Hachage, arbres, chemins & graphes

Philippe Chassaing[1] et Philippe Flajolet[2]

Mathématiques discrètes et continues se rencontrent et se complètent volontiers harmonieusement. C'est cette thèse que nous voudrions illustrer en discutant un problème classique aux ramifications nombreuses—l'analyse du hachage avec essais linéaires. L'exemple est issu de l'analyse d'algorithmes, domaine fondé par Knuth et qui se situe lui-même « à cheval » entre l'informatique, l'analyse combinatoire, et la théorie des probabilités. Lors de son traitement se croisent au fil du temps des approches très diverses, et l'on rencontrera des questions posées par Ramanujan à Hardy en 1913, un travail d'été de Knuth datant de 1962 et qui est à l'origine de l'analyse d'algorithmes en informatique, des recherches en analyse combinatoire du statisticien Kreweras, diverses rencontres avec les modèles de graphes aléatoires au sens d'Erdös et Rényi, un peu d'analyse complexe et d'analyse asymptotique, des arbres qu'on peut voir comme issus de processus de Galton-Watson particuliers, et, pour finir, un peu de processus, dont l'ineffable mouvement Brownien ! Tout ceci contribuant *in fine* à une compréhension très précise d'un modèle simple d'aléa discret.

1. Hachage

Nous ferons commencer l'histoire par Knuth en 1962 ; Knuth a alors 24 ans et hésite entre une carrière en mathématiques discrètes et une passion pour l'informatique très concrète. L'un des tous premiers problèmes quantitatifs de la naissante science informatique consiste à comprendre comment se comporte une certaine méthode d'accès à des données qui apparaît comme présentant au vu des simulations de très bonnes caractéristiques de complexité. L'École de Feller est aussi « sur le coup » (dixit Knuth). Knuth apportera très vite une solution[1] à ces questions (à partir de 1962) et, comme il le dit lui-même, c'est ce succès scientifique initial qui déterminera très largement la suite de sa carrière. Pour les informaticiens, cette étape marque les débuts de l'analyse mathématique d'algorithmes—entendre par là l'analyse des performances et de la complexité des algorithmes « discrets » de l'informatique.

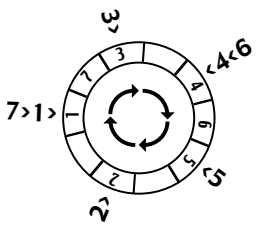
Voici donc le problème. Imaginez que vous voulez ranger dans un fichier les noms, numéros de téléphones, etc, de vos amis. Vous disposez de m tiroirs et de n amis ($n \leq m$). Dans un monde idéal, chaque ami verrait ses donnés rangées directement dans un tiroir qui lui est propre et dont le numéro est calculable (aisément) à partir du nom au moyen d'une fonction dite « de hachage ». L'utilisation du fichier est en ce cas simple et rapide : étant donné le nom x, on

[1]P. Chassaing : Institut Elie Cartan, Université Henri Poincaré B.P. 239, 54506 Vandoeuvre les Nancy Cedex
Philippe.Chassaing@iecn.u-nancy.fr
[2]P. Flajolet : Projet Algorithmes, INRIA Rocquencourt, F-78153 Le Chesnay
Philippe.Flajolet@inria.fr

[1]Le manuscrit est disponible en http ://algo.inria.fr/AofA/Research/11-97.html.

calcule la valeur numérique « hachée » $h(x)$, puis, dans la case $h(x)$, on trouve au premier coup les données désirées. Si n est bien plus petit que $\sqrt{m}$, le paradoxe des anniversaires (une salle de 23 personnes a de bonnes chances de contenir deux personnes ayant le même anniversaire) implique qu'avec bonne probabilité le rangement soit possible sans collisions. Une telle solution fondée sur un rangement « idéal » est cependant peu satisfaisante car il faudrait en gros prévoir un nombre de tiroirs qui soit de l'ordre du *carré* de votre nombre d'amis !

En termes informatiques, on dispose d'une table $T[1\,..\,m]$ dont chaque entrée peut contenir une donnée et des informations auxiliaires. Les données sont issues d'un univers $\mathcal{U}$ (par exemple les chaines alphabétiques de longueur au plus 20). On dispose d'une fonction h dite « fonction de hachage » : celle-ci[2] envoye le domaine $\mathcal{U}$, en général très grand (ici de cardinalité $2 \cdot 10^{29}$), sur un intervalle d'entiers $[1\,..\,m]$ beaucoup plus petit (le nombre m de cases ou tiroirs est en général entre 10^2 et 10^6). On range alors chaque donnée x à l'addresse $h(x)$ dans la table. Pour régler les collisions qui sont statistiquement inévitables, lorsque la case de $h(x)$ est déjà prise, on choisit de ranger x à la première case disponible parmi $1+h(x)$, $2+h(x)$, etc. De plus, dans la variante circulaire du procédé, on recommencera à partir de la première case (de numéro 1) suite à un échec sur la case m. Ci-dessous, un exemple où $(m, n) = (10, 7)$.



Il est en pratique très raisonnable de supposer que les valeurs hachées sont uniformément distribuées sur l'intervalle $[1, m]$. L'hypothèse est vérifiée pour peu que la fonction de hachage « mélange bien » les bits des données (voir par ex. Lum et al. pour des vérifications effectives sur bases de données industrielles). Knuth [18] décrit le problème en ces termes :

> « A certain one-way street has m parking spaces in a row numbered 1 to m. A man and his dozing wife drive by, and suddenly, she wakes up and orders him to park immediately. He dutifully parks at the first available space [...]. »

La question est alors de déterminer la distance attendue entre l'endroit où l'on parque et l'endroit où l'on souhaitait s'arrêter initialement. Naturellement, on s'attend à ce que parquer dans une rue quasi-déserte ($n = o(m)$) soit facile, mais que la situation se dégrade lorsque la rue se remplit ($n \approx m$). Mais comment ? On aura reconnu là une version discrète du problème de parking [30] lancé par le probabiliste Alfréd Rényi en 1958. Notons que le problème s'exprime aussi en termes de placement de n boules dans m urnes, chacune de capacité $b = 1$, mais avec la contrainte non classique de débordement à droite des urnes.

[2]Par exemple, $h(x)$ interprète x comme un nombre en base 26, puis le réduit modulo m.

2. L'analyse en moyenne

Une configuration de hachage est donc déterminée par une suite d'élements de $\{1,\dots,m\}$ ayant longueur n, et, d'après la discussion précédente, l'hypothèse d'équiprobabilité des m^n suites est bien justifiée. Le paramètre fondamental d'une configuration est le « déplacement total » $D_{m,n}$ défini comme la somme (sur x) des distances (circulaires) entre l'endroit où x est placé et la valeur de $h(x)$. Cette quantité est donc une variable aléatoire qui décrit précisément le coût cumulatif de construction de la table. Sa variation se situe entre 0 et $0+1+\cdots+(n-1)=\binom{n}{2}$.

