On Generating Functions of Generating Trees

Cyril Banderier (INRIA) Mireille Bousquet-Mélou (LaBRI)
Alain Denise (LRI) Philippe Flajolet (INRIA) Danièle Gardy (PRISM)
Dominique Gouyou-Beauchamps (LRI) *

Abstract

Generating trees describe conveniently certain families of combinatorial objects: each node of the tree corresponds to an object, and the branch leading to the node encodes the choices made in the construction of the object. Generating trees lead to a fast computation of enumeration sequences (sometimes, to explicit formulae as well) while providing efficient random generation algorithms. In this paper, we investigate the relationship between structural properties of the rules defining such trees and the rationality, algebraicity, or transcendence of the corresponding generating functions.

Résumé

Certaines méthodes d'énumération d'objets combinatoires utilisent des arbres infinis, ou arbres de génération, qui résument dans leurs branches et leurs noeuds les choix faits lors de la génération des objets. Les arbres de génération conduisent à des algorithmes de calcul des suites de dénombrement ainsi que de génération aléatoire qui sont rapides. Nous étudions les liens entre les propriétés structurelles de tels arbres, ou plutôt des systèmes de règles associés, et la nature (rationnelle, algébrique ou transcendante) de la série génératrice qui leur correspond ; cette série énumère les nœuds de niveau donné de l'arbre, *i.e.*, les objets de taille donnée.

1 Introduction

Only the simplest combinatorial structures — like binary strings, permutations, or *pure* involutions (*i.e.*, involutions with no fixed point) — admit product decompositions. In that case, the set Ω_n of objects of size n is isomorphic to a product set: $\Omega_n \cong [1, e_1] \times [1, e_2] \times \cdots \times [1, e_n]$. Two properties result from such a strong decomposability property: (i) enumeration is easy, since the cardinality of Ω_n is $e_1 e_2 \cdots e_n$; (ii) random generation is efficient since it reduces to a sequence of random independent draws from intervals. In that case, a simple infinite tree, called the *uniform generating tree* is determined by the e_j: the root has degree e_1, each of its e_1 descendents has degree e_2, and so on. This tree describes the sequence of all possible choices and the objects of size n are then in natural correspondence with the branches of length n, or equivalently with the nodes of generation n in the tree. The generating tree is thus fully described by its root degree (e_1) and by rewriting rules, here of the special form,

$$(e_j) \rightsquigarrow (e_{j+1})(e_{j+1}) \cdots (e_{j+1}) \equiv (e_{j+1})^{e_j},$$

where the power notation is used to express repetitions. For instance binary strings, permutations, or pure involutions are determined by

$$\begin{array}{rl} \mathcal{S}: & [(2),\ (2) \rightsquigarrow (2)(2)] \\ \mathcal{P}: & [(1),\ \{(j) \rightsquigarrow (j+1)^j\}_{j\geq 1}] \\ \mathcal{I}: & [(1),\ \{(2j-1) \rightsquigarrow (2j+1)^{2j-1}\}_{j\geq 1}]. \end{array}$$

*E-mails: Cyril.Banderier@inria.fr, bousquet@labri.u-bordeaux.fr, Alain.Denise@lri.fr, Philippe.Flajolet@inria.fr, Daniele.Gardy@prism.uvsq.fr, dgb@lri.fr

A powerful generalization of this idea consists in considering unconstrained *generating trees* where any set of rules

$$\Sigma = [(s_0),\ \{(k) \rightsquigarrow (e_{1,k})\,(e_{2,k})\,\cdots\,(e_{k,k})\}] \tag{1}$$

is allowed. Here, the *axiom* (s_0) specifies the degree of the root, while the *productions* list the degrees of the k descendents of a node labelled k. Obviously, much more leeway is available and there is hope to describe a much wider class of structures than those corresponding to product forms and uniform generating trees.

The idea of generating trees that we have just described has surfaced occasionally in the literature. West introduced it in the context of enumeration of permutations with forbidden subsequences [18, 19]; this idea has been further exploited in closely related problems [3, 4, 9, 10]. A major contribution in this area is due to Barcucci, Del Lungo, Pergola, and Pinzani [2, 5] who systematized the method under the name of ECO-systems (ECO stands for "Enumerating Combinatorial Objects"), while showing that a fairly large number of classical combinatorial structures are amenable to such descriptions by generating trees.

