

Combinatoire analytique et modèles d'urnes

Stage de Master 2 – MPRI
février – août 2010

Basile Morcrette*
encadré par Philippe Flajolet
Équipe-projet *Algorithms* - INRIA Rocquencourt
78153 Le Chesnay

Le contexte général

Les urnes sont apparues tout d'abord dans les travaux de Bernoulli et Laplace en probabilité, mais il faut attendre Pólya au début du XX^{ème} siècle afin d'obtenir un cadre général. Le modèle considéré est simplement une urne contenant des boules de couleurs différentes, ainsi qu'une règle d'évolution : à chaque instant, on tire au hasard une boule dans l'urne, et suivant sa couleur, on ajoute ou retire une ou plusieurs boules de l'urne, pas nécessairement de la même couleur. Le nombre de boules au départ ainsi que les règles sont fixées (déterministes), et on cherche à connaître la composition de l'urne après un temps très long. Par la diversité des règles, ces modèles décrivent une multitude de phénomènes : évolutions arborescentes, propagation d'épidémie, évolution d'un gaz, campagne électorale, ... Ces problèmes ont été traités par de nombreux probabilistes : Bagchi-Pal, Janson, Smythe, Mahmoud, Chauvin, Pouyanne. Depuis 2005, une nouvelle approche de ces problèmes est possible grâce à la combinatoire analytique. D'ailleurs, l'article de Flajolet–Dumas–Puyhaubert¹ traite des urnes bicolores équilibrées avec au moins un coefficient diagonal négatif (certaines boules peuvent être retirées).

Le problème étudié

On s'intéresse dans ce rapport aux urnes équilibrées, à deux couleurs et plus, et à leur étude par la combinatoire analytique. Pour les urnes à deux couleurs, on s'intéresse à celles ayant tous leurs coefficients positifs (on ne fait qu'ajouter des boules). Ces modèles n'ont pas été étudiés par la combinatoire analytique. Le but est de connaître le comportement limite de ces urnes, i.e. la composition asymptotique de l'urne. L'avantage du traitement par la combinatoire analytique est l'obtention de résultats précis sur les distributions limites, notamment des vitesses de convergence ainsi que des aspects locaux. L'intérêt pour ces modèles vient aussi du lien entre les urnes et certaines structures informatiques (arbres, graphes, réseaux).

*Basile.Morcrette@inria.fr

1. Philippe Flajolet, Philippe Dumas, et Vincent Puyhaubert. Some exactly solvable models of urn process theory. *Discrete Mathematics & Theoretical Computer Science*, pages 59–118, 2006

La contribution proposée

Partant de problèmes informatiques liés à des structures combinatoires de grande taille, on se propose de les traduire dans le formalisme des urnes, puis de développer des méthodes provenant de la combinatoire analytique², afin de déduire des propriétés fines sur la distribution limite de ces objets.

On étudie par exemple certaines propriétés d'attachement sur les arbres binaires selon une croissance uniforme aux feuilles (correspondant à des formules booléennes) : cela nous conduit à l'étude d'urnes de grandes dimensions (trois et plus). On obtient une loi de Poisson pour le comportement asymptotique.

On découvre une certaine classe d'urnes, dites de croissance préférentielle, dont les fonctions génératrices de dénombrement sont algébriques. On obtient leur comportement moyen.

On étudie enfin la modélisation par les urnes d'une certaine famille de k -arbres, structure informatique décrivant l'évolution de réseaux avec attachement préférentiel (réseaux sociaux). On se propose d'obtenir des résultats sur les limites avec vitesse de convergence. L'étude faite ici utilise les méthodes de combinatoire analytique selon ce schéma : fonctions génératrices et mise en équation ; exploitation des équations et analyse de singularités ; étude asymptotique des coefficients ; analyse fine pour les résultats en probabilité.

Les arguments en faveur de sa validité

Les exemples traités proviennent de questions liées aux structures informatiques telles que les arbres binaires étiquetés par des formules logiques et les k -arbres ordonnés croissants. Les résultats obtenus sont très encourageants : en allant au bout de la méthode analytique, on obtient une convergence en loi avec vitesse de convergence, des informations précises sur la queue de la distribution (grandes déviations), ainsi que des résultats de lois limites locales. Ces derniers résultats, inaccessibles par les méthodes probabilistes, incitent à poursuivre la classification de ces modèles par la combinatoire analytique.

Le bilan et les perspectives

On obtient des premiers résultats concernant des urnes de grande dimension (trois et plus), ce qui incite à généraliser dans ce sens.

Pour ce qui concerne les urnes bicolores positives, l'étude de différentes sous-classes confirme l'intérêt de la démarche générique par la combinatoire analytique : fonction génératrice et mise en équation, exploitation des équations et analyse de singularités, étude asymptotique des coefficients, analyse fine pour les résultats en probabilité. Cette étude soulève aussi quelques difficultés, notamment les problèmes de changement de régime dans l'asymptotique des fonctions complexes multivariées, liés aux singularités multiples.

Enfin, la classification des urnes bicolores additives, l'étude en dimension supérieure, mais aussi l'étude d'urnes non équilibrées sont des problèmes largement ouverts.

2. Le livre de référence dans le domaine est celui de Philippe Flajolet et Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009

COMBINATOIRE ANALYTIQUE ET MODÈLES D'URNES

BASILE MORCRETTE
PROJET ALGORITHMS, INRIA ROCQUENCOURT

Introduction

Quel rapport y a-t-il entre une campagne électorale désastreuse, la longueur d'une branche d'arbre, un gaz passant d'un compartiment à un autre et la propagation d'une épidémie ? *A priori*, peu de chose, si ce n'est qu'ils peuvent tous être décrits par un même formalisme : les *modèles d'urnes*. En effet, prenons une urne comportant un nombre fini de boules de deux couleurs différentes : noires et blanches. Ces boules peuvent très bien représenter les partisans de deux adversaires politiques (M. Black & M. White) ; on peut voir aussi les boules noires comme les feuilles d'un arbre et les boules blanches comme les noeuds internes ; ou encore les blanches comme les particules de gaz du compartiment gauche, les noires comme celles du compartiment droit ; ou encore les noires comme les individus sains, et les blanches comme les individus contaminés. Une fois ce formalisme simple adopté, on a unifié le système. Il peut maintenant évoluer : les deux hommes politiques sont tellement peu convaincants que lorsque l'un fait un discours, cela donne un nouveau partisan (précédemment indécis) pour son adversaire ; un arbre grandit (de façon binaire) en remplaçant une feuille en un noeud interne et deux feuilles (i.e. +1 blanche, +1 noire) ; une particule change de compartiment en changeant sa couleur ; un individu contaminé va en contaminer trois autres, tandis qu'un individu sain va guérir un contaminé. Ces évolutions (modification du nombre de boules selon la couleur tirée) peuvent donc être regroupées dans un formalisme commun, formalisme qui est celui des *urnes de Pólya équilibrées à deux couleurs*. Cependant, comme nos exemples le suggèrent, la nature des phénomènes quantitatifs en jeu diffère très largement d'une règle à l'autre. Avec la définition formelle des modèles d'urnes, donnée dans la section suivante, les modèles se résument aux matrices à coefficients entiers $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$. Notre objectif est précisément de contribuer à l'analyse quantitative et à la classification de ces modèles.

Ces urnes ont déjà eu un certain succès auprès des probabilistes. De fait, ces modèles ont été introduits depuis plus de deux siècles par Bernoulli et Laplace. Leur comportement est extrêmement bien connu pour certaines catégories de modèles, mais à cause de la richesse et de la diversité des règles, de nombreuses zones d'ombres subsistent. Depuis 2005 (cf. [FDP06], [FGP05]), la combinatoire analytique permet elle aussi d'aborder ces problèmes, et ceci avec une démarche assez générique dans le domaine (cf. [FS09]), fondée sur une description combinatoire, des équations sur les fonctions génératrices, une recherche d'asymptotique par analyse de singularités. Elle permet alors d'obtenir des résultats probabilistes forts, notamment sur les grandes déviations, et sur les lois locales limites.

Ce rapport, au travers d'exemples concrets, développe ces techniques analytiques, afin de comprendre comment peut se faire la généralisation à toutes les urnes de Pólya équilibrées à coefficients positifs. Le plan de l'étude est le suivant :

- La section 1 (p.3) introduit le formalisme adopté ainsi que les résultats principaux sur les urnes analytiques [FDP06], en particulier le théorème liant les urnes équilibrées à un système différentiel.
- La section 2 (p.4) répond à un problème posé par Cécile Mailler et Brigitte Chauvin¹ à propos de branches de longueur 1 dans les arbres. Cela fait intervenir une urne 3×3 et des techniques élémentaires de résolution. Une loi de Poisson est obtenue à la limite. Le résultat est généralisé aux sous-arbres à k feuilles, et l'on fait le lien avec des résultats connus sur les permutations.
- La section 3 (p.12) traite d'une classe d'urnes 2×2 dont les fonctions génératrices de dénombrement sont algébriques. Un premier défrichage est fait dans le but d'obtenir une convergence vers une loi normale, mais les problèmes de changement de régime asymptotique sont encore mal compris.
- La section 4 (p.17) fait un lien entre les urnes et les k -arbres, lesquels constituent une structure utile pour la compréhension de la croissance des réseaux de grande taille. Ce travail a été motivé par l'article [PS10] de Panholzer–Seitz² : on introduit une urne 2×2 afin de compter les lieux possibles de croissance dans ces arbres. Le modèle d'urne permet d'obtenir, grâce à une analyse fine, des résultats de convergence vers une loi normale, avec des précisions sur la vitesse de convergence, sur les grandes déviations et sur la loi locale limite.

Voici, à mon sens, les principales contributions de ce mémoire :

- Dans la section 2, je résous un problème lié au comptage de sous-arbres de taille fixée dans un arbre, en utilisant le formalisme des urnes. Cela permet d'illustrer facilement le théorème de la section 1, ainsi que les outils de base utilisés en combinatoire analytique : séries génératrices, extraction de coefficients et probabilités. Je propose également un lien bijectif entre l'urne comptant les sous-arbres et les cycles dans les permutations, puis je rappelle le lien entre permutations et records ; ce qui permet de rattacher les modèles d'urnes à des problèmes combinatoires bien connus.
- La section 3 est un travail en cours. J'ai découvert une classe d'urnes équilibrées bicolores additives qui se trouve être *algébriques*, dans le sens où les fonctions génératrices qui comptent les histoires sont algébriques. Cette classe d'urnes correspond à l'évolution de deux populations de façon croissante et préférentielle. L'étude de cette classe d'urne, qui est amorcée ici, est un premier pas pour la généralisation à l'ensemble des urnes additives. En effet, on met en lumière que des problèmes de discontinuité dans le comportement asymptotique apparaissent, et que ceci est lié à une confluence de singularités. Ce phénomène peu connu jusqu'alors (on trouve quelques éléments dans [FS09]³), mérite un travail théorique approfondi.
- Enfin, la section 4 est la partie la plus technique et la plus aboutie du rapport. Partant d'une modélisation d'une structure informatique, les *k*-arbres, j'obtiens une classe infinie d'urnes additives. Par l'utilisation de plusieurs techniques de combinatoire analytique, j'obtiens pour cette classe d'urne des résultats forts sur la distribution limite, en ce qui concerne les vitesses de convergence et les lois locales limites, jusqu'alors inaccessibles avec les traitements probabilistes.

1. Université de Versailles – UVSQ

2. Université de Vienne, Autriche

3. Il s'en présente certains, mais de nature différente, dans l'étude des cartes planaires.

Cela amorce un travail qui sera poursuivi en thèse à l'INRIA Rocquencourt, sous la direction de Philippe Flajolet.

1. Modèles d'urnes

1.1. **Modèle.** On considère le modèle d'urne de Pólya : on dispose d'une urne contenant des boules de différentes couleurs, ainsi que des règles d'évolution. Considérons le cas à deux couleurs : les boules sont de couleur x ou y . Les règles sont représentées par une matrice 2×2 :

$$(1) \quad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad \alpha, \delta \in \mathbb{Z}, \quad \beta, \gamma \in \mathbb{Z}_{\geq 0}.$$

À chaque instant, on tire une boule de l'urne, de façon uniforme. Si la couleur est x , on replace cette boule dans l'urne en ajoutant α boules x et β boules y . Si la couleur est y , on replace cette boule tout en ajoutant γ boules x et δ boules y . Notons que l'on peut avoir des coefficients négatifs sur la diagonale : cela signifie que l'on retire des boules de l'urne. On dit dans ce cas que l'urne est *soustractive*. (-1 signifie que l'on enlève la boule tirée). Si tous les coefficients sont positifs, on parle d'urne *additive*. On considère dans ce mémoire les urnes *équilibrées* définies par : $\alpha + \beta = \gamma + \delta$. Cela correspond à ajouter un nombre de boules constant à chaque étape. On note $\sigma := \alpha + \beta$ la "balance" de l'urne.