Le premier résultat de Knuth, obtenu il y a près de quarante ans, est le calcul de l'espérance du déplacement total sous forme exacte :

$$E\left[D_{m,n}\right] = n + \frac{n}{2}\left(\frac{n-1}{m} + \frac{(n-1)(n-2)}{m^2} + \dots\right). \tag{1}$$

On distingue classiquement le cas *épars* ($n\to+\infty,\ n/m=\alpha,\ 0<\alpha<1$) et le cas *presque plein* ($n\to+\infty,\ n=m-1$). Ceux-ci donnent alors lieu aux asymptotiques suivantes :

$$E\left[D_{m,n}\right] \sim \frac{n}{2}\left(1+\frac{1}{1-\alpha}\right), \quad E\left[D_{n+1,n}\right] \sim \frac{n}{2}\sqrt{\frac{\pi n}{2}}. \tag{2}$$

En substance (penser à $D_{m,n}-D_{m,n-1}$), ceci nous dit qu'il suffit en moyenne de $O(1)$ essais pour trouver une case vide tant que le taux de remplissage α est bien séparé de 1, $\alpha\le\alpha_0<1$, tandis que la situation se dégrade dès lors que α approche 1. Ce, au point que, pour une table pleine, le coût total de construction devient surlinéaire, de l'ordre de $n\sqrt{n}$.

L'analyse de Knuth pour (1) commence par un « lemme cyclique » : il y a $(m-n)m^{n-1}$ suites de hachage qui laissent la dernière case vide (preuve : grouper les suites par équivalence vis à vis des rotations). Le problème ainsi linéarisé fait jouer un rôle particulier aux « fonctions de parking » définies par le fait que la dernière case, m, reste vide, ainsi qu'aux allocations « presque pleines » (fonctions de parking satisfaisant de plus à $n=m-1$). Elle se poursuit par une récurrence sur n, en prenant en compte le dernier élément ajouté. Il est curieux que la simplification des formules donnant l'espérance (1) passe par une version du théorème binomial due à Abel, soit,

$$(x+y)^r = \sum_k \binom{r}{k} x(x-k)^{k-1}(y+k)^{r-k}.$$

Quant au problème asymptotique associé, la forme de l'espérance,

$$E[D_{n,n}] = \frac{n}{2}(Q(n)-1) \quad \text{où} \quad Q(n) := 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots,$$

devait s'avérer avoir une digne histoire. De fait, la quantité $Q(n)$ qui y figure apparaît déjà de manière déguisée dans la première lettre de Ramanujan à Hardy en date du 16 janvier 1913, sous forme de l'assertion suivante :

« $\frac{1}{2}e^n = 1+\frac{n}{1}+\frac{n^2}{2!}+\cdots+\frac{n^n}{n!}\theta$, with $\theta=\frac{1}{3}+\frac{4}{135+k}$, where k lies between $\frac{4}{45}$ and $\frac{2}{21}$. »

(Il s'agit donc en substance d'estimer très finement la probabilité qu'une variable de Poisson de grand paramètre n soit inférieure à sa moyenne.) Cette « conjecture » de Ramanujan, qualifiée par Berndt de « an ultimately famous problem », sera étudiée par Watson et Szegő en 1928–1929 avant de sombrer dans l'oubli et d'être finalement réglée en 1995 par l'asymptotique complexe élémentaire. Heureusement, une forme très faible de l'assertion de Ramanujan suffit à établir (2), de sorte qu'une preuve d'analyse réelle très élémentaire peut être donnée par Knuth, ce via la méthode de Laplace et la formule sommatoire d'Euler-Maclaurin.

3. Moments et distribution

Bien que la question, émanant de Knuth, de déterminer la variance de $D_{m,n}$ ait circulé dans la communauté d'analyse d'algorithmes, le problème ne reçoit que peu d'écho pendant deux décennies. En 1998, Flajolet, Poblete et Viola [13] donneront une solution en fournissant en fait des caractérisations complètes de la loi limite de $D_{m,n}$ Le résultat asymptotique en est le suivant : *dans le cas épars, la loi du déplacement total est asymptotiquement normale ; dans le cas (presque) plein, la loi limite coïncide avec la distribution de l'aire sous l'excursion brownienne.* Cette section discute l'angle « analytique » sous lequel ce résultat a d'abord été obtenu.

Le cas le plus intéressant mathématiquement est bien sûr celui d'une table pleine[3] et les allocations presque-pleines (seule la case finale est libre, $n = m-1$) jouent un rôle fondamental. En gros, la classe de ces allocations vérifie une décomposition quadratique : une allocation se décompose alors selon le dernier élement inséré n en une paire d'allocations ; le premier élément de cette paire a de plus une case distinguée, celle où n aurait voulu se placer. (Cette décomposition quadratique récursive évoque une structure d'arbre binaire enrichie de diverses informations.)

Notons $F_{n,k}$ le *nombre* d'allocations de $\mathcal{F}$ de n éléments ayant déplacement total égal à k ; on introduit, comme toujours en analyse combinatoire [10], les séries génératrices (de dénombrement) correspondantes :

$$F_n(q) := \sum_k F_{n,k} q^k, \qquad F(z,q) := \sum_{n,k} F_{n,k} q^k \frac{z^n}{n!}.$$

La décomposition quadratique correspond alors à une récurrence non linéaire

$$F_n(q) = \sum_{k=0}^{n-1} \binom{n-1}{k} F_k(q)(1+q+\cdots+q^k) F_{n-1-k}(q). \tag{3}$$

qui se traduit de manière plus synthétique sur la série génératrice double en

$$\frac{\partial}{\partial z} F(z,q) = (\Delta F(z,q)) \cdot F(z,q), \qquad \text{où} \quad \Delta h(z) := \frac{h(z) - q h(qz)}{1-q}. \tag{4}$$

[3]Le cas épars est plus simple techniquement : une table se décompose en effet en « ilots » qui sont des tables presque pleines ; les ilots étant bien séparés et de taille typique $O(1)$, un raisonnement *grosso modo* de type théorème central limite s'applique.

Cette *équation fonctionnelle* fondamentale est donc non linéaire, mixte aux dérivées partielles et aux différences finies. Les polynômes $F_n(q)$ ne peuvent être totalement élémentaires puisque le degré de $F_n(q)$ vaut $n(n-1)/2$.

L'article [13] poursuit en développant un petit calcul opérationnel mettant en jeu la dérivée partielle ∂_z, l'opérateur de différence, Δ, ainsi que ∂_q et l'opérateur U de spécialisation $q = 1$. Il en découle la possibilité de calculer par « pompage » les moments successifs via les séries associées (voir (9) ci-dessous). En particulier, on obtient l'écart type sous forme exacte, ce qui livre en passant l'asymptotique

$$\sigma(D_{n+1,n}) \sim \sqrt{\frac{10-3\pi}{24}} n^{3/2}. \tag{5}$$

Comme cet écart type est du même ordre que la moyenne, la distribution discrète est étale. Généralement on trouve, par « l'analyse de singularité » [12, 26] des fonctions génératrices associées (voir (9) ci-dessous pour l'aspect général de ces fonctions) le résultat suivant : *Le moment d'ordre k du déplacement total dans les tables de hachage pleines est asymptotique à $\mu_k (n/2)^{3k/2}$, où les constantes fondamentales μ_k se relient aux coefficients du développement à l'infini de la dérivée logarithmique de la fonction d'Airy :*

$$\begin{aligned} \mu_k &:= \frac{-\Gamma(-1/2)}{\Gamma((3r-1)/2))}\Omega_k \\ \Omega_k &: \frac{\mathrm{Ai}'(z)}{\mathrm{Ai}(z)} \underset{z\to+\infty}{\sim} \sum_{r=0}^{\infty} \frac{(-1)^r}{2^r \cdot r!}\Omega_r z^{-(3r-1)/2} \\ \mathrm{Ai}(z) &:= \frac{1}{\pi}\int_0^\infty \cos\left(\frac{1}{3}t^3 + zt\right) dt. \end{aligned} \tag{6}$$

Ceci s'obtient à partir de la décomposition quadratique : les μ_k vérifient eux-même une récurrence quadratique, image simplifiée de la décomposition combinatoire (4) ; un nouveau calcul de série génératrice livre alors (6).