A form equivalent to generating trees is well worth noting at this stage. Consider the set of *walks* on the integer half-line that start at point (s_0) and such that the only allowable transitions are those specified by Σ. Then, clearly, the set of such walks of length n is in bijective correspondence with branches of the tree. Thus, the model of generating trees is equivalent to walks of the form (1). The walks are only constrained by the consistency requirement of trees, namely, that the number of outgoing edges from point k on the half-line has to be exactly k. Such an alternative presentation in terms of walks implies that objects that admit generating trees can be *enumerated in cubic time*, given the rules in tabular form, and provided the $e_{i,k}$ are bounded linearly in k. (See below for details.)

Example 1. 123-*avoiding permutations.* The method of "local expansion" sometimes gives good results in the enumeration of permutations avoiding specified patterns. Consider for example the set $\mathfrak{S}_n(123)$ of permutations of length n that *avoid the pattern* 123: there exist no integers $i < j < k$ such that $\sigma(i) < \sigma(j) < \sigma(k)$. For instance, $\sigma = 4213$ belongs to $\mathfrak{S}_4(123)$ but $\sigma = 1324$ does not, as $\sigma(1) < \sigma(3) < \sigma(4)$.

Observe that if $\tau \in \mathfrak{S}_{n+1}(123)$, then the permutation σ obtained by erasing the entry $n+1$ from τ belongs to $\mathfrak{S}_n(123)$. Conversely, for every $\sigma \in \mathfrak{S}_n(123)$, insert the value $n+1$ in each possible place (this is the local expansion). For example, the permutation $\sigma = 213$ gives 4213, 2413 and 2143, by insertion of 4 in first, second and third place respectively. The permutation 2134, resulting of the insertion of 4 in the last place, does not belong to $\mathfrak{S}_4(123)$. This process can be described by a generating tree whose nodes are the permutations avoiding 123: the root is 1, and the children of any node σ are the permutations derived as above. Figure 1(a) presents the first four levels of this tree.

Let us now label the nodes by their number of children: we obtain the tree of figure 1(b). It can be proved that the k children of any node labelled k are labelled respectively $k+1, 2, 3, \ldots, k$. Thus the generating tree can be defined by giving only the value of the label of the root and the *succession rule* just defined. This can be written (after re-ordering the labels) as

$$[(2),\ \{(k) \rightsquigarrow (2)(3)\ldots(k-1)(k)(k+1)\}_{k\geq 2}]. \tag{2}$$

The equivalence with paths then implies that 123-avoiding permutations are equinumerous with "walks with returns" on the half-line, themselves isomorphic to Łukasiewicz codes of general trees. Thus, 123-avoiding permutations are counted by Catalan numbers. □

The main question addressed in this paper is the relationship between structural properties of the rules defining generating trees on the one hand, and properties of generating functions on the other hand. Since generating trees are associated with fast random generation algorithms and with enumeration sequences of relatively low computational complexity, there is an obvious interest in delineating as precisely as possible which combinatorial classes admit a generating tree specification. Generating functions that condense structural information in a simple analytic entity are prime candidates to be examined.

In the course of their investigations, Pinzani and his coauthors made a number of observations that were presented to us as conjectures in March 1998. This paper is devoted to bringing complete proofs of several of Pinzani's conjectures. Our main results are as follows.

