21

Variétés d'arbres croissants

Bruno Salvy
INRIA, Rocquencourt

[résumé par Michèle Soria]

On appelle arbre croissant un arbre étiqueté dont les étiquettes croissent le long des branches. Ces arbres ont été utilisés comme représentations de permutations, comme structures de données informatiques et comme modèles probabilistes dans diverses applications.
Nous présentons une approche générale permettant le calcul de paramètres de ces arbres. Les fonctions génératrices de ces paramètres sont reliées à une équation différentielle ordinaire simple, qui est non-linéaire et autonome. Les méthodes d'analyse de singularité permettent alors d'analyser asymptotiquement des paramètres comme le degré de la racine, le nombre de feuilles, la longueur de cheminement et les hauteurs des nœuds avec des hypothèses très faibles sur la famille d'arbres étudiée.

1 Introduction

Un *arbre étiqueté* de taille n est un arbre enraciné formé de n nœuds qui sont étiquetés par des entiers distincts de $\{1, \ldots, n\}$. On appelle *arbre croissant* un arbre étiqueté dont les étiquettes croissent le long des branches.
On étudie ici l'asymptotique de caractéristiques de ces arbres, via leur fonctions génératrices. Cette étude est à rapprocher de celle des "familles simples d'arbres" de Meir et Moon. Cependant, à cause de la contrainte de croissance des étiquettes, les équations définissant les fonctions génératrices ne sont plus des équations algébriques simples, mais des équations différentielles algébriques.
Pour un grand nombre de paramètres, les fonctions génératrices sont reliées à une équation différentielle ordinaire simple, qui est non-linéaire et autonome :

$$Y'(z) = \phi(Y(z)) \ .$$

La combinaison de méthodes algébriques et analytiques permet d'analyser dans un cadre uniforme un grand nombre de paramètres comme le degré de la racine, le nombre de feuilles, la longueur de cheminement et les hauteurs des nœuds avec des hypothèses très faibles sur la famille d'arbres étudiée.
Les résultats résumés ici font l'objet d'un article de François Bergeron, Philippe Flajolet et Bruno Salvy [1].

2 Énumération d'arbres croissants

Étant donné un multi-ensemble d'entiers Ω, contenant au moins 0 et un entier ≥ 2, on appelle *variété d'arbres croissants* planaires (ou non planaires) sur Ω l'ensemble de tous les arbres croissants planaires (ou non planaires) dont les nœuds ont un degré extérieur dans Ω.

La *fonction de degré* d'une variété d'arbres croissants planaires sur Ω est $\phi(u) = \sum_{d\in\Omega} u^d$; et pour les arbres non planaires $\phi(u) = \sum_{d\in\Omega} \frac{u^d}{d!}$.
Étant donnée $\mathcal{Y}$ une variété d'arbres croissants, on note $Y(z) = \sum Y_n \frac{z^n}{n!}$ sa série génératrice exponentielle, où Y_n est le nombre d'arbres croissants de taille n.
Former une forêt planaire de k arbres correspond à la série $Y^k(z)$ (et $\frac{1}{k!}Y^k(z)$ pour les forêts non planaires), et rajouter une racine avec étiquette minimale à une forêt de série $W(z)$ correspond à la série $\int_0^z W(z)$. On a donc le théorème suivant.

Théorème 1 *La série génératrice exponentielle $Y(z)$ d'une variété d'arbres définie par la fonction de degré ϕ est définie implicitement par*

$$Y'(z) = \phi(Y(z)),\ Y(0) = 0, \qquad \text{soit} \qquad \int_0^{Y(z)} \frac{du}{\phi(u)} = z.$$

Exemples. Lorsque $\int \frac{du}{\phi(u)}$ et son inverse s'expriment en termes de fonctions spéciales, $Y(z)$ peut avoir une forme explicite. Par exemple les arbres strictement binaires croissants (permutations alternées) ont pour fonction de degré $\phi(u) = 1 + u^2$, d'où $y(z) = \tan(z)$. Les arbres croissants d–aires ont pour fonction de degré $\phi(u) = (1+u)^d$, d'où $y(z) = -1 + [1-(d-1)z]^{-1/(d-1)}$. Les arbres récursifs (arbres planaires croissants sans contrainte d'arité) ont pour fonction de degré $\phi(u) = \exp(u)$ d'où $y(z) = \log\frac{1}{1-z}$. Pour les arbres récursifs non planaires la fonction de degré est $\phi(u) = \frac{1}{1-u}$, d'où $y(z) = 1 - \sqrt{1-2z}$. □

Étude asymptotique. Lorsque $Y(z)$ n'a pas de forme exacte, on a recours à l'analyse asymptotique pour obtenir un équivalent de Y_n quand n tend vers l'infini. La méthode consiste à trouver la (les) singularité(s) de module minimal de $Y(z)$, développer $Y(z)$ au voisinage de cette (ces) singularité(s), et enfin traduire ce développement sur les coefficients.