On peut dès lors effectuer la jonction avec l'aire sous l'excursion Brownienne normalisée puisque les moments en ont été calculés par Louchard en 1984 [24] et déjà reliés par lui à la fonction d'Airy. Par convergence des moments de la loi discrète, on est ainsi en mesure de conclure : *la loi de la variable $n^{-3/2} D_{n+1,n}$ converge vers la loi de l'aire sous l'excursion Brownienne.*

La densité du déplacement total est représentée à la figure 1. Le calcul de cette densité $\omega(x)$ est d'ailleurs permis par un calcul de Takács [35] (relatif à l'aire de l'excursion et publié en 1991) qui procède par double inversion de Laplace à partir de (6) et fournit :

$$\begin{aligned} \omega(x) &= \frac{2\sqrt{6}}{x^2}\sum_{k=1}^{\infty} e^{-v_k} v_k^{3/2} U\left(-\frac{5}{6}, \frac{4}{3}; v_k\right), \quad v_k = \frac{2\alpha_k^3}{27x^2}, \\ U(a,b;z) &:= \frac{1}{\Gamma(a)}\int_0^\infty e^{-zt} t^{a-1}(1+t)^{b-a-1}\, dt. \end{aligned}$$

Dans ces équations, $-\alpha_k$ parcourt l'ensemble des zéros de la fonction d'Airy $\mathrm{Ai}(z)$, tandis que $U(z)$ est une fonction hypergéométrique confluente [1] ; voir aussi [11] pour divers développements analytiques liés à ces formules.

Avec un synchronisme étonnant, Knuth qui avait déjà découvert (4) par de toutes autres voies en tire « directement », via un calcul simple mais astucieux, la solution explicite

$$F(z,1+q) = q\frac{\partial}{\partial z}\log\left(1+\sum_{n=1}^{\infty}(1+q)^{n(n-1)/2}\frac{z^n q^{-n}}{n!}\right), \tag{7}$$

publiée dans [19]. Tout spécialiste d'analyse combinatoire reconnaît alors à vue que $F(z,1+q)$ est une série génératrice double des *graphes connexes* selon le nombre de sommets et selon l'excès du nombre d'arêtes au nombre de sommets. Or E. M. Wright [36] avait dès 1977 déterminé la forme des séries correspondant à un excès donné. Il découle ainsi de la connection Knuth–Wright une caractérisation des séries génératices des moments factoriels du déplacement total. Un rôle central y est joué par la « fonction de Cayley »,

$$T(z) := \sum_{n\geq 1} n^{n-1}\frac{z^n}{n!} \quad \text{vérifiant} \quad T(z) = ze^{T(z)}, \tag{8}$$

laquelle énumère les arbres étiquetés enracinés à n sommets (Cayley, 1889). On accède ainsi aux moments du déplacement total par les relations [13, 19]

$$\frac{z}{k!}\left(\frac{\partial^k}{\partial q^k}F(z,q)\right)_{q=1} = \frac{A_{3k}(T(z))}{(1-T(z))^{3k-1}}, \tag{9}$$

pour une certaine famille de polynômes $\{A_{3k}\}$ fort bien étudiés par ailleurs par Knuth *et al.* dans « l'article géant sur la composante géante » concernant les graphes aléatoires [17]. (Les valeurs $A_{3k}(1)$ sont étroitement liées aux μ_k, Ω_k de (6).) L'estimation (5) en résulte et, bien que Knuth ne pousse pas l'observation dans cette direction, la loi limite de type Airy pour le déplacement sort encore facilement de ses calculs.

En conclusion, à ce stade de la discussion, deux approches analytiques, somme toute voisines, s'appliquent. L'une (Knuth) est exacte et révèle par le calcul explicite l'existence d'une *relation exacte* entre hachage et connexité des graphes ; l'autre (Flajolet-Poblete-Viola) est asymptotique et révèle au niveau des lois de probabilités une *relation limite* entre hachage et aire sous l'excursion Brownienne. En résumé : *« aire sous les excursions, connexité des graphes, et*

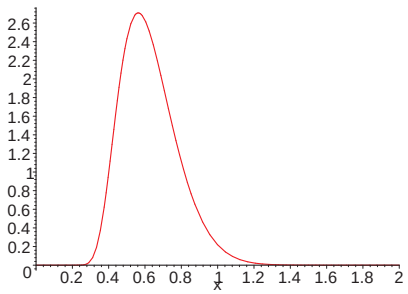


FIG. 1. La densité du déplacement total.

déplacement dans les tables de hachage sont étroitements liés par une structure analytique commune. »

On verra dans la suite de ce texte que ces relations s'insèrent dans un riche faisceau de transformations combinatoires. L'approfondissement de ces relations aurait pu d'ailleurs constituer une troisième voie d'attaque à l'analyse du hachage. En tout état de cause, les travaux combinatoires sont précieux. Ils permettent d'affiner notre connaissance du hachage et sont, comme nous allons le voir, exploités dans divers travaux probabilistes récents sur le sujet.

4. Combinatoire des arbres, graphes, et parkings

Après l'examen de ce que donnent les voies (voix ?) naturelles de l'analyse, nous nous proposons de revenir à la « troisième voie », celle de la combinatoire élémentaire. La démarche combinatoire éclaire de fait la relation entre les récentes constructions apparemment très differentes du coalescent additif, par Aldous et Pitman d'une part, par Bertoin d'autre part. Les articles sur le sujet constituant un puzzle complexe (des problèmes voisins sont attaqués par plusieurs communautés, probabiliste, combinatoire, ou informaticienne), on doit abandonner l'idée d'une perspective historique systématique. Ce qui suit est donc essentiellement une réorganisation d'idées de plusieurs auteurs, notamment Kreweras, Gessel, Spencer.

En 1980, le statisticien-combinatoricien Kreweras avait déjà découvert quelques propriétés surprenantes du polynôme $F_n(q)$ tel que défini par (3). Partant de la relation de récurrence, Kreweras remarque dans [22] que les polynomes F_n (qui sont unitaires et de degré $\binom{n}{2}$) satisfont aux relations et connexions combinatoires suivantes :
(10)

$$\begin{aligned}
F_n(-1) &= n![z^n](\tan z + \sec z) && \text{permutations alternantes de } [1,n]\\
F_n(0) &= n! && \text{permutations de } [1,n]\\
F_n(1) &= (n+1)^{n-1} && \text{arbres étiquetés à } n+1 \text{ sommets}\\
F_n(2) &= n![z^n]\log\left(1+\sum_{k=1}^{\infty} 2^{\binom{k}{2}}\frac{z^k}{k!}\right) && \text{graphes connexes à } n+1 \text{ sommets.}
\end{aligned}$$

(Une permutation alternante $\sigma = \sigma_1, \ldots, \sigma_n$ est telle que $\sigma_1 < \sigma_2 > \sigma_3 < \cdots$.)
L'article de Kreweras ainsi qu'un article contemporain de Gessel et Wang [16] explorent la nature de ces correspondances arbres–graphes et même la relation à des objets qui sont (implicitement) des fonctions de parking. Un rôle essentiel est joué par la notion de parcours de graphe, notion qui fait surface dans plusieurs travaux ultérieurs dont ceux de Spencer.