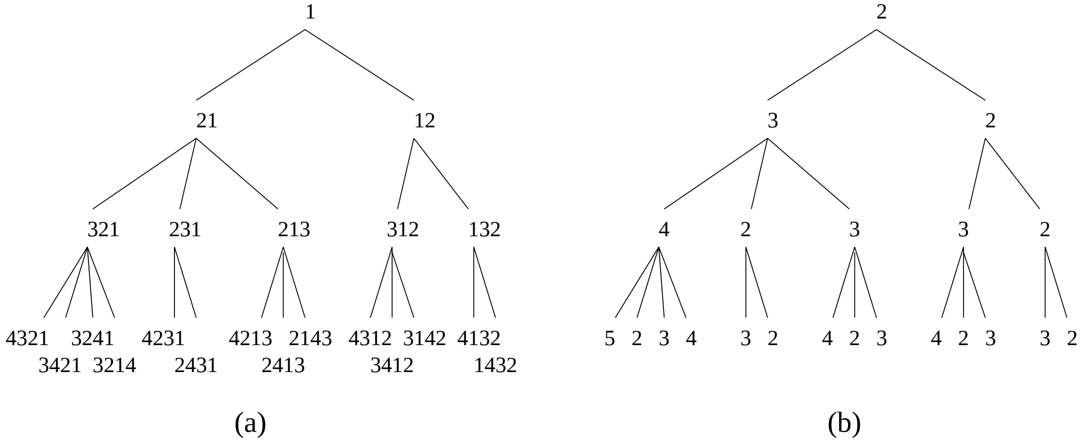

Figure 1: The generating tree of 123-avoiding permutations. (a) nodes labelled by the permutations. (b) nodes labelled by the numbers of children.

— *Rational systems.* Systems satisfying strong regularity conditions lead to rational generating functions (Section 2). This covers systems that have a finite number of allowed degrees, as well as systems like (a), (b), (c), and (d) in Example 2 below where the labels are constant except for a fixed number of labels that depend linearly and "uniformly" on k.

— *Algebraic systems.* Systems of a "factorial" form, *i.e.*, where a finite modification of the set $\{1,\ldots,k\}$ is reachable from k, lead to algebraic generating functions (Section 3). This includes in particular cases (f) and (g) in Example 2.

— *Transcendental systems.* One possible reason for a system to give a transcendental series is the fact that its coefficients grow too fast, so that its radius of convergence is zero. Transcendental generating functions are also associated with systems that are too "irregular" (Section 4). Instances are cases (e) and (h) of Example 2.

Example 2. *Particular generating tree systems.* Here is a list of examples recurring throughout this paper.

(a): $[(3),\{(k)\rightsquigarrow(3)^{k-3}(k+1)(k+2)(k+9)\}]$ (b): $[(3),\{(k)\rightsquigarrow(3)^{k-1}(3k+6)\}]$
(c): $[(2),\{(k)\rightsquigarrow(2)^{k-2}(2+(k \bmod 2))(k+1)\}]$ (d): $[(2),\{(k)\rightsquigarrow(2)^{k-2}(3-(k \bmod 2))(k+1)\}]$
(e): $[(3),\{(k)\rightsquigarrow(2)^{k-2}(3-[\exists p\!:k=2^p])(k+1)\}]$ (f): $[(2),\{(k)\rightsquigarrow(2)(3)\ldots(k-1)(k)(k+1)\}]$
(g): $[(1),\{(k)\rightsquigarrow(1)(2)\ldots(k-1)(k+1)\}]$ (h): $[(2),\{(k)\rightsquigarrow(2)(3)(k+2)^{k-2}\}]$

(In (e), we make use of Iverson's brackets: $[P]$ equals 1 if P is true, 0 otherwise.) □

Notations. From now on, we adopt functional notations for rewriting rules: systems will be of the form

$$[(s_0),\quad \{(k)\rightsquigarrow(e_1(k))\,(e_2(k))\ldots(e_k(k))\}]$$

where s_0 is a constant and each e_i is a function of k. Moreover, we assume that all the values appearing in the generating tree are positive.

In the generating tree, let f_n be the number of nodes at level n and s_n the sum of the labels of these nodes. (By convention, the root is at level 0, so that $f_0=1$.) In terms of walks, f_n is the number of walks of length n. The generating function associated to the system is

$$F(z)=\sum_{n\geq 0} f_n z^n.$$

Remark that $s_n=f_{n+1}$, and the f_n's are nondecreasing.

