre du caractère non-gaussian de la distribution limite.

Perspectives et conclusions

Il est rassurant de constater que nous retrouvons, via des outils totalement différents, des résultats bien connus sur le comportement d’algorithme récursif, par exemple le cas du Quicksort [6] (correspondant à $t_n = n$), algorithme de tri rapide d’une liste, dont on sait que sa complexité moyenne est en $n \ln n$, terme que l’on retrouve dans le développement obtenu au Théorème 7.

Plus généralement, les techniques d’analyse de singularités sont, on l’a vu, un outil très puissant pour relier le comportement d’une fonction complexe au voisinage de sa singularité, avec le comportement de ses coefficients de Taylor, en l’occurrence des coûts de complexité moyens. Néanmoins, si le schéma de raisonnement proposé par les travaux de Kapur semble assez générique, sa mise en application dépend fondamentalement du modèle de probabilité sous lequel “vivent” nos arbres. Par exemple, l’étude des moments et des lois limites sous le modèle uniforme conduit à des distributions qui ne sont plus nécessairement gaussiennes (comme c’était le cas pour notre étude avec $\alpha \leq 0.5$), mais uniquement déterminées par leurs moments, comme pour notre variable Y_β . De plus, un point intéressant de son étude est que la normalisation à effectuer est en $\frac{1}{n^\alpha}$, quelle que soit la valeur de α (différente de 0.5). Par ailleurs, cette étude du modèle uniforme est sensiblement plus technique et fait appel, notamment, aux développements asymptotiques des fonctions polylogarithmes Li_α et de leurs dérivées successives.

Dans notre cas, pour revenir au contexte initial présenté en introduction, celui des puces à ADN, et de la modélisation par des arbres aléatoires, cet aspect pose le problème sous-jacent des estimateurs choisis pour déterminer cette loi de probabilité. L’équipe de bio-informatique de Lille 2 travaille actuellement à la détermination de ces estimateurs (dans le cadre d’arbres m -aires, avec $m > 2$), mais une fois ce travail achevé, il restera à regarder comment se comportent nos équations de récurrence reliant f_n et t_n , selon que l’on se situe sous une loi estimée ou sous la loi “réelle”.

A Propriétés du polynôme caractéristique $\psi(\lambda)$

Le polynôme ψ est défini par $\psi(\lambda) = \lambda^{\overline{m-1}} - m! = \lambda(\lambda+1)\cdots(\lambda+m-2) - m!$. Il a pour racines racines $\lambda_1 = 2, \lambda_2, \dots, \lambda_{m-1}$, ordonnées suivant les parties réelles décroissantes.

Ce polynôme intervient dans la résolution de l'équation différentielle :

$$A^{(m-1)}(z) = B^{(m-1)}(z) + m!(1-z)^{-(m-1)}A(z).$$

Les différentes identités utilisées dans ce mémoire concernant ce polynôme sont résumées ici.

Identité 1 Si $\lambda \notin \{\lambda_1, \dots, \lambda_{m-1}\}$,

$$\sum_{j=1}^{m-1} \frac{1}{(\lambda - \lambda_j)\psi'(\lambda_j)} = \frac{1}{\psi(\lambda)}$$

Soit r et n des entiers naturels non nuls, nous notons $H_n^{(r)}$ le nombre harmonique d'ordre r

$$H_n^{(r)} = \sum_{j=1}^n \frac{1}{j^r}.$$

Pour $r = 1$ nous notons $H_n = H_n^{(1)}$ le nombre harmonique usuel.

Identité 2 Pour $0 \leq k \leq m-3$,

$$\sum_{j=1}^{m-1} \frac{\lambda_j^k}{\psi'(\lambda_j)} = 0.$$

Identité 3

$$\psi'(2) = m!(H_m - 1) \quad \text{and} \quad \psi''(2) = m![(H_m - 1)^2 - (H_m^{(2)} - 1)].$$

Identité 4

$$\sum_{j=2}^{m-1} \frac{1}{(\lambda_j - 2)\psi'(\lambda_j)} = \frac{1}{2(m!)} \left[1 - \frac{H_m^{(2)} - 1}{(H_m - 1)^2} \right].$$

B Programmes utilisés pour les simulations

Les programmes utilisés pour la simulation ont été réalisés en langage R.

Création d'un arbre à partir d'une liste

```
Arbre <- function(x)
{ z<-list(racine=c(0,0),gauche=c(),droit=c())
  z $ racine=c(x[1],length(x))
  z1=x[which(x<x[1])]
  z2=x[which(x>x[1])]
  if (length(z1) !=0)
    z $gauche=Arbre(z1)
  if (length(z2) !=0)
    z $droit=Arbre(z2)
  z
}
```

Création d'une permutation aléatoire de taille 10000

```
Permut <- function(a)
{
  for (i in 1 :10000)
  {
    t=floor(runif(1,1,i+1))
    v=a[t]
    a[t]=a[i]
    a[i]=v
  }
  end
  a
}
```

Calcul de la complexité X_n pour un arbre et une suite (t_n) donnés

Dans ce programme, l'entrée à fournir est un arbre.

```
Complexity <- function(a)
{
  if (length(a$racine)==2)
    o=Complexity(a$gauche)+Complexity(a$droit)+(a$racine[2]) ^ 0.5
  o
}
```

Il va de soi que la puissance 0.5 peut être remplacée par n'importe quelle puissance α .

Programme standard pour obtenir un échantillon de taille 5000 construit à partir d'arbres à 10000 clés

```
w=c()
for (i in 1 :5000)
  w=c(w,Complexity(Arbre(Permut(1 :10000))))
```

Références

- [1] CHERN, H.-H., ET HWANG, H.-K. Phase changes in random m -ary search trees and generalized Quicksort. *Random Structure algorithms*, 19(3-4) :316-358, 2001. Analysis of algorithms (Kynica Morska, 2000).
- [2] FILL, J.A., FLAJOLET, P. ET KAPUR, N. Singularity analysis, Hadamard products, and Tree recurrences, 2003, arXiv :math.CO/0306225.
- [3] FLAJOLET, P., ET ODLYZKO, A.M. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2) :216-240, 1990.
- [4] HENRICI, P. *Applied and computational complex analysis. Vol. 2.* Wiley Interscience [John Wiley & Sons], New York, 1977. Special functions-integral transforms-asymptotics-continued fractions.
- [5] KAPUR, N. Additive functionals on random search trees. PhD thesis, The Johns Hopkins University, 2003.
- [6] SEDGEWICK, R., ET FLAJOLET, P. *An Introduction to the Analysis of Algorithms.* Addison Wesley (1996)
- [7] TRANG, T. K-automates et modèles de Markov cachés : application à l'analyse des puces à ADN. Mémoire de D.E.A. en Mathématiques Appliquées, Lille, 2003.