Parcours de graphe.

Considérons un graphe étiqueté à $n+1$ sommets numérotés par $\{0, 1, 2, \ldots, n\}$. Le sommet 0 est vu comme une source d'exploration du graphe. Un algorithme de parcours de graphe est alors obtenu comme suit.

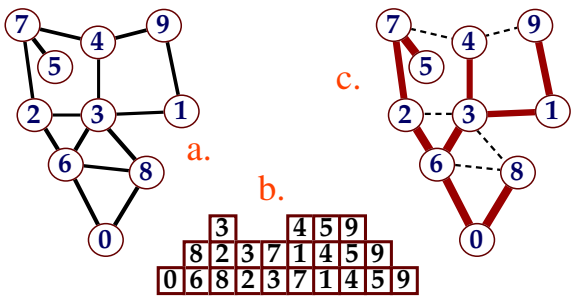


FIG. 2. Un graphe γ (a), les états $(\mathcal{F}_t)_{0\le t\le 9}$ de la file (b) et l'arbre couvrant (c) associés au parcours en largeur de γ.

ALGORITHME DU PARCOURS. On part du sommet 0. À tout instant, un sommet peut être dans deux états : l'état « inconnu » ou l'état « exploré ». Initialement, tout sommet (sauf la source) est « inconnu ». L'algorithme opère avec une file d'attente qui à chaque instant contient une liste de sommets déjà visités. Soit $\mathcal{F}_t$ l'état de la file à l'instant t. Initialement $t=0$ et $\mathcal{F}_0 := \{0\}$. La « boucle » principale de l'algorithme consiste alors à répéter jusqu'à épuisement l'operation suivante :

> Au temps t ($t=1,2,\ldots$), choisir un sommet $s_t \in \mathcal{F}_{t-1}$; l'enlever de $\mathcal{F}_{t-1}$. Soit A_t l'ensemble des voisins de s_t (selon l'adjacence du graphe) qui sont encore « inconnus » ; on remplace alors dans $\mathcal{F}_{t-1}$ l'élément s_t par les éléments de A_t, de sorte que
>
> $$\mathcal{F}_t = (\mathcal{F}_{t-1} \setminus \{s_t\}) \cup A_t.$$
>
> Au moment où les nouveaux sommets sont insérés dans $\mathcal{F}_t$, leur état passe du statut « inconnu » au statut « exploré ».

Clairement, ce schéma permet de parcourir tous les sommets d'un graphe connexe en leur rendant visite une fois et une seule. De fait, ce schéma associe à un graphe connexe γ un *arbre couvrant* τ, l'arbre dont les arêtes relient s_t aux élements de A_t, les arêtes de s_t vers les autres voisins étant inutilisées.

L'algorithme est complètement spécifié dès qu'une politique $\mathbf{\Pi}$ fixe le principe de selection de s_t ainsi que l'ordre d'insertion de ses successeurs $s' \in A_t$. Les principes « LIFO » (*Last-In-First-Out* : on choisit le $s \in \mathcal{F}$ le plus récent, ce qui se gère par une pile) ou « FIFO » (*First-In-First-Out* : on sert le plus ancien dans la file) sont par exemple des politiques correspondant aux parcours appelés « en profondeur d'abord » et « en largeur d'abord ». Le principe par « priorité » (choix du plus petit ou plus grand numéro d'abord) est une autre politique particulièrement intéressante du point de vue combinatoire.

Graphes, et parking.

Nous allons maintenant tirer quelques conséquences combinatoires du schéma de parcours. Ce qui suit est largement tiré d'une courte mais inspirante note de Spencer.

Principe d'équivalence graphes–parkings. *Soit $C_{n,k}$ le nombre de graphes connexes à n sommets et $n+k-1$ arêtes (on parle aussi de graphes d'excès k et les $C_{n,k}$ sont encore appelés « nombres de Wright »). Soit $D_{n+1,n}$ la variable aléatoire représentant le déplacement total dans les fonctions de parking à n voitures. Les moments binomiaux de $D_{n+1,n}$ et les nombres de Wright $C_{n,k}$ sont reliés par la relation*

$$\frac{C_{n,k}}{C_{n,0}} = E\left[\binom{D_{n,n-1}}{k}\right],$$

où l'on a par ailleurs que $C_{n,0} = n^{n-2}$.

Démonstration. La preuve se fonde sur le fait que graphes et parkings conduisent à des statistiques sur deux types de partitions ordonnées, lesquelles s'avèrent, *in fine*, isomorphes.

(*i*) *Des graphes aux partitions.* Le déroulement de l'algorithme de parcours d'un graphe γ à $n+1$ sommets $\{0,1,2,\ldots,n\}$ est décrit par son arbre couvrant τ, ou encore par la partition ordonnée $(A_k(\tau))_{1\le k\le n}$ de l'ensemble $\{1,2,\ldots,n\}$, induite par le parcours en largeur. Pour une telle partition, notons $x_k(\tau)$ le cardinal du $k^{\text{ème}}$ morceau, et posons

$$y_k(\tau) = x_1(\tau) + x_2(\tau) + \cdots + x_k(\tau) - k + 1 \ ; \tag{11}$$

$y_k(\tau)$ représente donc la longueur de la file d'attente après le $k^{\text{ème}}$ pas de l'algorithme de parcours, le graphe de $k \longrightarrow y_k(\tau)$ apparaissant par exemple au (b) de la Figure 2.

Un graphe connexe d'arbre couvrant τ a pour arêtes toutes les arêtes de τ, plus, éventuellement, certaines arêtes prises dans l'ensemble, à $(y_1(\tau)-1) + (y_2(\tau)-1) + \cdots + (y_n(\tau)-1)$ éléments, des arêtes qui joignent un sommet en tête de la file d'attente et un autre sommet de la file d'attente. Ainsi, sur l'exemple de la Figure 2, ont τ pour arbre couvrant les graphes contenant τ et dont les arêtes en excès sont prises dans l'ensemble à 12 éléments $\{(6,8),(8,2),(8,3),(2,3),(3,7),(7,1),(7,4),(1,4),(1,5),(4,5),(4,9),(5,9)\}$, correspondant, sur la Figure 2(b), aux 12 cases des deux étages supérieurs. Si par contre on rajoute l'arête $(6,5)$ à γ, on obtient un graphe γ' dont l'arbre couvrant n'est plus τ. Parmi les graphes connexes d'excès k , il y en a donc

$$\binom{y_1(\tau) + y_2(\tau) + \cdots + y_n(\tau) - n}{k}$$

dont l'arbre couvrant est τ. Ainsi

$$C_{n+1,k} \quad = \quad \sum_{\tau} \binom{y_1(\tau) + y_2(\tau) + \cdots + y_n(\tau) - n}{k}, \tag{12}$$

où τ parcourt l'ensemble des arbres à $n+1$ sommets.