Now let $f_{n,k}$ be the number of nodes at level n having label k (or the number of walks of length n ending at position k). The following generating functions will be also of interest:

$$F(z,u) = \sum_{n,k\geq 0} f_{n,k} z^n u^k \qquad \text{and} \quad F_k(z) = \sum_{n\geq 0} f_{n,k} z^n.$$

We have $F(z) = F(z,1)$ and $F(z) = \sum_{k\geq 1} F_k(z)$. Furthermore, the F_k's satisfy the relation

$$F_k(z) = [k = s_0] + z \sum_{i\geq 1} \pi_{i,k} F_i(z), \tag{3}$$

where $\pi_{i,k} = |\{j \leq i : e_j(i) = k\}|$ denotes the number of one-step transitions from i to k. This is equivalent to the following recurrence for the quantities $f_{n,k}$,

$$f_{0,s_0} = 1 \qquad \text{and} \quad f_{n+1,k} = \sum_{i\geq 1} \pi_{i,k} f_{n,i}, \tag{4}$$

that results from tracing all the paths that lead to k in $n+1$ steps.

Counting and random generation. The recurrence (4) gives rise to an algorithm that determines the successive values of the array $f_{n,k}$ by "forward propagation": For each n, i, propagate the contribution $f_{n,i}$ to $f_{n+1,k}$ whenever $e_j(i) = k$. Consider for this discussion "linearly bounded systems" where the states reachable in m steps have an index (a label) dominated by a linear function of m. (Systems where forward jumps are bounded by an absolute constant are for instance of this type.) Clearly, the forward propagation algorithm provides a counting algorithm of arithmetic complexity that is at most cubic. In that case, random generation can also be achieved in polynomial time, as we now show.

Let $g_{k,n}$ be the number of walks of length n that start from state k taken as axiom. The $g_{k,n}$ are then determined by a "backward" recurrence, $g_{k,n} = \sum_j g_{e_j(k),n-1}$, that traces all the possible continuations of a path given its initial step. Obviously, $f_n = g_{s_0,n}$, with s_0 the axiom. The $g_{k,n}$ form an array that is dual to the $f_{n,k}$ and, for a linearly bounded system, they can be determined in time $O(n^3)$, like before. Random generation is then achieved as follows: In order to generate an object of size n starting from state k, pick up a transition j with probability $g_{e_j(k),n-1}/g_{k,n}$, and generate recursively an object of size $n-1$ starting from state $e_j(k)$. The recursive procedure needs to set up the array $g_{k,n}$, which represents a preprocessing cost of $O(n^3)$ time and $O(n^2)$ storage. The cost of a single random generation is then $O(n^2)$ if a sequential search is used over the $O(n)$ possibilities of each of the n random drawings; the time complexity goes down to $O(n \log n)$ if binary search is used, but at the expense of an increase in storage complexity of $O(n^3)$ (arising from $O(n^2)$ arrays of size $O(n)$ that binary search requires).

2 The rational case

We give in this section four criteria implying that the generating function of a given ECO-system is rational. All the systems studied here have the following property: A bounded number of e_i's grow at most linearly in k, and the others are bounded by a constant.

Among these systems, the simplest ones are those in which *all* the e_i's are bounded.

Proposition 1 *If finitely many labels appear in the tree, then $F(z)$ is rational.*

Sketch of Proof. Only a finite number of F_k's are nonzero, and they are defined by linear equations like Equation (3) above. □

EXAMPLE 3. *Fibonacci sequence.* The system $[(1), \{(k) \rightsquigarrow (k)^{k-1}((k \bmod 2)+1)\}]$, which can be also written as $[(1), \{(1) \rightsquigarrow (2), (2) \rightsquigarrow (1)(2)\}]$, leads to $F(z) = \frac{1}{1-z-z^2} = 1 + z + 2z^2 + 3z^3 + 5z^4 + \cdots$, the well-known Fibonacci generating function. □

None of the systems of Example 2 satisfy directly the assumptions of Proposition 1. However, the proposition that follows can be applied to systems (a) and (b).

Proposition 2 *Let $\sigma(k) = e_1(k) + e_2(k) + \cdots + e_k(k)$. If σ is an affine function of k, say $\sigma(k) = \alpha k + \beta$, then the series $F(z)$ is rational. More precisely:*

$$F(z) \;=\; \frac{1 + (s_0 - \alpha)z}{1 - \alpha z - \beta z^2}.$$

Proof. Let $n \geq 0$ and let $k_1, k_2, \ldots k_{f_n}$ denote the labels of the f_n nodes at level n. Then

$$\begin{aligned} f_{n+2} \;=\; s_{n+1} &\;=\; (\alpha k_1 + \beta) + (\alpha k_2 + \beta) + \cdots + (\alpha k_{f_n} + \beta) \\ &\;=\; \alpha s_n + \beta f_n \;=\; \alpha f_{n+1} + \beta f_n. \end{aligned}$$