Partant d'une configuration initiale $\binom{a_0}{b_0}$, on fait évoluer le système par la matrice de règles (1). On cherche à connaître la distribution (exacte ou asymptotique) du nombre de boules de chaque couleur, après n étapes. Pour les urnes équilibrées, le nombre de boules à chaque instant est déterministe (et vaut $a_0 + b_0 + \sigma n$).

1.2. **L'apport probabiliste.** Ces modèles ont été largement étudiés avec des outils probabilistes depuis Pólya. En commençant par des exemples concrets venant notamment de problèmes de physique statistique (Ehrenfest, Friedman), la connaissance sur ces modèles s'est étoffée, notamment grâce à Bagchi et Pal [BP85], Smythe [Smy96], Janson [Jan06] [Jan08], Chauvin, Pouyanne, Sahnoun [CPS]... Des résultats sur les moments, et sur la convergence en loi ont été obtenus ; ainsi qu'un changement de régime suivant le rapport des valeurs propres de la matrice de l'urne. Il est ainsi connu que l'on a une convergence en loi vers une loi gaussienne pour une certaine classe d'urnes ("petites urnes"). Une introduction aux urnes par les probabilités se trouve dans [JK77], ou dans le plus récent [Mah08].

1.3. **L'apport analytique.** Depuis 2004 et les articles [FGP05] et [FDP06], une approche analytique et combinatoire permet d'étudier ces modèles sous un nouvel angle. L'élément crucial de cette approche analytique est un théorème de correspondance entre l'urne (1) et un système différentiel.

Définition 1. Une histoire de longueur n est une suite de n évolutions, obtenues par n tirages successifs dans l'urne. La fonction génératrice (exponentielle) des histoires de l'urne (1) est

$$(2) \quad H(x, y, z; a_0, b_0) = \sum_{n, a, b} H_n(a_0, b_0; a, b) x^a y^b \frac{z^n}{n!},$$

où $H_n(a_0, b_0; a, b)$ est le nombre d'histoires de longueur n commençant par la configuration (a_0, b_0) , et finissant dans une configuration (a, b) . On écrira $H(x, y, z)$ lorsqu'il n'y a pas de confusion possible sur la configuration initiale (a_0, b_0) .

Théorème 1. À l'urne (1), sous condition d'équilibre $\alpha + \beta = \gamma + \delta$, on associe le système différentiel suivant (le point correspond à la dérivation : $\dot{X}(t) = \frac{d}{dt}X(t)$) :

$$\begin{cases} \dot{X} &= X^{\alpha+1}Y^{\beta} \\ \dot{Y} &= X^{\gamma}Y^{\delta+1} \end{cases} .$$

Soient x_0 et y_0 deux complexes tels que $x_0y_0 \neq 0$. Soient $X(t, x_0, y_0)$ et $Y(t, x_0, y_0)$ les solutions du système différentiel avec les conditions initiales $X(t=0) = x_0$, $Y(t=0) = y_0$. Alors, la fonction génératrice des histoires s'écrit

$$(3) \quad H(x_0, y_0, z) = X(z, x_0, y_0)^{a_0} Y(z, x_0, y_0)^{b_0} .$$

On obtient ainsi une fonction génératrice trivariée contenant toute l'information de l'urne et de son évolution : $[x^a y^b z^n]H(x, y, z)$ est le nombre d'histoires de longueur n , se terminant avec a boules de couleur x , et b boules de couleur y . Comme les urnes considérées sont équilibrées, on sait que le nombre total de boules à l'instant n est $a_0 + b_0 + \sigma n$. Ainsi, si l'on connaît le nombre a de boules x , une simple soustraction nous donne le nombre b de boules y . Il est alors superflu de garder une dépendance en y dans la série génératrice H . On spécifie alors $y = 1$. Et il ne nous reste qu'une fonction bivariée $H(x, z)$.

Il importe de noter que c'est la condition d'équilibrage ($\alpha + \beta = \gamma + \delta$) qui fait que les modèles énumératifs et probabilistes sont équivalents : toutes les histoires de longueur n sont en effet alors équiprobables.

Remarque. Tout ceci reste valable pour des urnes multicolores à k couleurs. On utilise une fonction H à $k+1$ variables, et un système différentiel à k équations et k inconnues, avec k conditions initiales. Il y a alors également une variable superflue dans H , on peut se ramener à k variables.

Comme nous allons le voir dans les divers exemples qui vont suivre, l'accès à la fonction génératrice des histoires va nous permettre d'obtenir une asymptotique sur les coefficients. En normalisant, on obtiendra des informations précises, avec plusieurs termes d'approximation, sur la moyenne, la variance, et autres moments. Enfin, une analyse précise de cette fonction donnera des résultats sur la convergence en loi, les grandes déviations, et la loi locale limite.

2. Loi de Poisson dans les branches d'arbres

Le problème qui suit a été posé par Brigitte Chauvin et Cécile Mailler (UVSQ) à propos de formules booléennes représentées sous forme d'arbre binaire. On se place dans un modèle où seule l'implication et les littéraux sont autorisés. Soit un arbre binaire représentant une telle formule. Sur sa branche droite (la suite des fils droits), il n'y a que des implications, sauf la feuille terminale qui est un littéral. Sur cette branche, ce qui est à déterminer est le nombre de fils gauches qui sont directement des feuilles, i.e., des littéraux (cf fig.1, à gauche). Le modèle probabiliste choisi est celui d'arbres binaires non étiquetés à croissance uniforme aux feuilles ; i.e., pour passer d'un arbre de taille n à un arbre de taille $n+1$, on choisit au hasard uniformément une feuille, que l'on remplace par un noeud interne et 2 feuilles attachées à celui-ci. Ce modèle est équivalent à celui des arbres binaires de recherche. Un tel arbre sera appelé dans la suite *arbre aléatoire*.

On va résoudre un problème plus général, dont le résultat est le théorème 2 qui suit. La démonstration est donnée dans la section 2.2.

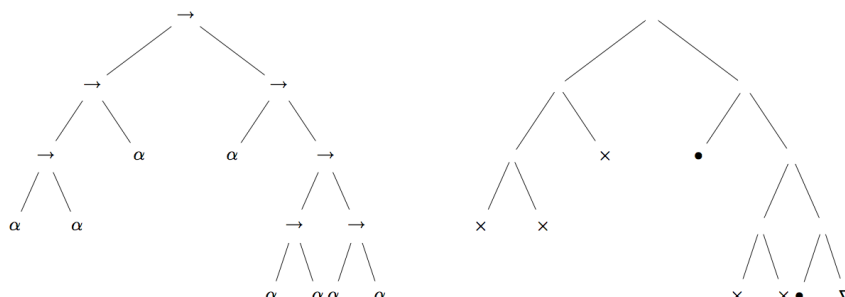


FIGURE 1. À gauche, l'arbre d'une formule : α représente un littéral quelconque. À droite, la traduction avec l'urne à trois couleurs. On veut compter les boules \bullet .

Définition 2. Soit X une variable aléatoire à valeur dans \mathbb{N} . On dit que X suit une loi de Poisson de paramètre λ si pour tout l dans \mathbb{N} ,

$$(4) \quad \mathbb{P}(X = l) = e^{-\lambda} \frac{\lambda^l}{l!}.$$

On note alors $X \sim \mathcal{P}(\lambda)$.

Théorème 2. Soit $U_{k,n}$ le nombre de sous-arbres gauches de taille k directement accrochés à la branche droite d'un arbre aléatoire de taille n . Alors, la variable $U_{k,n}$ converge en loi, $U_{k,n} \xrightarrow[n \rightarrow \infty]{} U_k$, où la limite est une variable de Poisson de paramètre $\frac{1}{k}$ (soit $U_k \sim \mathcal{P}(\frac{1}{k})$), avec une vitesse de convergence en $O\left(\frac{(2k)^n}{n!}\right)$.

Ce théorème est fortement relié au dénombrement des permutations selon les longueurs des k -cycles, et également aux espacements entre records en théorie statistique classique. Ces liens seront détaillés dans la section 2.3 : on explicitera la bijection entre les histoires d'urnes et les permutations, puis on rappellera le lien entre les permutations et les records. On peut alors retourner aux urnes en faisant le lien entre compter les cycles d'une permutation et compter les sous-arbres, sans se préoccuper de leur taille.

Dans la section 2.1, on montre le résultat pour le cas de base $k = 1$, i.e.

$$(5) \quad U_{1,n} \xrightarrow[n \rightarrow \infty]{d} U_1 \text{ où } U_1 \sim \mathcal{P}(1).$$

2.1. Le cas de l'urne 3×3 . Ici, l'objectif est de compter le nombre de sous-arbres gauches de taille 1 directement accrochés à la branche droite. Nous allons tout d'abord détailler la modélisation par urne. Pour cela, on utilise des boules de trois couleurs. Une boule "Déesse Mère" ∇ (qui correspond à la feuille la plus à droite : elle est toujours là, et il n'y en a qu'une). Des boules \bullet qui correspondent aux feuilles "directes" — les descendants directs de la branche droite. Et des boules \times qui représentent toutes les autres feuilles de l'arbre (cf fig.1). On débute l'évolution avec la "Déesse Mère" ∇ uniquement.

- Si on choisit cette boule ∇ , elle est conservée et l'on crée une feuille "directe" \bullet (1^{re} ligne de la matrice (6)).

- Si on choisit une boule \bullet , alors on transforme la feuille en arbre binaire à 2 feuilles, donc on crée $\times \times$, et on supprime \bullet (2^e ligne).
- Enfin, si on choisit \times , on crée $\times \times$, tout en supprimant la boule tirée (3^e ligne).

On obtient l'urne 3×3 suivante :

$$(6) \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix} .$$

On a une urne équilibrée de balance 1. Au départ, l'urne contient 1 boule de type x (∇), 0 boule de type y (\bullet), et 0 boule de type u (\times). Notons que cette urne est non classique à cause de son entrée diagonale négative.

Voici l'énoncé du théorème 2, spécialisé à $k = 1$:

Lemme 1. *Si $n \rightarrow \infty$, $\mathbb{P}(U_{1,n} = l) \rightarrow \frac{e^{-1}}{l!} \cdot U_{1,n}$ converge en loi vers $\mathcal{P}(1)$, avec une vitesse de convergence en $O\left(\frac{2^n}{n!}\right)$.*

PREUVE. On utilise la fonction génératrice exponentielle des histoires $H(x, y, u, z)$, comme en (2). Le système différentiel associé est

$$\begin{cases} \dot{X} &= XY \\ \dot{Y} &= U^2 \\ \dot{U} &= U^2 \end{cases} .$$

Ce système est triangulaire, ce qui permet d'obtenir successivement U , Y et X par intégration. On obtient :

$$U = \frac{u_0}{1 - u_0 z} \quad Y = \frac{u_0}{1 - u_0 z} + (y_0 - u_0) \quad X = \frac{x_0}{1 - u_0 z} \exp((y_0 - u_0)z) .$$

Par le théorème 1 de correspondance, la fonction génératrice est donc :

$$(7) \quad H(x, y, u, z) = X(x, y, u, z) = \frac{x}{1 - uz} \exp((y - u)z) .$$

z^n : n compte la longueur de l'histoire ; ici, le nombre de boules dans l'urne : $n + 1$ (c'est le nombre de feuilles dans l'arbre).

x^a : a compte le nombre de noeuds tout au bout de la branche : c'est toujours 1 . Cette information est donc inutile ; on peut donc prendre $x = 1$.

y^b : b compte le nombre de noeuds "directs" (la première génération de descendants de la branche directement feuilles) : c'est ce que l'on veut compter.

u^c : c compte le nombre de feuilles restantes : or $n = a + b + c$, donc $c = n - 1 - b$: cette information est également redondante ; donc on peut faire $u = 1$.