Le rayon de convergence de $Y(z)$ définie par le théorème 1 est

$$\rho = \int_0^\infty \frac{du}{\phi(u)}\ .$$

Lorsque ϕ est non périodique (i.e. il n'existe pas de série ψ telle que $\phi(u) = \psi(u^p)$, pour $p \geq 2$), ρ est l'unique singularité dominante (i.e. de module minimal) de $Y(z)$. Lorsque ϕ est de période $p \geq 2$, $Y(z) = zY^*(z^p)$, et la série $Y^*(z)$ a une unique singularité dominante en $\rho^{1/p}$.
Lorsque ϕ est un polynôme de degré $d \geq 2$ (la variété $\mathcal{Y}$ est alors dite *polynomiale*), en développant $1/\phi(u)$ au voisinage de l'infini, puis en intégrant, on trouve pour $Y \to \infty$

$$\int_Y^\infty \frac{du}{\phi(u)} \ = \ C.Y^{d-1}\Big(1 + O(\frac{1}{Y})\Big)\ .$$

Or $\int_Y^\infty \frac{du}{\phi(u)} = \rho - z$, et en inversant cette relation, on obtient le développement singulier de $Y(z)$ en fonction de $(\rho - z)^{1/(d-1)}$:

$$Y(z) = \lambda(1 - z/\rho)^{-1/(d-1)} + o((1 - z/\rho)^{-1/(d-1)})\ .$$

L'analyse de singularité permet de traduire l'asymptotique des fonctions sur leurs coefficients :

$$\begin{array}{rclcrcl} Y(z) & = & (1-z/\rho)^\alpha & \Rightarrow & [z^n]Y(z) & = & \frac{1}{\Gamma(-\alpha)}\rho^{-n}n^{-\alpha-1}\Big(1 + O(\frac{1}{n})\Big) \\ Y(z) & = & O((1-z/\rho)^\alpha) & \Rightarrow & [z^n]Y(z) & = & O(\rho^{-n}n^{-\alpha-1}) \end{array}$$

Et l'on a donc finalement

Théorème 2 *Le nombre d'arbres croissants de taille n dans une variété $\mathcal{Y}$ dont la fonction de degré est un polynôme de degré d vaut asymptotiquement*

$$\frac{Y_n}{n!} \sim \lambda \rho^{-n} n^{-(d-2)/(d-1)} \ .$$

Exemple. Les arbres ternaires croissants stricts ont pour fonction de degré $\phi(u) = 1 + u^3$. Il n'existe pas de solution explicite pour $Y(z)$, mais on détermine la singularité $\rho = 2\pi\sqrt{3}/9$, et $\frac{Y_n}{n!} \sim \frac{1}{\sqrt{2\pi}} \rho^{-n-1/2} n^{-1/2}$.

3 Distribution de paramètres

Les méthodes précédentes s'appliquent aussi à l'étude de différents paramètres sur les variétés d'arbres. Un paramètre est *inductif* s'il est défini par une relation

$$s[t] = f_{|t|} + \sum_{\tau \propto t} s[\tau] \ ,$$

pour une suite (f_n), et où la somme est prise sur tous les sous-arbres τ à la racine de l'arbre t. Citons comme exemples de paramètres inductifs la taille ($f_n = 1$), le nombre de feuilles ($f_n = \delta_{n,1}$) et la longueur de cheminement ($f_n = n$).
Étant donnés un paramètre inductif s et une variété $\mathcal{Y}$, la série génératrice exponentielle de s est $S(z) = \sum_{t \in \mathcal{Y}} s[t] \frac{z^{|t|}}{|t|!}$.
Introduisant la relation de définition de s dans la série double $Y_s(z,u) = \sum_{t \in \mathcal{Y}} u^{s[t]} z^{|t|}$, on obtient

$$S(z) = F(z) + \int_0^z S(t) \phi'(Y(t)) dt,$$

où la série $F(z)$ est un produit de Hadamard : $F(z) = \sum f_n Y_n z^n / n!$. La résolution de l'équation différentielle induite donne le résultat suivant.