(*ii*) *Des fonctions de parking aux partitions.* Une fonction parking f de n voitures sur $n+1$ places est également décrite par une partition ordonnée de l'ensemble $\{1,2,\ldots,n\}$ en n morceaux, notée $(B_k(f))_{1\le k\le n}$, avec

$$B_k(f) = f^{-1}(k) \ ;$$

$B_k(f)$ est l'ensemble des voitures, de cardinal noté $x_k(f)$, dont la place k est la première tentative. Définissons $y_k(f)$ de manière analogue à (11) :

$$y_k(f) = x_1(f) + x_2(f) + \cdots + x_k(f) - k + 1; \tag{13}$$

$y_k(f)$ représente alors le nombre de voitures ayant tenté, avec ou sans succès, de se garer sur la $k^{\text{ème}}$ place. Le déplacement total coïncidant avec le nombre total de tentatives infructueuses, on a

$$y_1(f) + y_2(f) + \cdots + y_n(f) - n = D_{n,n+1}(f),$$

et, naturellement, le moment factoriel du déplacement total s'exprime par

$$E\left[\binom{D_{n,n+1}}{k}\right] = \frac{1}{(n+1)^{n-1}} \sum_f \binom{y_1(f) + y_2(f) + \cdots + y_n(f) - n}{k}, \tag{14}$$

où f parcourt l'ensemble des fonctions de parking. (Rappel : on a vu à la section 2 que $(n+1)^{n-1}$ dénombre les fonctions de parking.)

(*iii*) *L'équivalence.* Le point clé est que les ensembles de partitions "admissibles", d'une part l'ensemble des partitions $(A_k)_{1\le k\le n}$ issues du parcours d'un arbre (ou plus généralement du parcours d'un graphe connexe) et d'autre part l'ensemble des partitions $(B_k)_{1\le k\le n}$ issues d'une fonction parking, sont confondus. De fait, la condition pour qu'une partition soit admissible dans l'un ou l'autre des sens du terme, est que

$$y_k \ge 1, \qquad 1 \le k \le n. \tag{15}$$

En effet, pour un parcours d'arbre ou de graphe, y_k représente la longueur de la file d'attente après le $k^{\text{ème}}$ pas, et l'inégalité (15) traduit la contrainte de connexité sur le graphe ou l'arbre ; pour une fonction parking, l'inégalité (15) traduit que la seule place vide est la place $n+1$ (ou 0, indifféremment).

Ainsi chacun des trois parcours cités plus haut (le parcours en largeur en particulier) induit une bijection arbres $\leftrightarrow$ parking, en associant à l'arbre τ la fonction de parking f_τ définie par

$$B_k(f_\tau) = A_k(\tau).$$

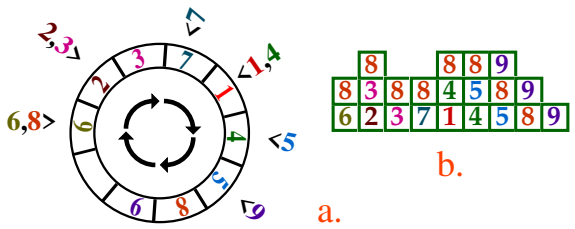


FIG. 3. Le parking induit par la partition ordonnée $(\{6,8\},\{2,3\},\varnothing,\{7\},\{1,4\},\{5\},\{9\},\varnothing,\varnothing)$, apparaissant déja Figure 2 pour l'arbre γ (a), et les piles des voitures ayant essayé de se garer sur les places 1, 2, etc .., dans l'ordre chronologique (b).

Les relations (12) et (14) induisent alors l'identité

$$C_{n+1,k} = C_{n+1,0} E\left[\binom{D_{n,n+1}}{k}\right],$$

laquelle exprime précisément le principe d'équivalence graphes-parking. □

Cette explication bijective est due à Spencer [33], quoiqu'il ne soit pas fait mention explicite du parking dans son article. Des bijections arbres-parking, voisines sont connues et étudiées depuis des travaux de Pollack, Schützenberger, Foata et Riordan, Françon dans les années 1970. Si l'on part directement de la définition de $F_n(q)$ comme polynôme énumérateur du déplacement dans les fonctions de parking de taille n, la série génératrice de probabilité du déplacement vaut $F_n(q)/F_n(1)$ et son moment binomial vaut

$$E\left[\binom{D_{n,n+1}}{k}\right] = \frac{1}{k!}\frac{F_n^{(k)}(1)}{F_n(1)}$$

Le développement de Taylor

$$F_n(1+q) = \sum_{k\geq 0} \frac{1}{k!} F^{(k)}(1) q^k$$

permet alors de transcrire la relation (12) comme suit : *Le polynôme énumérateur des déplacements de parking, F_n, est tel que $F_n(1+q)$ énumère les graphes connexes de taille $n+1$ selon l'excès.* On retrouve de la sorte par la combinatoire le résultat de Knuth (7) et l'on a au passage largement généralisé l'observation de Kreweras (10) concernant $F_n(1)$ et $F_n(2)$.

Arbres et parking.

Les arbres sont manifestement présents dans toute cette discussion. D'abord, le nombre d'arbres étiquetés de taille n, valant n^{n-2}, est présent dans le dénombrement des allocations presque-pleines ($(n+1)^{n-1}$) et dans l'analyse en moyenne de Knuth à la section 2. Ensuite, la série génératrice de ces nombres, la fonction $T(z)$ de Cayley de (8) est à la base de l'analyse de variance, et des moments d'ordre supérieur, du déplacement total (section 4 et équations (8), (9)). Enfin, le ré-examen combinatoire du problème de parking fait jouer, comme on vient de le voir, un rôle privilégié aux arbres couvrants.

Il est dans ces conditions naturel d'attendre l'existence des formulations arborescentes équivalentes au problème de parking. De fait, on peut s'appuyer sur un article de Gessel et Wang [16] pour dédire une relation directe entre distribution des inversions dans les arbres de Cayley et déplacement total dans le parking. (Via des travaux plus anciens de Mallows et Riordan [25], les quantités en jeu se relient ultimement aux moments de la loi log-normale !) On retrouvera encore les arbres à la section 6 consacrée à l'analyse fine des phénomènes de coalescence.

5. Parking et excursion Brownienne

Vers le milieu du siècle dernier, un grand nombre de chercheurs issus d'horizons très divers (logique, combinatoire, théorie des probabilités, informatique, linguistique formelle, etc) découvrent différents aspects des riches relations existant entre les arbres et les chemins dans le plan—plus précisément, les excursions. On dispose effectivement d'une interprétation du parking en terme d'excursions, selon le principe suivant :

Principe d'équivalence parkings-excursions. *La loi du déplacement total dans les fonctions de parking de taille n coïncide, à décalage près, avec la loi de l'aire sous l'excursion de Poisson. La loi limite du déplacement est ainsi la loi de l'aire sous l'excursion Brownienne.*

Suivant Spencer [33], on remarque en effet que le nombre de fonctions de parking associées à une séquence $(y_1, y_2, \dots, y_n)$ satisfaisant à (15) est

$$\frac{n!}{x_1! x_2! \dots x_n!},$$

de probabilité proportionnelle à

$$\prod_{i=1}^{n} \frac{e^{-1}}{x_i!}.$$

Donc la loi de la séquence $(y_i)_{0\le i\le n+1}$ est celle d'une marche aléatoire partant de $y_0 = 1$, d' incréments $x_i - 1$, conditionnée à rester positive aux instants 1, 2, ..., n, et à revenir en zéro à l'instant $n+1$, les x_i suivant la loi de Poisson de moyenne 1. Ceci définit « l'excursion de Poisson ». Or on sait depuis Kaigh [23], que

$$\left(\frac{y_{\lfloor nt\rfloor}}{\sqrt{n}}\right)_{0\le t\le 1} \xrightarrow{\mathcal{L}} (e(t))_{0\le t\le 1}, \tag{16}$$

où $(e(t))$ désigne l'excursion Brownienne, et $\xrightarrow{\mathcal{L}}$ indique la convergence en loi. Il découle donc, comme conséquence immédiate,

$$\frac{D_{n+1,n}}{n\sqrt{n}} \xrightarrow{\mathcal{L}} \int_0^1 e(t)\, dt.$$

L'apparition des moments de l'excursion brownienne

$$\widehat{\mu}_k = E\left[\left(\int_0^1 e(t)\, dt\right)^k\right]$$

dans l'asymptotique (6) des moments du déplacement total en hachage, observée calculatoirement par Flajolet et al., devient ainsi explicable, et un peu de travail conduit à une démonstration probabiliste alternative [8].