Finalement, il reste à étudier

$$(8) \quad H(1, y, 1, z) = \frac{1}{1 - z} \exp((y - 1)z) = \frac{e^{-z}}{1 - z} e^{yz} .$$

La probabilité recherchée est le nombre d'histoires de longueur n ayant l boules "directes" (de type y), divisé par le nombre d'histoires de longueur n , c'est-à-dire,

$$(9) \quad \mathbb{P}(U_{1,n} = l) = \frac{n! [z^n y^l] H(1, y, 1, z)}{n! [z^n] H(1, 1, 1, z)} ,$$

où $[z^n y^l] H(1, y, 1, z)$ désigne le coefficient de $z^n y^l$ dans la série génératrice. On a, comme il est attendu, $[z^n] H(1, 1, 1, z) = [z^n] (1 - z)^{-1} = 1$, ainsi que

$$H(1, y, 1, z) = \left(\sum_{n \geq 0} z^n \right) \left(\sum_{m \geq 0} (y-1)^m \frac{z^m}{m!} \right) = \sum_{n \geq 0} \left(\sum_{j=0}^n \frac{(y-1)^j}{j!} \right) z^n,$$

avec un peu de calcul,

$$[z^n y^l] H(1, y, 1, z) = \frac{1}{l!} \sum_{j=0}^{n-l} \frac{(-1)^j}{j!} = \frac{e^{-1}}{l!} - \frac{1}{l!} \sum_{j \geq n-l+1} \frac{(-1)^j}{j!}.$$

Finalement,

$$(10) \quad \mathbb{P}(U_{1,n} = l) = \frac{e^{-1}}{l!} - \frac{1}{l!} \sum_{j \geq n-l+1} \frac{(-1)^j}{j!}.$$

On en déduit le résultat annoncé. Le calcul de l'erreur se fait par une majoration de $\mathbb{P}(U_{1,n} \leq l) - \mathbb{P}(U_1 \leq l)$. \square

Remarque. La fonction génératrice est la même que celle des permutations où l'on compte le nombre de points fixes :

$$(11) \quad H(y, z) = \frac{1}{1-z} \exp((y-1)z) = \exp\left(\ln\left(\frac{1}{1-z}\right) - z + yz\right).$$

On identifie cette formule à la description combinatoire symbolique suivante, avec les notations de [FS09] :

$$(12) \quad \mathcal{H} = \text{SET}(\text{CYC}_{>1}(\mathcal{Z}) \cup \mathcal{Y}\mathcal{Z}).$$

On retrouve ainsi la probabilité d'avoir un dérangement (i.e. une permutation de \mathcal{S}_n sans point fixe). Cela correspond au cas $l = 0$. Dans notre cas, cela revient à regarder la probabilité d'avoir une histoire qui se termine dans une configuration sans boule \bullet , ou encore la probabilité qu'un arbre aléatoire ne possède pas de feuille "directe" sur sa branche droite.

$$\mathbb{P}(U_{1,n} = 0) = \sum_{j=0}^n \frac{(-1)^j}{j!}.$$

Ainsi, la probabilité d'avoir un dérangement, si $n \rightarrow \infty$, tend vers e^{-1} (cf [FS09] p.622). On reviendra sur ces aspects bijectifs dans la sous-section 2.3.

2.2. Généralisation par une urne $(k+2) \times (k+2)$. Voici la preuve du théorème général 2.

PREUVE. On a désormais des boules de $k+2$ couleurs différentes; une boule spéciale ∇ (toujours notre "Déesse Mère"); k types de boule $\bullet_1, \dots, \bullet_k$ comptant les sous-arbres à 1, 2, \dots , k feuilles; enfin, un type \times pour les sous-arbres ayant plus de k feuilles. Les règles de l'urne sont les suivantes :

$$\begin{array}{lcl} \nabla & \rightarrow & \bullet_1 \nabla \\ \bullet_1 & \rightarrow & \bullet_2 \bullet_2 \\ \bullet_2 \bullet_2 & \rightarrow & \bullet_3 \bullet_3 \bullet_3 \\ & & \vdots \\ \underbrace{\bullet_k \dots \bullet_k}_{k \text{ fois}} & \rightarrow & \underbrace{\times \times \dots \times}_{k+1 \text{ fois}} \\ \times & \rightarrow & \times \times . \end{array}$$

La matrice de l'urne s'écrit :

$$(13) \quad \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 2 & \ddots & & \\ \vdots & \ddots & -2 & 3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & -k & k+1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}.$$

2.2.1. *Fonction génératrice.* On associe les variables X, U_1, \dots, U_k, V aux couleurs $\nabla, \bullet_1, \dots, \bullet_k, \times$. Le système différentiel associé est

$$\begin{cases} \dot{X} &= XU_1 \\ \dot{U}_1 &= U_2^2 \\ \dot{U}_2 &= U_2^{-1}U_3^3 \\ &\vdots \\ \dot{U}_j &= U_j^{-j+1}U_{j+1}^{j+1} \\ &\vdots \\ \dot{U}_k &= U_k^{-k+1}V^{k+1} \\ \dot{V} &= V^2. \end{cases}$$

Ce système est triangulaire. Pour le résoudre, on part de la dernière ligne : V s'intègre directement en $\frac{v_0}{1-v_0z}$, et on prend $v_0 = 1$ (cette information est superflue, car l'urne est équilibrée).

$$V(z) = \frac{1}{1-z}.$$

On peut ainsi en déduire U_k par intégration de $\dot{U}_k U_k^{k-1} = V^{k+1}$:

$$U_k^k = k \int \frac{dz}{(1-z)^{k+1}} = \frac{1}{(1-z)^k} + (U_k^k(0) - 1).$$

Puis, on remarque la propriété suivante : pour $j = 1 \dots k-1$, $\dot{U}_j U_j^{j-1} = U_{j+1}^{j+1}$. Cela s'intègre en $U_j^j = j \int U_{j+1}^{j+1}$. En itérant cette dernière égalité, on obtient

$$U_j^j = j(j+1) \dots (k-1) \underbrace{\int \int \dots \int}_{k-j \text{ fois}} U_k^k.$$

En intégrant $k-1$ fois U_k , on trouve

$$U_1 = \frac{1}{1-z} + \sum_{j=1}^k (U_j^j(0) - 1) z^{j-1}.$$

Enfin $\dot{X} X^{-1} = U_1$, d'où

$$X(z) = x_0 \cdot \exp \left(\ln \left(\frac{1}{1-z} \right) + \sum_{j=1}^k (U_j^j(0) - 1) \frac{z^j}{j} \right).$$

En posant $u_j = U_j^j(0)$, et en prenant $x_0 = 1$ (la "Déesse Mère" ne disparaît jamais!), on a

$$X(u_1, \dots, u_k, z) = \exp \left(\ln \left(\frac{1}{1-z} \right) + (u_1 - 1)z + (u_2 - 1)\frac{z^2}{2} + \dots + (u_k - 1)\frac{z^k}{k} \right)$$

$$(14) \quad = \exp \left(u_1 z + u_2 \frac{z^2}{2} + \dots + u_k \frac{z^k}{k} + \sum_{j>k} \frac{z^j}{j} \right).$$

En posant $u_j = U_j^j(0)$, on compte ainsi directement les arbres à j feuilles. On retrouve au passage la série génératrice des permutations où u_j compte le nombre de j -cycles (cycles de longueur j), et si on extrait les coefficients, on trouve une formule originellement due à Cauchy (cf. [FS09]).

2.2.2. *Probabilité limite.* Soit $X_k(u_k, z) := X(1, \dots, 1, u_k, z) = \frac{1}{1-z} \exp \left((u_k - 1) \frac{z^k}{k} \right)$.

On s'intéresse à la loi limite du nombre de sous-arbres à k feuilles, donc au coefficient en $u_k^l z^n$ dans $X_k(u_k, z)$:

$$\begin{aligned} X_k(u_k, z) &= \left(\sum_{n \geq 0} z^n \right) \left(\sum_{m \geq 0} \left(\frac{u_k - 1}{k} \right)^m \frac{z^{km}}{m!} \right). \\ [z^n] X_k(u_k, z) &= \sum_{j=0}^{\lfloor n/k \rfloor} \frac{\left(\frac{u_k - 1}{k} \right)^j}{j!}. \end{aligned}$$

Avec les mêmes calculs que dans la section précédente, on a

$$[u_k^l z^n] X_k(u_k, z) = \frac{e^{-1/k}}{k^l l!} - \frac{1}{k^l l!} \sum_{j \geq \lfloor n/k \rfloor - l + 1} \frac{(-1/k)^j}{j!}.$$

Comme précédemment, $[z^n] X_k(1, z) = [z^n] (1-z)^{-1} = 1$, donc la variable aléatoire $U_{k,n}$ vérifie

$$\mathbb{P}(U_{k,n} = l) = \frac{[u_k^l z^n] X_k(u_k, z)}{[z^n] X_k(1, z)} = \frac{e^{-1/k}}{k^l l!} - \frac{1}{k^l l!} \sum_{j \geq \lfloor n/k \rfloor - l + 1} \frac{(-1/k)^j}{j!}.$$

Donc si n tend vers l'infini, $\mathbb{P}(U_{k,n} = l)$ tend vers $\frac{e^{-1/k}}{k^l l!}$. Ce qui conclut la preuve : quand $n \rightarrow \infty$, alors $U_{n,k} \xrightarrow[n \rightarrow \infty]{} U_k$, où $U_k \sim \mathcal{P}(k^{-1})$. La vitesse de convergence s'obtient par une majoration de $\mathbb{P}(U_{k,n} \leq l) - \mathbb{P}(U_k \leq l)$. On trouve $O\left(\frac{(2k)^n}{n!}\right)$. \square

Remarque. Par analyse du pôle en $z = 1$, $[z^n] \frac{1}{1-z} e^{(u_k-1)\frac{z^k}{k}} \sim e^{\frac{u_k-1}{k}}$, or $e^{(u-1)\lambda}$ est la fonction génératrice de probabilité d'une variable aléatoire de Poisson de paramètre λ . Le résultat s'obtient alors alternativement, par le théorème de continuité des fonctions génératrices de probabilité. ([FS09] p.625).

2.3. Un peu de combinatoire bijective. On a pu remarquer que les séries génératrices obtenues sont les mêmes que celles qui comptent les permutations selon leurs k -cycles (formules (11), (12), et (14)). Nous allons tout d'abord donner une bijection entre les histoires d'urnes comptant les arbres aléatoires et les permutations. Puis on rappellera le lien entre records et permutations, ce qui permet de retrouver des résultats bien connus en combinatoire et en théorie statistique. On terminera la section en donnant la construction d'une urne faisant le lien entre les histoires et le nombre de cycles dans une permutation.

Proposition 1. *Il existe une bijection constructive entre les histoires de longueur n de l'urne (13) comptant les sous-arbres de taille $1, 2, \dots, k$ et les permutations de taille n où l'on compte le nombre de cycles de longueur $1, 2, \dots, k$.*

Remarque. Pour $k = 1$, on retrouve la bijection (cf. formule 12) entre les permutations de taille n ayant p point fixes et les histoires de l'urne (13) de longueur n terminant dans une configuration avec p boules \bullet_1 .

Pour la preuve de la proposition, on va utiliser une urne plus simple, de taille 2. Elle est donnée par les règles suivantes :

$$(15) \quad \begin{array}{l} \nabla \rightarrow \bullet \nabla \\ \bullet \rightarrow \bullet \bullet \end{array} .$$

Le point de départ est ∇ . Le système différentiel associé est

$$\begin{cases} \dot{X} = XY \\ \dot{Y} = Y^2 \end{cases} .$$

Ainsi, la série génératrice de cette urne est

$$X(z) = \frac{1}{1-z} = \exp \left(\ln \left(\frac{1}{1-z} \right) \right) .$$

Il s'agit en fait d'une grammaire qui génère les arbres binaires enracinés, et où l'on marque le noeud terminal de la branche la plus à droite.

Lemme 2. *Il existe une construction bijective entre les histoires de l'urne (15) de longueur n et les permutations de taille n .*

PREUVE. Une histoire de longueur n (suite des évolutions) permet de construire une permutation, et réciproquement. Nous allons détailler la construction bijective. Notons tout d'abord que c'est toute l'histoire de l'urne qui permet de donner une unique permutation, et non pas la configuration finale (évidemment), ni la configuration d'arbre finale. La permutation se construit pas à pas : lorsque l'on regarde une histoire de longueur i , cela permet d'ajouter i dans la permutation : on construit une permutation de \mathcal{S}_i à partir d'une permutation de \mathcal{S}_{i-1} .