Théorème 3 *La série génératrice exponentielle d'un paramètre inductif s défini par $s[t] = f_{|t|} + \sum_{\tau \propto t} s[\tau]$, sur une variété $\mathcal{Y}$ est*

$$S(z) = Y'(z) \int_0^z \frac{F'(t)}{Y'(t)} dt, \tag{1}$$

où $F(z)$ est définie à partir de (f_n) et $\mathcal{Y}$ par $F(z) = \sum f_n Y_n \frac{z^n}{n!}$.

Étude asymptotique. L'équation (1) agit comme un "transformateur de singularité". Pour les paramètres inductifs *élémentaires*, i.e. tels que $f_n = C n^\alpha \log^r n$, la singularité de S est aussi en ρ, et l'analyse asymptotique résulte de celle de $Y(z)$ combinée à l'analyse de singularité sur les produits d'Hadamard (voir article [1]).

Théorème 4 *Soit s un paramètre inductif sur la variété polynomiale $\mathcal{Y}$, tel que $f_n = C n^\alpha \log^r n$. La valeur moyenne de s sur les éléments de taille n de $\mathcal{Y}$ vaut asymptotiquement*

$$\overline{S_n} \sim \lambda n \psi(n) \qquad \text{avec} \qquad \psi(n) = \begin{cases} 1 & si\ \alpha < 1, \\ \log^{r+1} n & si\ \alpha = 1, \\ n^{\alpha-1} \log^r n & si\ \alpha > 1. \end{cases}$$

EXEMPLE. La longueur de cheminement est un paramètre inductif, avec $f_n = n$ donc $\alpha = 1$ et $r = 0$; d'où pour toutes les familles d'arbres croissants $\overline{S_n} \sim \lambda n \log n$. En particulier $\lambda = 2$ pour les arbres binaires (Quicksort) et strictement binaires.

La même méthode s'applique à des paramètres qui ne sont pas exactement inductifs élémentaires. Par exemple le nombre moyen de nœuds d'arité i dans un arbre de taille n d'une variété $\mathcal{Y}$ vaut :

$$\overline{S_n^{(i)}} \sim \lambda_i n \quad \text{avec} \quad \lambda_i = \frac{\phi}{\rho} \int_0^\rho \frac{Y^i(t)}{Y'(t)} dt.$$

Et la probabilité que le degré de la racine soit j dans un arbre de taille n d'une variété polynomiale (de degré d) $\mathcal{Y}$ est

$$\pi_{nj} \sim \frac{\alpha_j}{n^{(d-j)/(d-1)}}.$$

De plus une extension de la méthode permet de trouver des distributions limites de paramètres. Par exemple, la profondeur d'un nœud aléatoire : asymptotiquement gaussienne avec moyenne et variance en $\log n$:

$$\mu_n = \frac{d}{d-1} \log n + O(1), \quad \sigma_n^2 = \frac{d}{d-1} \log n + O(1).$$

4 Extensions

Les résultats présentés peuvent être étendus dans diverses directions. Il est possible de considérer des schémas algébriques plus larges, par exemple

$$Y' = H(z, Y(z)) \quad \mapsto \quad S(z) = U(z) \int_0^z \frac{F'(t)}{U(t)} dt.$$

Les variétés non polynomiales peuvent aussi être traitées dans ce cadre. Lorsque la fonction de degré, à coefficients positifs, devient singulière en $\sigma > 0$ (ϕ entière, ou avec des singularités à distance finie), le théorème 2 se généralise : le comportement singulier de $Y(z)$ est obtenu en inversant le développement de $\sum du/\phi(u)$ autour de sa singularité dominante $\rho = \int_0^\infty du/\phi(u)$. Et la forme asymptotique des coefficients de $Y(z)$ se déduit par analyse de singularité. On peut ainsi par exemple traiter les arbres récursifs planaires et non planaires.
L'approche développée sur des expressions intégrales agissant comme des "transformateurs de singularité" est très générale, et permet de traiter de manière unifiée un grand nombre de problèmes statistiques.

Références

[1] F. Bergeron, Ph. Flajolet, and B. Salvy. Varieties of increasing trees. In J.-C. Raoult, editor, *CAAP'92*, volume 581 of *Lecture Notes in Computer Science*, pages 24–48, 1992. Proceedings of the 17th Colloquium on Trees in Algebra and Programming, Rennes, France, February 1992.