Pour conclure signalons, dans cette orbite d'idées, une tentative encore largement heuristique de Spencer [34]. Si l'on suit la chaîne reliant l'aire de l'excursion aux graphes, on peut espérer estimer efficacement le nombre de graphes connexes d'excès k et taille n par l'observation des moments de l'aire sous l'excursion de Poisson. Les régimes interessants, liés au célèbre modèle de graphes

aléatoires d'Erdős et Rényi [6], correspondent à n et k tendant simultanément vers l'infini de manière approximativement proportionnelle. Des résultats asymptotiques sur les problèmes de graphes sont connus grâce à un article techniquement délicat de Bender, Canfield, et McKay [4]. Spencer présente dans [34] divers modèles et conjectures relatifs au comportement aléatoire de l'excursion de Poisson et susceptibles de constituer une alternative interessante à l'approche analytique de [4].

6. Parking et coalescence

En rapport avec le déplacement total, l'analyse de l'évolution de la taille des « ilots » (blocs de voitures garées consécutivement), au fur et à mesure des arrivées, présente l'intérêt de quantifier le coût d'insertion ou de recherche d'un élément. La taille du plus grand bloc, par exemple, représente une mesure du coût de rangement d'un élément dans le pire des cas. Les phénomènes de transition entre la phase éparse (avec $n = \alpha m$ et $\alpha < 1$) et la phase « dense » (où $n \sim m$) présentent un intérêt tout particulier ; un ensemble de blocs très « dilués » se « coagule » progressivement en un ilot unique qui finit par occuper l'ensemble de la place mémoire disponible, lorsque le rapport n/m devient égal à 1.

Transition de phase.

La loi de la taille $R^{m,n}$ du bloc de la $n^{\text{ème}}$ voiture, lors de son arrivée, est d'un abord facile, et son comportement asymptotique est instructif : on trouve directement, par les méthodes évoquées à la section 2,

$$\Pr\left(R^{m,n} = k\right) \quad = \quad \binom{n-1}{k-1} \frac{(k+1)^{k-1}\ (m-k-1)^{n-k-1}(m-n-1)}{m^{n-1}}.$$

Lorsque $n/m = \alpha$ et que n et m sont grands, on obtient sans peine, sous condition que α soit fixé et vérifie $\alpha < 1$:

$$\Pr\left(R^{m,n} = k\right) \quad \sim \quad (k+1)^{k-1} e^{-\alpha(k+1)} \frac{\alpha^{k-1}}{(k-1)!}(1-\alpha).$$

Ainsi, pour $\alpha < 1$, la configuration de parking s'organise en un grand nombre d'ilots de taille typique $O(1)$, et dont la distribution est bien caractérisée. (Pittel montre par ailleurs que le plus grand ilot conserve en probabilité une taille voisine de $\log n$; voir [29].)

La distribution de $R^{m,n}$ fuit vers l'infini, lorsque α se rapproche de 1. Le cas $\alpha = 1$ est précisément celui que nous avons déjà longuement étudié (fonctions de parking « pleines » ou « presque pleines »). Lorsque m et n tendent vers $+\infty$ tout en restant voisins, une région critique se dessine lorque $m - n$ est de l'ordre de $\sqrt{m}$. On pose $m - n = \beta\sqrt{m}$, et

$$\Pr\left(R^{m,n} = k\right) = \frac{1}{m}\ f\left(\beta, \frac{k}{m}\right) + O\left(m^{-3/2}\right),$$

où f est définie par

$$f(\beta, x) = \frac{\beta}{\sqrt{2\pi}}\ x^{-1/2}(1-x)^{-3/2} \exp\left(-\frac{\beta^2 x}{2(1-x)}\right) 1_{]0,1[}(x).$$

Si N suit la loi normale centrée réduite, f est en particulier la densité de probabilité de $\frac{N^2}{\beta^2+N^2}$; autrement dit, si $m-n \sim \beta\sqrt{m}$, on a

$$\frac{R^{m,n}}{m} \xrightarrow{\mathcal{L}} \frac{N^2}{\beta^2+N^2}.$$

Bien sûr, cette loi limite se concentre en 0 quand β tend vers $+\infty$, et en 1 quand β tends vers 0. En résumé, on a :

$$\beta = +\infty \ (\sqrt{m} = o(m-n)) \ : \qquad \frac{R^{m,n}}{m} \xrightarrow{P} 0,$$
$$\beta = 0 \ (\sqrt{m} = o(m-n)) \ : \qquad \frac{R^{m,n}}{m} \xrightarrow{P} 1.$$

En clair, quand $m-n \sim \sqrt{m}$, les blocs sont suffisamment proches les uns des autres ; leur taille augmente alors moins par l'adjonction de voitures fraîchement arrivées que par la coalescence avec d'autres blocs. Le plus grand bloc augmente de $O(m)$ voitures quand $\sqrt{m}$ voitures arrivent.

La construction de Bertoin.

Il se trouve que la loi limite de $R^{m,n}$ est apparue récemment dans la littérature, comme la loi conditionnelle du premier terme de la permutation, biaisée par la taille (voir [28] pour cette notion), des sauts du processus stable d'indice $1/2$ [28], conditionné, ou des morceaux du coalescent additif standard (cf. [3, relation (27)]). Cela incite à aller plus loin et à tenter de décrire précisément le comportement asymptotique, quand m tends vers l'infini, de la taille et de la position de chaque bloc au fur et à mesure des arrivées, ainsi que la manière dont ces blocs s'agglomèrent jusqu'à n'en former finalement qu'un seul : il s'avère que cette description est aisée à l'aide de quelques opérations simples sur l'excursion Brownienne [9].

Pour une fonction parking aléatoire de $m-1$ voitures sur m places, notons $B_k^{m,\ell} = \left(B_{k,1}^{m,\ell}, B_{k,2}^{m,\ell}\right)$ le couple formé de la taille du $k^{\text{ème}}$ plus grand bloc de voitures garées consécutivement et de la position de la première voiture de ce $k^{\text{ème}}$ bloc lorsque $m-\ell$ voitures se sont garées et qu'il reste encore ℓ places vides ; notons également $y_k^{m,\ell}$ le nombre de visites sur la place k ; notons finalement

$$B^{m,\ell} = \left(B_k^{m,\ell}\right)_{k\geqslant 1}$$

la suite des tailles et positions des blocs, complétée par une suite infinie de couples $(0,0)$. On considère les suites renormalisées

$$Z_a^{(m)}(t) = \frac{y_{\lceil mt\rceil}^{m,\lceil\beta\sqrt{m}\rceil}}{\sqrt{m}}, \qquad \Theta^{(m)}(\beta) = \frac{B^{m,\lceil\beta\sqrt{m}\rceil}}{m},$$
$$Z^{(m)} = \left(Z_\beta^{(m)}(t)\right)_{0\leq\beta,\,0\leq t\leq 1}, \qquad \Theta^{(m)} = \left(\Theta^{(m)}(\beta)\right)_{\beta\geq 0}.$$

Ces suites constituent respectivement les « profils de visite » et « profils de placement » du parking.