Partant de ∇ , la seule action possible donne $\bullet \nabla$: cette histoire correspond à la permutation (1) de \mathcal{S}_1 . Les \bullet correspondent aux éléments de la permutation. On note \bullet^i lorsque la boule \bullet correspond à un élément i d'un permutation. Ici, $\bullet^1 \nabla$. Le principe pour la suite est simple : si on choisit ∇ à l'étape i , cela va créer un nouveau cycle dans la permutation $:(i)$, i.e. dans la permutation σ à l'étape $i-1$, on passe à σ' de l'étape i en posant

$$\sigma' = (i)\sigma .$$

Sinon, la boule frappée \bullet correspond à un élément j , et alors, on remplace ce j par ji , i.e. dans la permutation σ à l'étape $i-1$, on passe à σ' de l'étape i en posant

$$\sigma'(j) = i, \sigma'(i) = \sigma(j), \text{ et } \forall l \neq i, j \sigma'(l) = \sigma(l) .$$

Par exemple, l'histoire suivante :

$$\begin{array}{c} \nabla \xrightarrow{\nabla} \bullet^1 \nabla \xrightarrow{\nabla} \bullet^1 \bullet^2 \nabla \xrightarrow{\bullet^1} \bullet^1 \bullet^3 \bullet^2 \nabla \xrightarrow{\bullet^2} \bullet^1 \bullet^3 \bullet^2 \bullet^4 \nabla \\ \xrightarrow{\bullet^2} \bullet^1 \bullet^3 \bullet^2 \bullet^5 \bullet^4 \nabla \xrightarrow{\nabla} \bullet^1 \bullet^3 \bullet^2 \bullet^5 \bullet^4 \bullet^6 \nabla \end{array}$$

correspond de façon unique à la construction de la permutation (1 3)(2 5 4)(6). En effet, la construction donne

$$\emptyset \xrightarrow{\nabla}_1 (1) \xrightarrow{\nabla}_2 (1)(2) \xrightarrow{\bullet^1}_3 (1\ 3)(2) \xrightarrow{\bullet^2}_4 (1\ 3)(2\ 4) \xrightarrow{\bullet^2}_5 (1\ 3)(2\ 5\ 4) \xrightarrow{\nabla}_6 (1\ 3)(2\ 5\ 4)(6) .$$

Pour la réciproque, il suffit de se donner une permutation σ de taille n . On la regarde suivant sa décomposition en cycles à supports disjoints. On organise chaque cycle en plaçant le plus petit élément

en tête, puis on classe les différents cycles selon leur élément de tête dans l'ordre croissant. Cela donne une écriture canonique pour chaque permutation. De cette écriture, on en déduit une histoire d'urne qui démarre par (1), i.e. $\nabla \xrightarrow{\nabla} \bullet \nabla$. À l'étape $i - 1$, on a construit une histoire H_{i-1} . Pour passer à l'étape i , on regarde où est placé l'élément i dans la permutation : si i est en tête d'un cycle, alors l'évolution de l'urne sera en choisissant ∇ (on crée ainsi un nouveau cycle). Sinon on regarde l'élément qui précède i : appelons-le j : si $j < i$ alors il y a déjà une boule \bullet^j dans l'histoire de l'urne : c'est elle qui est sélectionnée. En revanche, si $j > i$, on regarde le premier élément avant j dans le cycle de la permutation qui est plus petit que i , et c'est celui-ci qui est sélectionné. (Son existence vient du fait que le premier élément de chaque cycle est le plus petit parmi les éléments du cycle).

Par exemple, pour (1 3)(2 5 4)(6). 1 et 2 sont en tête de cycle, donc on commence par $\xrightarrow{\nabla} \nabla$, puis 3 est juste après 1, donc on choisit $\xrightarrow{\bullet^1}$. Ensuite, 4 est après 5 (trop grand), lui même après 2, donc on a la transition $\xrightarrow{\bullet^2}$; puis 5 est après 2, donc on fait $\xrightarrow{\bullet^2}$. Pour 6, on termine par $\xrightarrow{\nabla}$. \square

La preuve de la proposition 1 découle de la construction donnée dans la preuve du lemme ci-dessus : si désormais, on a différentes couleurs de boules qui marquent le nombre de feuilles (de 1 à k) dans les sous arbres gauches de la branche droite, cela correspond exactement à marquer les cycles (de longueur 1 à k) dans les permutations.

Records et permutations. Il y a un lien particulier entre les records dans les permutations et le nombre de cycles :

Définition 3. *Un record dans une permutation est un élément qui n'a pas de plus grand à sa gauche : par exemple, 3 2 5 4 1 8 9 6 7.*

Proposition 2. *Les permutations de \mathcal{S}_n à k records sont en bijection avec les permutations à k cycles.*

PREUVE. En effet, il suffit de découper suivant les records : ainsi on obtient pour notre exemple : (3 2) (5 4 1) (8) (9 6 7). Une permutation à k records donne une permutation à k cycles, telle qu'il apparaît dans chaque cycle le plus grand élément en tête, et les cycles sont ordonnés selon les éléments de tête croissants.

Réciproquement, se donner une permutation à k cycles, c'est se donner une permutation à k records : il suffit de placer en tête de chaque cycle l'élément le plus grand, puis d'ordonner les cycles par éléments de tête croissants, enfin on oublie les parenthèses. \square

Ce résultat peut être affiné tout en utilisant la même construction bijective :

Proposition 3. *Les permutations avec s_1 cycles de longueur 1 (points fixes), s_2 cycles de longueur 2 (transpositions), \dots , s_k cycles de longueur k sont en bijection avec les permutations dont les records sont espacés s_1 fois de 1, s_2 fois de 2, \dots , s_k fois de k .*

Pour résumer cette partie, en utilisant les propositions 1 et 3 on peut donc énoncer le théorème suivant :

Théorème 3. *Pour k fixé, on a une bijection entre les histoires de l'urne (13) de longueur n finissant avec p_1 boules \bullet_1, \dots, p_k boules \bullet_k , les permutations de taille n avec p_1 points fixes, p_2 transpositions, \dots , p_k k -cycles, et les permutations de taille n avec p_1 records espacés de 1, \dots , p_k records espacés de k .*

Remarque. On retrouve ainsi un résultat bien connu en combinatoire : dans une permutation de taille très grande ($n \rightarrow \infty$), le nombre de records espacés de k suit une loi de Poisson de paramètre $\frac{1}{k}$. C'est également le cas pour le nombre de k -cycles dans une permutation : $\mathcal{P}(\frac{1}{k})$.

Pour revenir aux urnes, il est possible de compter directement le nombre de sous-arbres gauches de la branche droite, ce qui revient par la bijection, à compter le nombre de cycles d'une permutation. Examinons la série génératrice qui dénombre les permutations et où la variable u marque les cycles :

$$H(z, u) = \exp\left(u \ln\left(\frac{1}{1-z}\right)\right) = (1-z)^{-u}.$$

Cela correspond à la description combinatoire, selon les notations de [FS09], $\text{PERM} = \text{SET}(u\text{CYC}(\mathcal{Z}))$.

Le lien avec les arbres avec "modèle de croissance uniforme" est le suivant : à chaque nouveau cycle créé (i.e. lorsque l'on frappe ∇ , la Déesse Mère), il faut compter 1 (avec une boule \bullet qui ne disparaît plus et n'engendre pas de \bullet : une sorte de "Déesse Fille") . Ainsi les règles sont

$$\begin{array}{l} \nabla \rightarrow \bullet \nabla \\ \bullet \rightarrow \times \bullet \\ \times \rightarrow \times \times \end{array} \quad \text{d'où l'urne} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Cela donne donc $H(z, u)$ comme fonction génératrice (z marque la longueur et u marque le nombre de \bullet). On retrouve ainsi le résultat suivant bien connu :

Proposition 4. *Le nombre moyen de cycles dans une permutation de taille n est H_n , asymptotique à $\ln(n)$ lorsque $n \rightarrow \infty$, où $H_n = \sum_{j=1}^n \frac{1}{j}$.*

3. Urnes de croissance préférentielle

La théorie analytique des urnes a été développée pour les urnes bicolores soustractives (c'est-à-dire qu'au moins une des entrées diagonales de la matrice est négative), et pour quelques modèles triangulaires (cf [FDP06]). C'est un problème largement ouvert de l'étendre aux urnes "additives" (où tous les coefficients sont positifs). Le premier pas dans cette direction est constitué par une classe de modèles que j'ai trouvée qui a l'avantage de donner lieu à une solution explicite : les fonctions génératrices associées sont de fait des fonctions *algébriques*.

On considère donc dans cette partie la classe d'urnes équilibrées 2×2 suivante :

$$(16) \quad \begin{pmatrix} 2\alpha & \beta \\ \alpha & \alpha + \beta \end{pmatrix}, \text{ avec } \alpha > 0, \beta \geq 0.$$

EXEMPLE : Une première instanciation avec $\alpha = \beta = 1$ correspond aux règles d'évolution suivante :

$$\begin{array}{l} \mathbf{x} \rightarrow \mathbf{x} \mathbf{x} \mathbf{x} \mathbf{y} \\ \mathbf{y} \rightarrow \mathbf{x} \mathbf{y} \mathbf{y} \end{array} \quad \text{correspondant à l'urne} \quad \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

On peut y voir un modèle de croissance de population avec deux types d'individus (\mathbf{x} et \mathbf{y}). Chaque type a trois descendants et favorise son propre type avec un rapport 2/1.

Au vu de ce dernier exemple, on appellera désormais cette classe de modèles *urnes de croissance préférentielle*. Le résultat principal de cette section est précisément le fait que les fonctions génératrices sont algébriques. L'exploitation analytique sera effectuée ultérieurement, mais une loi gaussienne est attendue pour la loi limite. Dans la mesure où ces problèmes n'ont pas été traités en utilisant la théorie analytique, la compréhension de ces modèles sera utile pour la généralisation. Notons que ces urnes de croissance préférentielle ont deux paramètres, α et β , ce qui présage une classe assez expressive. L'analyse partielle qui suit la preuve de l'analyticité montre que les avancées sont ramenées ici à comprendre un changement de régime asymptotique et des transitions de singularités.

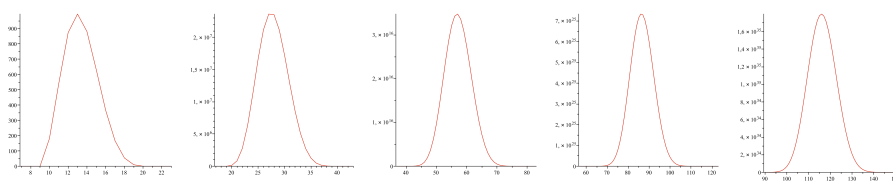


FIGURE 2. Nombre d'histoires de longueur n se finissant avec k boules x , pour l'urne $\alpha = \beta = 1$, avec $n = 10, 20, 40, 60, 80$. En abscisse, nombre de boules x ; en ordonnée, nombre d'histoires. La forme gaussienne apparaît rapidement.

Les premières caractéristiques sont les suivantes : la “balance” de l'urne — la somme des termes d'une ligne — est $\sigma = 2\alpha + \beta$. L'indice de dissymétrie — la différence des termes dans une colonne — est $p = -\alpha$. Remarquons que $-p$ et σ sont les deux valeurs propres de la matrice. Le rapport de ces deux valeurs propres est $\rho = -\frac{p}{\sigma} = \frac{\alpha}{2\alpha + \beta} \leq \frac{1}{2}$. D'après les analyses probabilistes de Smythe [Smy96] et Janson [Jan04], on s'attend à obtenir une loi limite gaussienne. On étudie l'évolution de cette urne dans le temps. On commence le processus, avec 0 boule de couleur x , et 1 boule de couleur y .

3.1. Une fonction génératrice algébrique.

Théorème 4. *La fonction génératrice exponentielle bivariée des histoires $H(x, z)$ est algébrique, et est solution de l'équation fonctionnelle en y :*

$$(17) \quad (z - A - B(x))y^{2\alpha+\beta} + B(x)y^\alpha + A = 0$$

$$\text{où } A := \frac{1}{2\alpha + \beta} \text{ et } B(x) := \frac{x^{-\alpha} - 1}{\alpha + \beta}.$$

PREUVE. La série exponentielle des histoires est

$$H(x, y, z) = \sum_{n,a,b} H_n(a, b) x^a y^b \frac{z^n}{n!}.$$

Le système différentiel associé à l'urne est

$$\begin{cases} \dot{X} &= X^{2\alpha+1} Y^\beta \\ \dot{Y} &= X^\alpha Y^{\alpha+\beta+1}. \end{cases}$$

Soient $X(x, y, z)$ et $Y(x, y, z)$ les solutions associées au système, alors par le théorème 1, comme $a_0 = 0$ et $b_0 = 1$, on a $H(x, y, z) = Y(x, y, z)$. En réécrivant le système,

$$\frac{\dot{X}}{X^{\alpha+1}} = X^\alpha Y^\beta = \frac{\dot{Y}}{Y^{\alpha+1}} \text{ donc } X^{-\alpha} - Y^{-\alpha} = x^{-\alpha} - y^{-\alpha}.$$

Alors,

$$\frac{\dot{Y}}{Y^{\alpha+\beta+1}}(Y^{-\alpha} + x^{-\alpha} - y^{-\alpha}) = 1.$$

En intégrant par rapport à z et en prenant $-K(x, y)$ comme constante d'intégration,

$$\frac{1}{2\alpha + \beta} Y^{-(2\alpha+\beta)} + \frac{x^{-\alpha} - y^{-\alpha}}{\alpha + \beta} Y^{-(\alpha+\beta)} = -(z - K(x, y)).$$

On peut prendre $y = 1$ (variable superflue car l'urne est équilibrée), ce qui donne l'équation annoncée. La fonction génératrice $H(x, z)$ est donc algébrique, car solution d'un polynôme de degré $2\alpha + \beta$. \square

3.2. Moyenne et variance. De cette équation polynomiale, on peut en premier lieu extraire des informations précises sur la moyenne et la variance de la composition de l'urne.

Proposition 5. *Soit X_n la variable aléatoire comptant le nombre de boules de couleur \mathbf{x} dans l'urne (16) au temps n . Alors*

$$(18) \quad \mathbb{E}(X_n) = \frac{\alpha(2\alpha + \beta)}{\alpha + \beta} n + \frac{\alpha}{\alpha + \beta} \frac{\Gamma(\frac{1}{2\alpha+\beta})}{\Gamma(\frac{\alpha+1}{2\alpha+\beta})} n^{\frac{\alpha}{2\alpha+\beta}} + \frac{\alpha}{\alpha + \beta} + O\left(n^{\frac{\alpha}{2\alpha+\beta}-1}\right),$$

$$(19) \quad \mathbb{V}(X_n) = \frac{\alpha^3(2\alpha + \beta)}{(\alpha + \beta)^2} n + O(n^{\frac{\alpha+\beta}{2\alpha+\beta}}) \text{ où } \Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt.$$

PREUVE. Pour cela, on différencie par rapport à x l'équation (17), puis on spécifie $x = 1$; on l'exprime sous forme de développement asymptotique, puis on normalise par le développement de $(1 - \sigma z)^{-1/\sigma}$. Pour la variance, on différencie deux fois l'équation, et on use des mêmes ingrédients. \square

Remarque. Ici, en plus de l'asymptotique, on a accès également à l'expression exacte de la moyenne et de la variance. On ne donne que l'expression de la moyenne; celle de la variance s'obtient de la même façon en utilisant l'expression exacte de la dérivée seconde, prise en $x = 1$.

$$\mathbb{E}(X_n) = \frac{[z^n] \frac{\partial H(x, z)}{\partial x} \Big|_{x=1}}{[z^n] (1 - \sigma z)^{-1/\sigma}} \text{ et } \frac{\partial H(x, z)}{\partial x} \Big|_{x=1} = \frac{\alpha}{\alpha + \beta} \left((1 - \sigma z)^{-(1+\frac{1}{\sigma})} + (1 - \sigma z)^{-\frac{\alpha+1}{\sigma}} \right).$$

Pour notre exemple de l'urne $(\frac{2}{1} \frac{1}{2})$, la moyenne et la variance du nombre X_n de boules de couleur \mathbf{x} sont

$$\mathbb{E}(X_n) = \frac{3}{2}n + \frac{\Gamma(1/3)}{2\Gamma(2/3)}n^{1/3} + \frac{1}{2} + O(n^{-2/3}) \text{ et } \mathbb{V}(X_n) = \frac{3}{4}n + O(n^{2/3}).$$

L'équation de la série génératrice $H(x, z)$ est : $(z - K(x))H^3 + \frac{1-x}{2x}H + \frac{1}{3} = 0$, où $K(x) = \frac{1}{3} - \frac{x-1}{2x} = \frac{1}{2x} - \frac{1}{6}$.

3.3. Vers la loi normale. Dans la démarche classique de la combinatoire analytique, après avoir obtenu une expression de la fonction génératrice de dénombrement, on s'intéresse aux singularités de cette fonction. En effet, ce sont ces singularités qui déterminent l'asymptotique des coefficients. Voici donc quelques observations sur les singularités de la fonction $H(x, z)$.

Singularités. Une singularité de l'équation (17) est

$$\rho(x) := A + B(x) = \frac{1}{2\alpha + \beta} - \frac{x^\alpha - 1}{x^\alpha(\alpha + \beta)}.$$

En effet, pour $z = \rho(x)$, $(z - \rho(x))$ annule le coefficient du monôme dominant du polynôme annulateur de $H(x, z)$.

Propriétés de $\rho(x)$:

$$\rho(x) = \frac{1}{\sigma} - \frac{x^\alpha - 1}{x^\alpha(\alpha + \beta)}, \quad \lim_{x \rightarrow 0} \rho(x) = +\infty \text{ et } \rho(x) \sim_{x \rightarrow 0} \frac{1}{x^\alpha(\alpha + \beta)}.$$

$$\rho(x) = \frac{1}{\sigma} \text{ pour tout } x = \exp\left(\frac{2ik\pi}{\alpha}\right), \text{ pour } k \in \{0 \dots \alpha - 1\}.$$

$$\rho(x) = 0 \text{ pour } x = \left(\frac{-\sigma}{p}\right)^{1/\alpha} \exp\left(\frac{2ik\pi}{\alpha}\right), \text{ pour } k \in \{0 \dots \alpha - 1\}.$$

Lemme 3. $H(z, x)$ se développe en série au voisinage de la singularité $\rho(x)$ pour z , et au voisinage de $x = 1$.

PREUVE. Notons $\kappa := (\frac{-\sigma}{p})^{1/\alpha}$, ainsi pour α pair, $\rho(1) = \rho(-1) = \sigma^{-1}$, d'où $\Re(\rho(x)) > 0$ si $|x| < \kappa$. Pour α impair, $\rho(-1) < 0$ et $\Re(\rho(x)) > 0$ si $\Re(x) > 0$ et $|x| < \kappa$. Ici, on a $\frac{-\sigma}{p} > 2$, donc $\kappa > \sqrt[3]{2} > 1$ et on peut alors développer en série au voisinage de $x = 1$. \square

Pour la suite de l'étude, on pose $B(x) = \frac{x^{-\alpha} - 1}{\alpha + \beta}$. Voici un lemme qui traite du cas "critique" lorsque x tend vers 0. On remarque ainsi que l'équation reste cohérente dans ce cas dégénéré.

Lemme 4. $H(x_0 = 0, y_0 = 1, z) = x_0^{a_0}$:

- si $a_0 = 0$, $H(0, 1, z) = 1$: il n'y a qu'une histoire qui se termine dans la configuration sans boule \mathbf{x} : c'est celle qui commence sans boule \mathbf{x} .
- si $a_0 \neq 0$, $H(0, 1, z) = 0$: il est impossible de faire disparaître des boules de l'urne : si on commence avec des boules \mathbf{x} , on ne peut pas les enlever.

PREUVE. $Y(z, x)$ vérifie l'équation

$$Y^{2\alpha+\beta} + \frac{B(x)}{z - \rho(x)} Y^\alpha + \frac{\sigma^{-1}}{z - \rho(x)} = 0 \text{ avec } B(x) = \frac{x^{-\alpha} - 1}{\alpha + \beta}.$$

Prenons la limite quand x tend vers 0 : $\frac{B(x)}{z - \rho(x)} \rightarrow -1$ et $\frac{\sigma^{-1}}{z - \rho(x)} \rightarrow 0$, alors l'équation devient $Y^{\alpha+\beta} = 1$. On obtient $Y(z, 0, 1) = y_0 = 1$, $X(z, 0, 1) = 0$. Ainsi, $H(x_0 = 0, y_0 = 1, z) = x_0^{a_0}$. \square

Pour comprendre l'influence du paramètre x dans l'équation, et pour étudier ses singularités, on reprend notre exemple de l'urne avec $\alpha = \beta = 1$.

Lemme 5. *Les singularités de l'équation (17) vérifiée par la fonction génératrice sont*

$$\rho(x) := B(x) + \sigma^{-1} \text{ et } \tilde{\rho}(x) = \rho(x) - \frac{4B(x)^3}{27\sigma^{-2}}.$$

De plus, on a le schéma suivant :

- si $x > 1$, $\rho(x) < \tilde{\rho}(x) < \sigma^{-1}$,
- si $x < 1$, $\sigma^{-1} < \tilde{\rho}(x) < \rho(x)$,
- si $x = 1$, $\rho(1) = \tilde{\rho}(1) = \sigma^{-1}$.

PREUVE. Soit F la fonction bivariable suivante : $F(z, Y(z)) = (z - B(x) - \sigma^{-1})Y^3 + B(x)Y + \sigma^{-1}$. On résout le système où z est l'inconnue :

$$\left\{ \begin{array}{l} F(z, Y(z)) = 0 \\ \frac{\partial}{\partial Y} F(z, Y(z)) = 0 \end{array} \right. \text{ i.e. } \left\{ \begin{array}{l} (z - \rho(x))Y^3 + B(x)Y + C = 0 \\ 3(z - \rho(x))Y^2 + B(x) = 0 \end{array} \right. .$$

On calcule le résultant de F et de $\frac{\partial}{\partial Y} F$ afin de trouver les singularités et l'on obtient

$$(z - B(x) - \sigma^{-1})^2 \left(z - \left(B(x) + \sigma^{-1} - \frac{4B(x)^3}{27\sigma^{-2}} \right) \right).$$

Les zéros de ce polynôme en z donnent les singularités $\rho(x)$ et $\tilde{\rho}(x)$. Pour la position des singularités, on utilise l'écriture de $\tilde{\rho}$ suivante : $\tilde{\rho}(x) = \sigma^{-1} + B(x) \left(1 - \frac{4B(x)^2}{27\sigma^{-2}} \right)$ et l'expression $\left(1 - \frac{4B(x)^2}{27\sigma^{-2}} \right)$ est positive dans un voisinage de $x = 1$. \square

Une fois les singularités trouvées, l'objectif est de trouver une expression pour $Y(z)$ dans un voisinage de $z = \rho(x)$, et $z = \tilde{\rho}(x)$. Le dernier lemme répond à cette question.

Lemme 6. *Soit x voisin de 1, avec $x \neq 1$.*

Dans un voisinage de $z = \tilde{\rho}(x)$, on a

$$Y(z) = \frac{-3\sigma^{-1}}{2B(x)} + \frac{9\sigma^{-2}}{4B(x)^{5/2}} \sqrt{z - \tilde{\rho}(x)} + \varepsilon(x, z),$$

où $\varepsilon(x, z)$ est tel que pour x voisin de 1 et $x \neq 1$, $\varepsilon(x, z) = o(z - \tilde{\rho}(x))$, pour $z \rightarrow \tilde{\rho}(x)$.

Dans un voisinage de $z = \rho(x)$, on a

$$Y(z) = \sqrt{\frac{B(x)}{z - \rho(x)}} + \varepsilon(x, z),$$

où $\varepsilon(x, z)$ est tel que pour x voisin de 1 et $x \neq 1$, $\varepsilon(x, z) = o(z - \rho(x))$, pour $z \rightarrow \rho(x)$.

PREUVE. Ces expressions asymptotiques sont obtenues par la méthode de Newton-Puiseux (cf [FS09]). Cela consiste en un changement de variable et une recherche d'exposant et de coefficient dominant.

Pour $z = \tilde{\rho}(x)$, on trouve $Y(z = \tilde{\rho}(x)) = \frac{-3\sigma^{-1}}{2B(x)}$. On applique donc le changement de variable suivant :

$\mathcal{Z} = z - \tilde{\rho}(x)$, $\mathcal{Y} = Y + \frac{3\sigma^{-1}}{2B(x)}$; on recherche ensuite l'exposant ξ tel que $\mathcal{Y} = c\mathcal{Z}^\xi (1 + o(1))$. On trouve

$\xi = 1/2$ et $c = \frac{9\sigma^{-2}}{4B(x)^{5/2}}$. Alors,

$$Y(z) = \frac{-3\sigma^{-1}}{2B(x)} + \frac{9\sigma^{-2}}{4B(x)^{5/2}} \sqrt{z - \tilde{\rho}(x)} + \varepsilon(x, z).$$

Regardons Y^{-1} :

$$(20) \quad Y^{-1}(z) = \frac{-2B(x)}{3\sigma} - \sqrt{\frac{z - \tilde{\rho}(x)}{B(x)}} + \epsilon(x, z).$$

Pour $z = \rho(x)$, on applique la même méthode, cette fois avec $\mathcal{Z} = z - \rho(x)$, $\mathcal{Y} = Y^{-1}$. Le développement de Newton-Puiseux donne

$$(21) \quad Y^{-1}(z) = \sqrt{\frac{z - \rho(x)}{B(x)}} + \epsilon(x, z).$$

□

Le problème majeur qui se pose est le suivant : lorsque $x \rightarrow 1$, on devrait obtenir une expression de $Y(z)$ proportionnelle à $(C-z)^{-1/3}$. En effet, l'équation de départ lorsque $x = 1$, donne $B(x) = 0$ et $K(x) = \rho(x) = -1/3$, et donc :

$$Y^{-3} = (1 - 3z).$$

On retrouve le fait que, dans le cas univarié (lorsque $x = 1$),

$$H(z) = (1 - \sigma z)^{-(a_0+b_0)/\sigma}.$$

Lorsque $x = 1$, on doit donc obtenir un développement asymptotique en $z = \sigma^{-1} = 1/3$ avec une racine cubique. Or ce que l'on obtient dans le cas bivarié (x varie autour de 1, mais reste différent de 1), donne des asymptotiques en racine carrée. Les résultats ne sont pas incompatibles dans la mesure où le coefficient $B(x)$ s'annule lorsque $x = 1$. En effet, dans les développements asymptotiques donnés, on a considéré chaque singularité séparément. Or, lorsque x tend vers 1, dans un voisinage de $x = 1$, les deux singularités $\rho(x)$ et $\tilde{\rho}(x)$ se retrouvent dans le domaine de z , puisqu'elles se rejoignent en $x = 1$. Les deux contributions autour des deux singularités sont donc à prendre en compte. On remarque d'ailleurs que les deux expressions de Y^{-1} (20 et 21) sont similaires, ce qui laisse à penser que les deux termes principaux se compensent, laissant ainsi apparaître une asymptotique dominante en racine cubique. Dans des contextes différents, des questions de confluence de singularités et de changement de régime apparaissent dans [Dur04], [BFSS01] et [FS09] (p. 704–714).