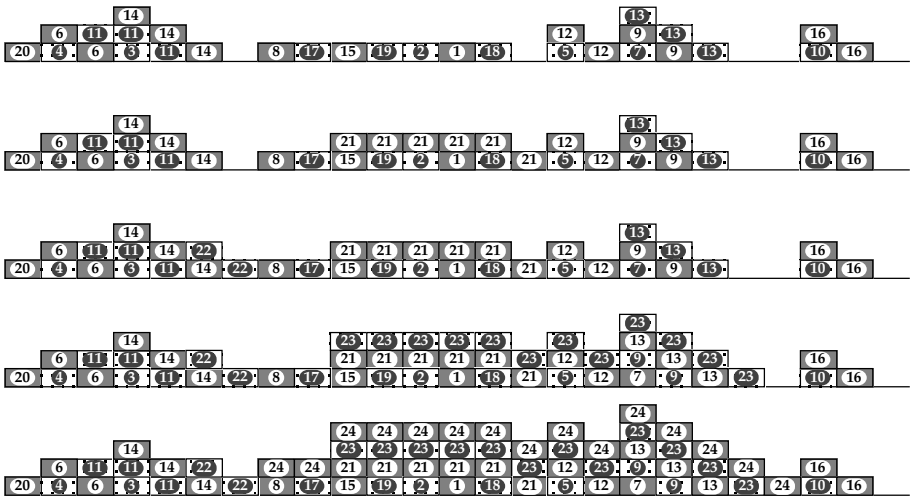


FIG. 4. Les $\left(y_k^{m,\ell}\right)_{0\leq k\leq 25}$ pour $m = 25$ et $\ell = 5, 4, 3, 2, 1$.

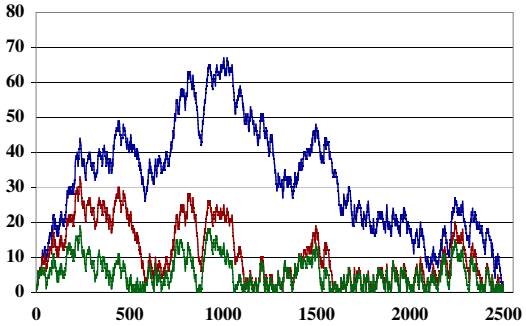


FIG. 5. $t \longrightarrow Z_\beta^{(m)}(t)$ pour $m = 2500$ places
et $\beta = 0, 2, 4$ successivement (1 puis 100 et 200 places vides)

Soit par ailleurs, comme précédemment, $(e(t))$ l'excursion Brownienne. On définit alors « l'excursion élaguée » construite selon la règle

$$\begin{aligned} Z_\beta(t) &= e(t) - \beta t + \sup_{t-1\leqslant s\leqslant t} (\beta s - e(s)), \\ Z &= (Z_\beta(t))_{0\leq\beta,\, 0\leq t\leq 1}\,, \end{aligned}$$

où β est un paramètre de contrôle. On note alors $\Theta(\beta)$ la suite des largeurs et positions des excursions du processus $t \to Z_\beta(t)$.

Avec ces notations, en utilisant un couplage *parking–processus empirique*, on peut construire, sur un espace de probabilité approprié, des copies de $Z^{(m)}$ et de Z, telles que soit vérifié le principe suivant (qui généralise (16)) :

Convergence du profil. *Dans la région critique, le profil des visites du parking converge vers l'excursion Brownienne élaguée,*

$$\Pr\left(Z^{(m)} \to Z\right) = 1,$$

où $Z^{(m)}$ converge vers Z au sens de la convergence uniforme sur tout compact de $[0,+\infty) \times [0,1]$. En conséquence, les distributions finies-dimensionelles du profil des placements du parking, $\Theta^{(m)}$, convergent vers les distributions finies-dimensionelles de la suite Θ des largeurs et positions des excursions de l'excursion élaguée.

Le processus des largeurs, $\Theta_2 = (\Theta_{k,2}(\beta))_{k\geq 1,\beta\geq 0}$, sans les positions, a été étudié pour lui-même par Bertoin [5], et comme on s'y attend au vu de la loi limite de $R^{m,n}$, Bertoin a démontré que Θ_2 était, à un changement de temps déterministe près, une version du *coalescent additif standard*, dont la première construction, à l'aide du CRT (« continuum random tree »), est due à Aldous et Pitman [3]. Comme le CRT est construit à l'aide de l'excursion Brownienne [2, Corollary 22], on pourrait espérer que l'effet de l'opération

$$(\Psi_\beta f)(t) = f(t) - \beta t + \sup_{t-1\leqslant s\leqslant t} (\beta s - f(s))$$

sur l'excursion Brownienne e se traduise par une fragmentation du CRT associé à e, qui soit exactement la fragmentation décrite par Aldous et Pitman. Ce n'est apparemment pas le cas. On constate cependant, voir [9, Section 8], que pour une bijection parking-forêt bien choisie, la fragmentation des blocs du parking au fur et à mesure du départ des voitures (en inversant le cours du temps) se traduit par une fragmentation de l'arbre correspondant en une forêt, qui est exactement la version discrète du coalescent additif, utilisée par Aldous et Pitman [3, Section 1] pour construire le coalescent additif standard, par un passage à la limite.

Le plus grand bloc.

Revenons sur un problème abordé, sans être totalement élucidé, dans [3, Section 3.4], la loi de la taille $Y_1(\beta)$ du plus gros morceau du coalescent additif,

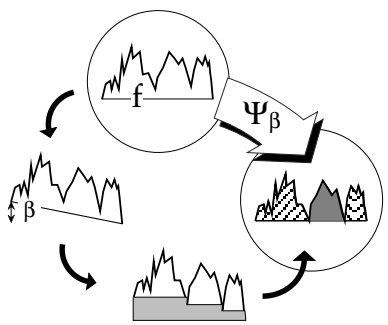


FIG. 6. Fragmentation des excursions par l'opérateur Ψ_β.

qui est aussi, par une chaîne d'équivalences évoquée ci-dessus, la loi limite de la taille du plus grand bloc d'un parking à m places contenant $m-\beta\sqrt{m}$ voitures (avec nos notations, la loi de $\Theta_{1,2}(\beta)$).

Dans un article de 1977 [27] (où, incidemment, le phénomène de transition de phase ci-dessus apparaissait déja en filigrane), Pavlov obtenait, en combinant astucieusement arguments probabilistes et combinatoires, des résultats asymptotiques très complets sur la loi de la taille du plus grand arbre d'une "forêt" tirée au hasard parmi les forêts à $m-n$ arbres et m nœuds (ou feuilles), pour les différents régimes de convergence du couple (m,n) vers $+\infty$. Les bijections arbres $\leftrightarrow$ parking explicitées plus haut permettent de transposer immédiatement les résultats de Pavlov au cadre du parking, donc au coalescent additif, et donnent l'expression suivante pour $\Pr(Y_1(\beta)\le x)$:

$$1+\beta^3 e^{\beta^2/2}\sum_{k\ge 1}\frac{(-1)^k}{k!}\int_{D(\beta,x,k)}\frac{\beta^{2k}\exp\{-\beta^4/2(\beta^2-x_1-\cdots-x_k)\}dx_1\ldots dx_k}{(2\pi)^{k/2}(x_1\ldots x_k(\beta^2-x_1-\cdots-x_k))^{3/2}},$$

où

$$D(\beta,x,k)=\left\{(x_i)_{1\le i\le k}\ :\ x_i\ge\beta^2 x, 1\le i\le k,\ \text{and}\ \sum x_i\le\beta^2\right\}.$$

Le lecteur trouvera dans le livre de Kolchin [20] un compte-rendu synthétique des travaux analytico-probabilistes de l'École Russe sur ces sujets.