4. Urnes et k -arbres

Dans cette dernière section, on s'intéresse à une autre classe de modèles d'urnes, avec un paramètre. Ces urnes sont également des urnes 2×2 équilibrées et "additives". Ici, il n'apparaît pas de confluence de singularités. On réussit à obtenir des développements asymptotiques valables dans des domaines complexes autour de la singularité. Par la théorie analytique, on arrive alors à en déduire une loi limite pour le nombre de boules d'une certaine couleur, ainsi que des résultats jusqu'alors inaccessibles avec les méthodes probabilistes, tels que des vitesses de convergence, des propriétés de grandes déviations, et des lois locales limites. Mais commençons par motiver l'étude de cette classe d'urnes.

Les k -arbres croissants reflètent l'évolution du Web. Les pages Web sont numérotées dans leur ordre de création, et lorsqu'une nouvelle page Web est créé, elle possède des liens vers k

pages Web déjà existantes. Le choix de ces pages Web “mères” est considéré uniforme. Les 1-arbres ordonnés croissants correspondent davantage aux réseaux sociaux, avec un modèle d’attachement préférentiel : plus une personne est influente (plus ses connaissances sont élevées), plus elle a de chance d’attirer un nouvel utilisateur (cf fig.3). Les k -arbres ordonnés croissants correspondent au même type d’évolution, mais l’attachement se fait directement sur une clique de k utilisateurs (un groupe de k personnes qui se connaissent). C’est ce dernier modèle que nous allons étudier. D’autres types de k -arbres sont étudiés dans [Dar10] et [DHBS10].

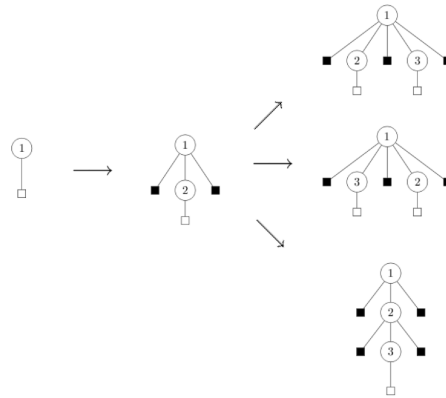


FIGURE 3. 1-arbres ordonnés croissants, ou arbres planaires croissants de Catalan de taille 1, 2, et 3, avec les boules représentées, correspondant aux différentes possibilités d’ajouter un sommet.

4.1. La modélisation des k -arbres par une urne 2×2 . L’article de Panholzer et Seitz⁴ [PS10] est à l’origine de ce travail. En effet, les auteurs travaillent sur un modèle de réseaux à grande échelle : les k -arbres ordonnés croissants et étudient un certain nombre de paramètres, avec des méthodes analytiques. Certaines formules, similaires à celles trouvées lors de travaux sur les urnes analytiques, nous ont poussé à faire le lien entre les 2 modèles.

Définition 4. Un k -arbre se construit récursivement : Soit c est une k -clique (graphe complet à k sommets); soit il existe un sommet f tel que f a exactement k voisins qui forment une k -clique, et $T \setminus f$ est un k -arbre.

On parle de k -arbre ordonné si on peut mettre un ordre sur les fils d’une même k -clique : on a un premier fils, un deuxième, Cela équivaut à introduire une notion de gauche et droite pour les fils d’une même k -clique.

Enfin, la notion de croissance est la même que pour les arbres classiques : on étiquette tous les sommets de l’arbre, et un sommet fils doit avoir une étiquette plus grande que tous ses parents (ici, un sommet fils a k sommets parents, qui forment une k -clique). On parle alors de k -arbre ordonné croissant.

Une k -clique est sans descendance lorsqu’elle n’a pas de sommet fils. Pour $k = 1$, on retrouve la définition des arbres planaires croissants de Catalan (dits PORT) (cf fig.3).

4. AofA 2010 - Université de Vienne

Dans toute la suite, on utilise la terminologie *k*-*arbre* pour désigner les *k*-arbres ordonnés croissants. On cherche donc à décrire la croissance d'un *k*-arbre par un modèle d'urne. Pour cela, on utilise deux types de boules : \square marque la possibilité de branchement à une clique sans descendance, et \blacksquare marque les autres possibilités de branchement (à des cliques ayant au moins un descendant). Le plus petit *k*-arbre possible est simplement une *k*-clique, que l'on nommera *k*-clique *racine*. Elle est composée de *k* sommets tous reliés, notés $0_1, \dots, 0_k$. Cette *k*-clique racine a une seule possibilité de croissance : c'est d'ajouter un sommet étiqueté par 1, relié à tous les sommets racines. Cette première étape correspond à la configuration initiale de notre modèle d'urne : au départ, l'urne ne contient qu'une boule \square . D'après la règle de construction des *k*-arbres ordonnés croissants, on a les règles suivantes pour notre urne bicolore : à l'étape *n*,

- si on tire \square , on a donc tiré une *k*-clique *K* sans descendance. En ajoutant un sommet fils, étiqueté par *n*, à celle-ci, cela va créer *k* nouvelles *k*-cliques (sans descendance), chacune contenant le nouveau sommet *n* et (*k* – 1) sommets de *K*. Cela crée donc *k* boules \square . En revanche, il faut enlever l'ancien \square tiré, puisque désormais la clique *K* a un fils (*n*). Ensuite, il faut ajouter deux boules \blacksquare , correspondant aux nouvelles possibilités d'attachement à *K* : à gauche ou à droite du sommet *n*. Finalement, on ajoute (*k* – 1) boules \square et deux boules \blacksquare .
- si on tire \blacksquare , c'est le même principe ; en ajoutant un nouveau sommet *n*, cela ajoute *k* nouvelles *k*-cliques sans descendance, chacune contenant ce sommet feuille *n* ; on ajoute donc *k* \square . Et on a deux nouvelles possibilités d'attachement : à gauche ou à droite du nouveau sommet *n*, soit deux nouvelles boules \blacksquare . Enfin, il faut enlever la boule tirée, d'où un bilan de *k* boules \square ajoutées, et une \blacksquare .

L'urne correspondant aux *k*-arbres est ainsi

$$(22) \quad U_k := \begin{pmatrix} k-1 & 2 \\ k & 1 \end{pmatrix}.$$

Voici une proposition préliminaire sur la forme de la fonction génératrice.

Proposition 6. *La fonction génératrice bivariée des histoires de l'urne U_k , $H(x, z)$ vérifie*

$$(23) \quad \frac{1}{H^k(H-b)^2} = \frac{a_1}{H-b} + \frac{a_2}{(H-b)^2} + \frac{c_1}{H} + \dots + \frac{c_k}{H^k}$$

avec

$$a_1 = \frac{-k}{b^{k+1}}, \quad a_2 = \frac{1}{b^k}, \quad c_1 = \frac{k}{b^{k+1}}, \quad \dots, \quad c_i = \frac{k-i+1}{b^{k-i+2}}, \quad \dots, \quad c_k = \frac{1}{b^2}.$$

et $b = x - 1$.

PREUVE. La configuration initiale est $a_0 = 1$ et $b_0 = 0$. Par le théorème 1,

$$H(x, z) = X(x, z).$$

Le système différentiel associé est : $\dot{X} = \dot{Y} = X^k Y^2$. Ainsi, $\dot{Y} Y^{-2} = X^k$. En intégrant puis inversant, $Y(z) = (1 - \int_0^z X^k(u) du)^{-1}$. Finalement,

$$X(z) = Y(z) + x_0 - 1 = x_0 - 1 + \frac{1}{1 - \int_0^z X^k(u) du} \quad (\text{on pose } y_0 = 1).$$

Pour résoudre l'équation, on pose

$$b := x_0 - 1.$$

Ainsi, $(X - b)^{-1} = 1 - \int_0^z X^k(u) du$. On dérive par rapport à z :

$$\frac{X'}{X^k(X - b)^2} = 1.$$

La décomposition en éléments simples donne le résultat. \square

Remarque. Définissons $T(z) := X(z)$ et $S(z) := \int_0^z X^k(u) du$, ce qui permet d'obtenir la description suivante :

$$T(z) = \frac{1}{1 - S(z)} \text{ et } S'(z) = T(z)^k. \quad T(0) = 1, S(0) = 0.$$

On retrouve ainsi la description combinatoire des k -arbres : $X = \text{SEQ}(Z^\square \star X^k)$, en utilisant les règles de combinatoire symbolique classique [FS09]. Ici, $Z^\square \star X^k$ est le produit d'un atome étiqueté et de k éléments de X .

Toute la section est consacrée à l'étude de cette urne. La démarche est la suivante : on traite tout d'abord du cas $k = 1$ en détail. Cela donne lieu à une loi limite gaussienne (théorème 5 p.20), à une loi locale limite (théorème 6 p.20), et à un principe de grandes déviations (théorème 7 p.24). Ceci est ensuite généralisé à tous les cas dans la dernière partie, ce qui donne le théorème général (théorème 8 p.25).

4.2. Loi normale, loi locale limite et grandes déviations pour les 1-arbres. Soit X_n la variable aléatoire comptant le nombre de boules \square dans l'urne U_1 après n étapes. Cela correspond, dans le cas $k = 1$, au nombre de feuilles dans un PORT de taille n (les 1-cliques sans descendance dans un 1-arbre sont les feuilles).

Théorème 5. (Loi limite normale) La variable aléatoire X_n a une moyenne μ_n et une variance σ_n^2 , et la variable aléatoire normalisée $\frac{X_n - (2/3)n}{\sqrt{n/3}}$ converge en loi vers une loi gaussienne centrée réduite $\mathcal{N}(0, 1)$, avec vitesse de convergence $O\left(\frac{1}{\sqrt{n}}\right)$.

$$\mu_n = \mathbb{E}(X_n) = \frac{2}{3}(n + 1/2) + O\left(\frac{1}{n}\right), \quad \sigma_n^2 = \mathbb{V}(X_n) = \frac{1}{9}(n + 1/2) + O\left(\frac{1}{n}\right),$$

$$\mathbb{P}\left\{\frac{X_n - \frac{2}{3}n}{\sqrt{\frac{n}{9}}} \leq t\right\} = \Phi(t) + O\left(\frac{1}{\sqrt{n}}\right), \quad \text{où } \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{v^2}{2}} dv.$$

Théorème 6. (Loi limite locale) Notons $p_{n,k} = \mathbb{P}\{X_n = k\}$. La distribution des X_n satisfait une loi locale limite de type gaussienne avec vitesse de convergence $O\left(\frac{1}{\sqrt{n}}\right)$, i.e.

$$\sup_{t \in \mathbb{R}} \left| \frac{\sqrt{n}}{3} p_{n, \lfloor 2n/3 + t\sqrt{n}/3 \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leq \frac{1}{\sqrt{n}}.$$

Remarque. Le fait que $\frac{X_n - (2/3)n}{\sqrt{n/3}}$ converge en loi vers une loi gaussienne centrée réduite est un résultat déjà connu, donc la démonstration fait appel à des méthodes probabilistes. On peut trouver ce résultat dans [Mah08] p.160, ou [Mah03].

Ici, les démonstrations de ces théorèmes reposent sur les théorèmes de Quasi-Puissances, et de Loi-Locale Limite de [FS09]. Le point clé est l'obtention d'une approximation de la

fonction génératrice de probabilité de X_n , définie par :

$$p_n(x) = \frac{n! [z^n] X(z, x)}{n! [z^n] X(z, 1)}.$$

Pour le théorème des quasi-puissances, il faut une approximation uniforme dans un voisinage de $x = 1$. Pour le théorème de Loi-Locale-Limite, il faut en plus étudier sur tout le cercle unité $|x| = 1$, afin d'obtenir une borne sur $p_n(x)$.

Rappelons que X et H représentent la même fonction. On travaille avec les variables z et $b := x_0 - 1$. On veut observer le comportement de la fonction X dans un voisinage de 1 pour x_0 . On fait donc varier x_0 autour de 1. Dans la suite, on utilisera simplement la notation x .

Proposition 7. Soit $\tilde{X}(b, z) = 1 + \frac{b}{X(b, z) - b}$. Alors \tilde{X} vérifie

$$(24) \quad \tilde{X}e^{-\tilde{X}} = e^{-1}e^{-b^2(K(b)-z)}, \quad \text{où } K(b) = -\frac{1}{b^2} \ln(1+b) + \frac{1}{b}.$$

PREUVE. Reprenons l'équation (23) avec $k = 1$.

$$\frac{X'}{X(X-b)^2} = 1 = -\frac{1}{b^2} \frac{X'}{X-b} + \frac{1}{b} \frac{X'}{(X-b)^2} + \frac{1}{b^2} \frac{X'}{X}.$$

Alors, l'intégration donne

$$z - K(b) = -\frac{1}{b^2} \ln(X-b) - \frac{1}{b} \frac{1}{X-b} + \frac{1}{b^2} \ln(X) = -\frac{1}{b^2} \ln\left(1 - \frac{b}{X}\right) + \frac{1}{b} \frac{1}{X-b}.$$

$K(b)$ est la constante d'intégration déterminée par la condition initiale $X(x, 0) = x$. On trouve $K(b) = -\frac{1}{b^2} \ln(1+b) + \frac{1}{b}$. Par le changement de variable $\tilde{X} = 1 + \frac{b}{X-b}$, i.e. $X = \frac{-b\tilde{X}}{1-\tilde{X}}$, on trouve une écriture plus appropriée :

$$1 + b^2(K(b) - z) = \tilde{X} - \ln(\tilde{X}).$$

En composant par l'exponentielle, on découvre l'équation fonctionnelle suivante :

$$\tilde{X} \exp(-\tilde{X}) = \exp(-1 - b^2(K(b) - z)).$$

□

Remarquons simplement que si $b \rightarrow 0$ (i.e. $x \rightarrow 1$), $K(b) = \frac{1}{2} - \frac{b}{3} + o(b)$. De plus, l'équation vérifiée par \tilde{X} introduit la fonction $x \mapsto xe^{-x}$, inverse de la fonction d'arbre.

Définition 5 (Définition – Propriété). ([FS09] p.403) Considérons la fonction suivante $f(y) = ye^{-y}$. La fonction inverse de f est T , la fonction d'arbre. T est définie pour $z \in \mathbb{C}$ tel que $|z| < e^{-1}$. En fait, T est analytique sur \mathbb{C} privée de la demi-droite $[e^{-1}, \infty[$, et on a

$$T(z) = 1 - \sqrt{2}(1 - ez)^{1/2} + \frac{2}{3}(1 - ez) + O((1 - ez)^{3/2}).$$

$$T(e^{-1}) = 1 \quad \text{et} \quad T(0) = 0.$$

Remarque. La fonction T est appelée fonction d'arbre car $T(z) = z \exp(T(z))$, et la description combinatoire des arbres est $\mathcal{T} = \mathcal{Z} \text{ SET}(\mathcal{T})$ ⁵.

Définition 6. Un Δ -domaine $\Delta(R, \phi)$ en z_0 est un domaine du plan complexe défini par les z tels que $|z| < R$ et $|\arg(z - z_0)| > \phi$.

5. Cela correspond à la description des arbres non planaires étiquetés, dits arbres de Cayley

Lemme 7. *Pour tout b dans un voisinage de 0 et z dans un Δ -domaine $\Delta(R, \phi)$ en $K(b)$, on a*

$$(25) \quad X(z, b) = (2K(b) - 2z)^{-1/2} - \frac{2}{3}b + \frac{1}{12}b^2 \sqrt{2K(b) - 2z} + b^2 \cdot o(\sqrt{2K(b) - 2z}).$$

On peut remarquer que lorsque $b \rightarrow 0$ (i.e. $x \rightarrow 1$), on retrouve la fonction génératrice classique de l'urne $(1 - 2z)^{-1/2}$ (ici, $\sigma = 2$).

PREUVE. On normalise \tilde{X} en effectuant le changement de variable suivant $\zeta = K(b)z$. Ainsi (24) se réécrit

$$\widehat{X}_b(z) = \tilde{X}(zK(b)) \quad \text{et} \quad \widehat{X}_b(z)e^{-\widehat{X}_b(z)} = e^{-1} \exp(-b^2 K(b)(1-z)) =: a(b, z).$$

Si $a(b, z)$ est dans un domaine valide, on peut écrire $\widehat{X} = T(a(b, z))$. Discutons alors de la validité de l'inverse. Commençons par fixer un Δ -domaine pour z , $\Delta(R, \phi)$ en 1 et voyons ce qu'il devient par la transformation $a(b, z)$. Par $z \mapsto -(1-z)$, on obtient $\Delta(R, \phi)$ en 0; puis $\Delta(b^2 K(b)R, \phi)$ en 0, après multiplication par $b^2 K(b)$. Enfin, en composant par l'exponentielle, on a $\Delta(\exp(b^2 K(b)R), \phi)$ en 1. Ainsi, pour b assez petit, le Δ -domaine est inclus dans le domaine de validité de l'approximation de T . Il est alors possible d'avoir une approximation de \widehat{X} valide pour b dans un voisinage de 0, et z dans le Δ -domaine fixé $\Delta(R, \phi)$ en 1 :

$$\widehat{X}(z, b) = 1 - \sqrt{2}(1 - e^{-b^2 K(b)(1-z)})^{1/2} + \frac{2}{3}(1 - e^{-b^2 K(b)(1-z)}) + O((1 - e^{-b^2 K(b)(1-z)})^{3/2}).$$

Mais

$$1 - e^{-s} = s + O(s^2) \quad \sqrt{1 - e^{-s}} = \sqrt{s} + O(s^{3/2}) \quad (1 - e^{-s})^{3/2} = O(s^{3/2}).$$

Finalement,

$$\widehat{X}(z, b) = 1 - \sqrt{2}(b^2 K(b)(1-z))^{1/2} + \frac{2}{3}(b^2 K(b)(1-z)) + O((b^2 K(b)(1-z))^{3/2}).$$

Donc, en revenant à $X(z, b)$,

$$X(z, b) = (2K(b) - 2z)^{-1/2} - \frac{2}{3}b + \frac{1}{12}b^2 \sqrt{2K(b) - 2z} + b^2 \cdot o(\sqrt{2K(b) - 2z}).$$

□

Lemme 8. *La fonction génératrice de probabilité de la variable aléatoire X_n (qui compte le nombre de boules \square après n étapes) est donnée, pour b dans un voisinage de 0, par*

$$(26) \quad p_n(b) = \left(\frac{1}{2K(b)} \right)^{n+1/2} \left(1 + O\left(\frac{1}{n} \right) \right).$$

PREUVE. Appliquons maintenant l'analyse de singularité ([FS09]) : la singularité principale est en $z = K(b)$. Alors asymptotiquement, en reprenant le résultat du lemme 7,

$$n! [z^n] X(z, b) = \frac{K(b)^{-n}}{\sqrt{2K(b)}} \frac{n^{-1/2}}{\Gamma(1/2)} \left(1 + O\left(\frac{1}{n} \right) \right).$$

Et la probabilité $p_n(b)$ est $\frac{n! [z^n] X(z, b)}{n! [z^n] X(z, 0)}$, d'où, $p_n(b) = \left(\frac{1}{2K(b)} \right)^{n+1/2} \left(1 + O\left(\frac{1}{n} \right) \right)$. □

On peut donc conclure la preuve du théorème 5.

PREUVE. (Th.5) Le lemme 8 est l'hypothèse principale pour appliquer le théorème des quasi-puissances ([FS09] p.645). Il reste à vérifier la condition de "variabilité". soit $B(b) := (2K(b))^{-1}$. $B(0) = 1$ et B est analytique dans un voisinage de 0.

$$B''(0) + B'(0) - (B'(0))^2 = 1/9 \neq 0 \quad \text{et} \quad B'(0) = \frac{2}{3}.$$

□

Loi locale limite :

PREUVE. (Th.6) Regardons ce qui se passe lorsque x se promène sur le cercle unité tout entier : $|x| = 1$, i.e. $b = -1 + e^{i\theta}$, $\theta \in]-\pi, \pi[$.

Par le lemme 8, on a une approximation de la fonction génératrice de probabilité $p_n(x)$, pour x dans un petit voisinage \mathcal{V} de 1. Il reste simplement à majorer $p_n(x)$ lorsque $|x| = 1$ et x en dehors de \mathcal{V} . On utilise la même stratégie que dans la démonstration du lemme 7. Il faut simplement remarquer que pour tout x de module 1, $-b^2K(b)$ est bien défini et est borné : $|-b^2K(b)| < 4$. (cf fig.4 gauche). Par l'approximation de $p_n(u)$, on déduit que pour n suffisamment grand : $|p_n(u)| < (2K(u))^{-n-1/2}(1+\epsilon)$. Comme pour tout x de module 1 et hors du voisinage \mathcal{V} , on a $2K(u) > \kappa > 1$ (cf. fig.4 droite), on peut borner, pour n suffisamment grand, $|p_n(u)| < \kappa^{-n}$. Toutes les hypothèses du théorème de Loi-locale-limite sont ainsi vérifiées ([FS09] p.696). □

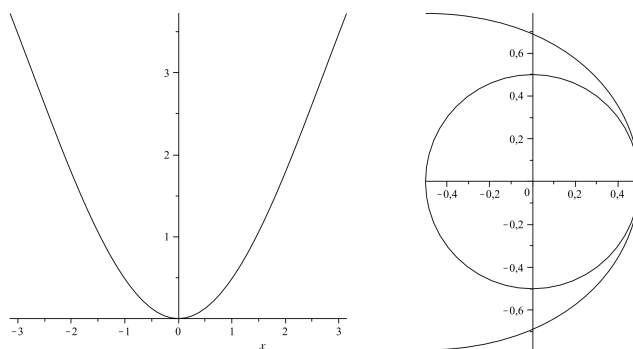


FIGURE 4. À gauche, graphe de $|b^2K(b)|$ pour $|b+1| = 1$. À droite, graphe de $K(b)$ et cercle de rayon $1/2$ pour $|b+1| = 1$.

Grandes déviations :

Il est possible d'obtenir un résultat de grandes déviations sur la variable aléatoire étudiée X_n .

Définition 7. (*Propriété de grandes déviations*) Soit a_n une suite tendant vers l'infini. Une suite X_n de variables aléatoires, de moyenne $\mathbb{E}_n \sim \mu a_n$, satisfait une propriété de grandes déviations, relative à l'intervalle $[t_0, t_1]$ contenant μ , s'il existe une fonction $W(t)$ telle que $W(t) > 0$ pour $t \neq \mu$, et pour n suffisamment grand,

$$\forall t, t_0 < t < \mu, \frac{1}{a_n} \log \mathbb{P}(X_n \leq ta_n) = -W(t) + o(1) \quad (\text{queue gauche})$$

$$\forall t, \mu < t < t_1, \frac{1}{a_n} \log \mathbb{P}(X_n \geq ta_n) = -W(t) + o(1) \quad (\text{queue droite}).$$

$W(t)$ est la fonction de taux, et a_n est le facteur d'échelle.

Théorème 7. Pour tout ν , $0.1 < \nu < 1$, la suite de variables aléatoires X_n satisfait une propriété de grandes déviations relative à l'intervalle $[t_0, t_1]$, avec un facteur d'échelle n , et un taux de grandes déviations $W(t)$, avec $x_0 = \nu$, $x_1 = 2 - \nu$

$$t_0 = x_0 \frac{B'(x_0)}{B(x_0)} \quad \text{et} \quad t_1 = x_1 \frac{B'(x_1)}{B(x_1)}$$

$$W(t) = - \min_{x \in [x_0, x_1]} \log \left(\frac{B(x)}{x^t} \right), \quad \text{où } B(x) = (2K(x))^{-1}.$$

Pour $\nu = 0.1$, on obtient $t_0 = 0.42$ et $t_1 = 0.735$, et $W(t)$ est représenté au centre sur la figure 5.

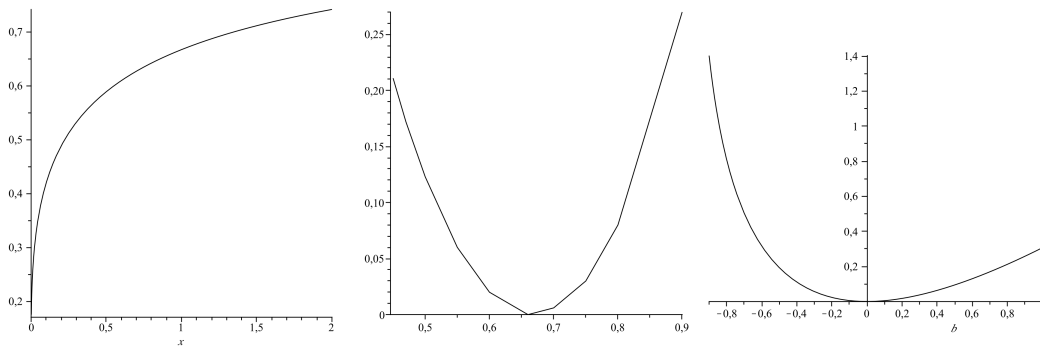


FIGURE 5. À gauche, graphe de $x \mapsto x \frac{B'(x)}{B(x)}$. Au centre, graphe de $W(t)$. À droite, graphe de $b^2 K(b)$ pour $b = -0.9 \dots 0.9$

PREUVE. (Th.7) Pour les grandes déviations, une expression de $p_n(x)$ est nécessaire pour $x \in [x_0, x_1]$, avec $0 < x_0 < 1 < x_1$,

$$p_n(x) = \frac{n! [z^n] X(z, x)}{n! [z^n] X(z, 1)}.$$

En fait, on va prendre $b \in]-0.9, 0.9[$, soit $x_0 = 0.1$, $x_1 = 1.9$. Soit ν , $0.1 < \nu < 1$. Il existe ϵ , $\epsilon < 1.5$, tel que

$$\forall b \in [-1 + \nu, 1 - \nu], \quad 0 \leq b^2 K(b) \leq \epsilon \quad (\text{cf fig.5, à droite}).$$

Le reste s'en suit comme pour la démonstration du lemme 7. On utilise le développement asymptotique de la fonction d'arbre, puis on retrouve $X(x, z)$, et enfin $p_n(x)$.

$$p_n(x) = \left(\frac{1}{2K(x)} \right)^{n+1/2} \left(1 + O\left(\frac{1}{n}\right) \right).$$

De plus, $B(x) = (2K(x))^{-1}$ est analytique sur tout $[\nu, 2 - \nu]$. Il reste alors à appliquer le théorème des grandes déviations ([FS09] p.700). \square

4.3. Généralisation aux k -arbres. Dans cette section, on généralise le travail fait dans la section 4.2, afin de couvrir toute la classe d'urnes. Cela donne les premiers résultats de convergence vers une loi normale avec loi locale limite et principe de grandes déviations pour toute une classe d'urnes, à coefficient tous positifs.