7. Conclusion

Nous avons voulu montrer sur un cas particulier, celui du hachage, une surprenante convergence d'approches combinatoires, probabilistes, et analytiques. De tels cas, issus de l'analyse d'algorithmes informatiques sont en fait très nombreux. En France, dans l'orbite de la théorie des probabilités, le groupe de Versailles (Chauvin, Rouault, *et al.*) a pu par exemple montrer tout l'intérêt d'une approche par martingale dans l'étude des très classiques arbres binaires de recherche (voir par exemple la thèse de Jabbour, 2000). G. Louchard à Bruxelles a au fil des ans établi que le comportement d'algorithmes fort divers se quantifie très précisément grâce aux propriétés du mouvement Brownien ; voir aussi les travaux de l'un d'entre nous et de J.-F. Marckert. Ces questions et bien d'autres[4] sont notamment débattues au sein du groupe de travail ALÉA lequel tient une réunion annuelle sur le thème de l'aléa discret et de ses applications informatiques. L'importance de ces questions découle d'un besoin très réel de compréhension, validation, et optimisation des principales méthodes de traitement de données non-numériques.

[4]Dans ce texte, l'accent a largement été mis sur les aspects probabilistes. Les méthodes combinatoires incluent un riche ensemble de transformations bijectives de structures, la théorie des séries formelles et l'algèbre des séries, etc. Les méthodes analytiques reposent désormais très largement sur l'analyse complexe, les équations fonctionnelles, l'analyse de singularité, l'analyse fonctionnelle, etc.

8. Références

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover (1973).
[2] D. Aldous, *The continuum random tree III.* Ann. Probab. 21, 248-289 (1993).
[3] D. J. Aldous & Jim Pitman, *The standard additive coalescent.* Ann. of Proba. 26, 1703-1726 (1998).
[4] E.A. Bender, E.R. Canfield et B.D. McKay, *The asymptotic number of labeled connected graphs with a given number of vertices and edges.* Random Structures Algorithms 1 (1990), no. 2, 127–169.
[5] J. Bertoin, *A fragmentation process connected to Brownian motion.* Probab. Theory Relat. Fields 117, no. 2, 289-301 (2000).
[6] Béla Bollobás, *Random Graphs*, Academic Press (1985).
[7] A. Cayley, (1889) *A theorem on trees.* Quarterly Journal of Pure and Applied Math. 23, 376-378.
[8] P. Chassaing & J.F. Marckert, *Parking functions, empirical processes and the width of rooted labeled trees.* The Electronic Journal of Combinatorics 8 (2001), R14.
[9] P. Chassaing & G. Louchard, *Phase transition for parking blocks, Brownian excursion and coalescence.* To appear in Random Structures Algorithms, available at : http ://www.iecn.u-nancy.fr/∼chassain
[10] L. Comtet, *Advanced Combinatorics*, Reidel, Dordrecht (1974).
[11] P. Flajolet & G. Louchard, *Analytic Variations on the Airy Distribution*, *Algorithmica*, Special Issue on Analysis of Algorithms, 31, 361-377 (2001).
[12] P. Flajolet and A. M. Odlyzko, *Singularity Analysis of Generating Functions*, SIAM Journal on Algebraic and Discrete Methods 3, No. 2, 216-240 (1990).
[13] P. Flajolet, P. Poblete, A. Viola, *On the Analysis of Linear Probing Hashing*, Algorithmica Vol. 22, No. 4, 490-515 (1998).
[14] D. Foata & J. Riordan, *Mappings of acyclic and parking functions.* Aequationes math. 10, 10-22 (1974).
[15] J. Françon, *Acyclic and parking functions.* J. Combinat. Theory Ser. A 18, 27-35 (1975).
[16] I. Gessel and Da Lun Wang, *Depth-first search as a combinatorial correspondence*, Journal of Combinatorial Theory, Series A 26, No. 3, 308–313 (1979).
[17] S. Janson, D.E. Knuth, T. Łuczak, & B. Pittel, *The birth of the giant component.* Random Structures Algorithms 4, No. 3, 231-358 (1993).
[18] D.E. Knuth, *The art of computer programming. Vol. 3 : sorting and searching*, 2nd ed., Addison-Wesley (1997).
[19] D. E. Knuth, *Linear Probing and Graphs*, Algorithmica 22, No. 4, 561-568 (1998).
[20] V. F. Kolchin, *Random Graphs*, In Encyclopedia of Mathematics and its Applications, Cambridge University Press (1999).
[21] A.G. Konheim, B. Weiss, *An occupancy discipline and applications*, SIAM J. Appl. Math. 14, No. 6, 1266-1274 (1966).
[22] G. Kreweras, *Une famille de polynômes ayant plusieurs propriétés énumeratives*, Period. Math. Hungar. 11, no. 4, 309–320 (1980).
[23] W. D. Kaigh, *An invariance principle for random walk conditioned by a late return to zero*, Ann. Probability 4, no. 1, 115–121 (1976).
[24] G. Louchard, *The Brownian excursion : a numerical analysis*, Computers and Mathematics with Applications 10, No. 6, 413–417 (1984).
[25] C. L. Mallows and John Riordan, *The inversion enumerator for labeled trees*, Bulletin of the American Mathematical Society 74, 92–94 (1968).
[26] A. M. Odlyzko, *Asymptotic Enumeration Methods*, In Handbook of Combinatorics II, R. Graham, M. Grötschel, and L. Lovász, Eds, Elsevier, Amsterdam, 1063–1229 (1995).
[27] Yu. L. Pavlov, *The asymptotic distribution of maximum tree size in a random forest.* Th. Probab. Appl. 22, 509-520 (1977).
[28] M. Perman, J. Pitman, M. Yor, *Size-biased sampling of Poisson point processes and excursions.* Probab. Theory Related Fields 92, No. 1, 21-39 (1992).
[29] B. Pittel, *Linear probing : the probable largest search time grows logarithmically with the number of records.* J. Algorithms 8, No. 2, 236-249 (1987).

[30] A. Rényi, *On a one-dimensional problem concerning random space filling*. Magyar Tud. Akad. Mat. Kutató Int. Közl. 3, No. 1/2, 109–127 (1958).
[31] J. Riordan, *Ballots and trees*. J. Combinatorial Theory 6, 408-411 (1969).
[32] M. P. Schutzenberger, *On an enumeration problem*. J. Combinatorial Theory 4, 219-221 (1968).
[33] J.H. Spencer, *Enumerating graphs and Brownian motion*. Commun. Pure Appl. Math. 50, No. 3, 291-294 (1997).
[34] J.H. Spencer, *Ultrahigh moments for a Brownian excursion*. Mathematics and computer science (Versailles, 2000), 323–328, Trends Math., Birkhäuser, Basel, 2000.
[35] L.Takács, *On a probability problem connected with railway traffic*, Journal of Applied Mathematics and Stochastic Analysis 4, No. 1, 1-27 (1991).
[36] E. M. Wright, *The number of connected sparsely edged graphs,* J. Graph Theory 1 (1977), no. 4, 317–330.