Théorème 8. *Soit X_n le nombre de boules \square dans l'urne U_k des k -arbres après n étapes. Alors,*

– X_n est de moyenne $\mu_n = \mu(n + \frac{1}{k+1}) + O(\frac{1}{n})$ et de variance $\sigma_n^2 = \sigma^2(n + \frac{1}{k+1}) + O(\frac{1}{n})$,

$$\text{et } \frac{X_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1), \text{ avec vitesse de convergence } O\left(\frac{1}{\sqrt{n}}\right).$$

– Soit $p_{n,l} = \mathbb{P}(X_n = l)$. Alors X_n satisfait une loi locale limite gaussienne avec vitesse de convergence $O\left(\frac{1}{\sqrt{n}}\right)$, i.e.

$$\sup_{t \in \mathbb{R}} \left| \sigma\sqrt{n} p_{n, \lfloor \mu n + t\sigma\sqrt{n} \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leq \frac{1}{\sqrt{n}}.$$

– Pour tout ν , $0.1 < \nu < 1$, la suite de variables aléatoires X_n satisfait une propriété de grandes déviations relative à l'intervalle $[t_0, t_1]$, avec un facteur d'échelle n , et un taux de grandes déviations $W(t)$, avec $x_0 = \nu$, $x_1 = 2 - \nu$

$$t_0 = x_0 \frac{B'(x_0)}{B(x_0)} \quad \text{et} \quad t_1 = x_1 \frac{B'(x_1)}{B(x_1)}$$

$$W(t) = - \min_{x \in [x_0, x_1]} \log \left(\frac{B(x)}{x^t} \right).$$

Les constantes non nulles μ et σ^2 sont définies par

$$\mu := \frac{k(k+1)}{k+2} \quad \text{et} \quad \sigma^2 := B''(1) + B'(1) - B'(1)^2 \quad \text{où} \quad B(x) := ((k+1)K_k(x))^{-1},$$

$$\text{et} \quad K_k(x) := \frac{-k}{(x-1)^{k+1}} \ln(x) + \frac{1}{(k-1)x^{k-1}(x-1)^2} + \dots + \frac{k-1}{x(x-1)^k} + \frac{1}{(x-1)^k}.$$

On compte ici les boules \square dans l'urne U_k . Cela correspond, par notre modélisation, au nombre de k -cliques sans descendance dans les k -arbres ordonnés croissants. Ce théorème répond donc à l'étude du paramètre *nombre de k -cliques "feuilles"* (sans descendance).

PREUVE. (Th.8) Intégrons l'équation (23) de la proposition 6,

$$K_k(b) - z = \frac{k}{b^{k+1}} (\ln(X-b) - \ln(X)) + \frac{1}{b^k (X-b)} + \frac{k-1}{b^k X} + \frac{k-2}{2b^{k-1} X^2} + \dots + \frac{1}{(k-1)b^2 X^{k-1}}$$

avec

$$(27) \quad K_k(b) := \frac{-k}{b^{k+1}} \ln(1+b) + \frac{1}{(k-1)b^2(b+1)^{k-1}} + \dots + \frac{k-1}{b^k(b+1)} + \frac{1}{b^k}.$$

Au voisinage de 0, on a

$$K_k(b) = \frac{1}{k+1} - \frac{k}{k+2}b + \frac{k(k+1)}{k+3}b^2 + O(b^3).$$

On réécrit :

$$\frac{b^{k+1}}{k} (K_k(b) - z) = -\ln \left(\frac{X}{X-b} \right) + \frac{b}{k(X-b)} + \sum_{i=1}^{k-1} \frac{b^i (k-i)}{i k X^i}.$$

On effectue le même changement de variable que pour les 1-arbres (prop.7) : $\widehat{X} = \frac{X}{X-b}$, i.e. $X = \frac{-b\widehat{X}}{1-\widehat{X}}$. Cela simplifie grandement l'équation :

$$1 + b^{k+1}(K_k(b) - z) = -k \ln(\widehat{X}) + \widehat{X} + \sum_{i=1}^{k-1} \left(\frac{k}{i} - 1 \right) (1 - \widehat{X}^{-1})^i.$$

En passant à l'exponentielle,

$$(28) \quad \exp(-1 - b^{k+1}(K_k(b) - z)) = \widehat{X}^k e^{-\widehat{X}} \exp\left(\sum_{i=1}^{k-1} \left(1 - \frac{k}{i}\right) (1 - \widehat{X}^{-1})^i\right).$$

Soient $a(b, z) := e^{-1} \exp(-b^{k+1}(K_k(b) - z))$ et

$$f_k(x) := x^k e^{-x} \prod_{i=1}^{k-1} \exp\left(\left(1 - \frac{k}{i}\right) (1 - x^{-1})^i\right).$$

Alors (28) se réécrit $f_k(\widehat{X}) = a(b, z)$. On va inverser, de la même façon que dans la preuve du lemme 7.

Voici quelques propriétés de $f_k(x)$

- (i) $f_1(x) = x e^{-x}$.
- (ii) $f'_k(x) = -(x-1)^k e^{-x} \prod_{i=1}^{k-1} \exp\left(\left(1 - \frac{k}{i}\right) (1 - x^{-1})^i\right)$.
- (iii) $\forall j = 1..k, f_k^{(j)}(1) = 0$, $f_k(1) = e^{-1}$, $f_k^{(k+1)}(1) = -k! e^{-1}$.
- (iv) Au voisinage de $x = 1$, on peut écrire

$$f_k(x) = e^{-1} - \frac{1}{k+1} (x-1)^{k+1} e^{-1} + \frac{(k-1)}{(k+1)(k+2)} (x-1)^{k+2} e^{-1} - \frac{k(k+1)}{2(k+3)} (x-1)^{k+3} e^{-1} + O((x-1)^{k+4}).$$

En inversant au voisinage de $x = 1$, on obtient une expression pour \widehat{X} ,

$$\widehat{X} = 1 - b(k+1)^{\frac{1}{k+1}} (K_k(b) - z)^{\frac{1}{k+1}} + o(K_k(b) - z).$$

En repassant à X , on obtient

$$(29) \quad X(z, b) = ((k+1)(K_k(b) - z))^{-\frac{1}{k+1}} + b O(K_k(b) - z).$$

Cette approximation est valable, pour les mêmes raisons que précédemment : $K_k(b)$ est proche de $(k+1)^{-1}$ pour b voisin de 0, on peut choisir z dans un Δ -domaine contenant 0 et avec la singularité en $K_k(b)$. $\exp(-1 - b^{k+1}(K_k(b) - z)) \sim e^{-1}$ pour b proche de 0, uniformément en z , $|z| < (k+1)^{-1}$. La fonction génératrice de probabilités de X_n est

$$(30) \quad p_n(b) = \frac{n! [z^n] X(z, b)}{n! [z^n] X(z, 0)} = \left(\frac{1}{(k+1)K_k(b)} \right)^{n+\frac{1}{k+1}} \left(1 + O\left(\frac{1}{n}\right) \right).$$

$B_k(x) := ((k+1)K_k(x-1))^{-1}$ est analytique au voisinage $x = 1$. $B'_k(1) = \frac{k(k+1)}{k+2}$ et $B''_k(1) + B'_k(1) - B'_k(1)^2 \neq 0$ se calcule pour chaque valeur de k .

- Les mêmes résultats sont donc obtenus, en appliquant les théorèmes IX.8 IX.14 et IX.15 de [FS09] :
- convergence vers une loi gaussienne avec vitesse de convergence,
 - loi locale limite,
 - principe de grandes déviations.

□

Conclusion

Ce rapport illustre divers problèmes liés aux modèles d'urnes que nous avons traités par la combinatoire analytique. Différents types d'urnes sont évoqués ; d'une part, des urnes bicolores additives ; d'autre part, des urnes en grande dimension (supérieure ou égale à 3). Dans ces deux cas, peu de résultats avaient été obtenus par le traitement analytique.

L'étude de l'urne 3×3 liée aux branches d'arbre, ainsi que sa généralisation aux urnes de dimensions supérieures ($k + 2$), ouvrent des perspectives dans la classification des urnes de grande dimension. On peut se demander si une structure se dégage lorsque l'on considère des urnes 3×3 comportant une boule "Déesse Mère" (c'est-à-dire les urnes dont la première colonne est nulle).

L'étude de la classe générale d'urnes bicolores additives "algébriques" soulève des problèmes de confluence de singularités, donnant lieu à des discontinuités de régime asymptotique (cas $x = 1$ et $x \neq 1$). Voici une étape dans la compréhension des modèles généraux d'urnes additives.

Enfin, on a exhibé une première classe (infinie) de modèles d'urnes additives bicolores pour lesquelles des résultats probabilistes forts sont obtenus : convergence en loi vers une loi gaussienne avec vitesse de convergence, loi locale limite et principe de grandes déviations. Cette classe est obtenue par la modélisation des k -arbres ordonnés croissants. Cela nous amène à une piste de réflexion qui est celle de la modélisation des arbres par des urnes. Dans quels cas le modèle probabiliste de croissance sur les arbres correspond-il à l'énumération des histoires d'une urne ?

Ces perspectives viennent directement de notre étude, mais d'autres questions subsistent encore. La classification des urnes bicolores additives, l'étude en dimension supérieure, mais aussi l'étude d'urnes non équilibrées sont des problèmes largement ouverts.

RÉFÉRENCES

- [BFSS01] Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria. Random maps, coalescing saddles, singularity analysis, and Airy phenomena. *Random Structures & Algorithms* **19**, n° 3/4, pages 194–246, 2001.
- [BP85] A. Bagchi and A.K. Pal. Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM Journal on Algebraic and Discrete Methods* **6**, n° 3, pages 394–405, 1985.
- [CPS] Brigitte Chauvin, Nicolas Pouyanne, and Réda Sahnoun. Limit distributions for large Pólya urns. À paraître dans *Annals of Applied Probability*.
- [Dar10] Alexis Darrasse. *Structures arborescentes complexes : analyse combinatoire, génération aléatoire et applications*. PhD thesis, Université Pierre et Marie Curie, 2010.
- [DHBS10] Alexis Darrasse, Hsien-Kuei Hwang, Olivier Bodini, and Michèle Soria. The connectivity-profile of random increasing k -trees. *Workshop on Analytic Algorithmics and Combinatorics (ANALCO10)*, 2010.
- [Dur04] Marianne Durand. *Combinatoire analytique et algorithme des ensembles de données*. PhD thesis, École Polytechnique, 2004.
- [FDP06] Philippe Flajolet, Philippe Dumas, and Vincent Puyhaubert. Some exactly solvable models of urn process theory. *Discrete Mathematics & Theoretical Computer Science*, pages 59–118, 2006.
- [FGP05] Philippe Flajolet, Joaquim Gabarro, and Helmut Pekari. Analytic urns. *Annals of Probability* **33**, n° 3, pages 1200–1233, 2005.
- [FS09] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

- [Jan04] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Processes and their Applications* **110**, pages 177–245, 2004.
- [Jan06] Svante Janson. Limit theorems for triangular urn schemes. *Probability Theory Related Fields* **134**, n° 3, pages 417–452, 2006.
- [Jan08] Svante Janson. Plane recursive trees, stirling permutations and an urn model. *DMTCS Proceedings*, 2008.
- [JK77] Norman L. Johnson and Samuel Kotz. *Urn Models and Their Application*. John Wiley & Sons, 1977.
- [Mah03] Hosam Mahmoud. Pólya Urn Models and Connections to Random Trees : A Review. *JIRSS* **2**, n° 1, pages 53–114, 2003.
- [Mah08] Hosam Mahmoud. *Pólya Urn Models*. Chapman & Hall/CRC Press, 2008.
- [PS10] Alois Panholzer and Georg Seitz. Ordered increasing k-trees : Introduction and analysis of a preferential attachment network model. *DMTCS Proceedings of AofA'10*, 2010.
- [Smy96] R.T. Smythe. Central limit theorems for urn models. *Stochastic Processes and their Applications* **65**, pages 115–137, 1996.