We know that $f_0 = 1$. The result follows. □

Example 4. *Bisection of Fibonacci sequence.* The system $[(2), \{(k) \rightsquigarrow (2)^{k-1}(k+1)\}]$ gives $F(z) = \frac{1-z}{1-3z+z^2} = 1 + 2z + 5z^2 + \cdots$, the generating function for every odd entry in the Fibonacci sequence. (Changing the axiom to (3) leads to the other half of the Fibonacci sequence.) Systems $[(2), \{(k) \rightsquigarrow (1)^{k-1}(2k)\}]$, as well as $[(2), \{(k) \rightsquigarrow (2)^{k-2}(3 - (k \bmod 2))(k + (k \bmod 2))\}]$ and $[(2), \{(k) \rightsquigarrow (2)^{k-2}(3 - [k \text{ is prime}])(k + [k \text{ is prime}])\}]$ lead to the same function $F(z)$ since $\sigma(k) = 3k - 1$ and $s_0 = 2$ in all cases. However, the generating trees are different, as are the bivariate functions $F(z, u)$. □

Proposition 2 can be slightly generalised. For example, let us consider a system having the following properties: (i) the system can be decomposed into two productions, one for even k and one for odd k, such that the corresponding functions σ_0 and σ_1 are affine and have the same leading coefficient α, say $\sigma_0(k) = \alpha k + \beta_0$ and $\sigma_1(k) = \alpha k + \beta_1$; (ii) there exists a constant c such that exactly c odd labels occur in the right-hand side of each rule. An argument similar to the proof of Proposition 2 leads to the following result:

Proposition 3 *If a system satisfies properties (i) and (ii) above, then*

$$F(z) = \frac{1 + (s_0 - \alpha)z + (s_1 - \alpha s_0 - \beta_0)z^2}{1 - \alpha z - \beta_0 z^2 - c(\beta_1 - \beta_0)z^3}.$$

For example, system (c) in Example 2 can be rewritten $[(2), \{(2k) \rightsquigarrow (2)^{2k-2}(2)(2k+1), (2k+1) \rightsquigarrow (2)^{2k-1}(3)(2k+2)\}]$. It satisfies properties (i) and (ii) above with $\alpha = 3$, $\beta_0 = -1$, $\beta_1 = 0$ and $c = 1$. Consequently, its generating function is $F(z) = \frac{1-z}{1-3z+z^2-z^3}$.

System (d), although very close to (c), does not satisfy property (ii) above, so that Proposition 3 does not apply. We then consider systems of the form

$$[(s_0), \{(k) \rightsquigarrow (c_1(k))(c_2(k)) \ldots (c_{k-K}(k))(k + a_1)^{\lambda_1}(k + a_2)^{\lambda_2} \ldots (k + a_m)^{\lambda_m}\}] \tag{5}$$

where $0 < a_1 < a_2 < \cdots < a_m$ and the $c_i(k)$ are uniformly bounded by a constant $C \geq s_0$.

Proposition 4 *Consider the system* (5)*, and let $\pi_{i,k} = |\{j \leq i : e_j(i) = k\}|$. If all the series*

$$\sum_{j \geq 1} \pi_{j,k} \; t^j$$

for $k \leq C$ are rational, then so is the series $F(z)$.

Sketch of Proof. We form an infinite system of equations defining the series $F_k(z)$ by writing (3) for all $k \geq 1$. The bottom part of the system $(k > C)$ is diagonal, and the solution of the corresponding equations yields, for $k \geq 1$:

$$F_k(z) = \sum_{i=1}^{C} P_{i,k}(z) F_i(z) \tag{6}$$

where the $P_{i,k}$ are polynomials in z defined by the following recurrence: for all $i \le C$,

$$P_{i,k}(z) = \begin{cases} [k = i] & \text{if } k \le C, \\ z \sum_{\ell=1}^{m} \lambda_\ell P_{i,k-a_\ell}(z) & \text{if } k > C, \end{cases} \tag{7}$$

with the convention $P_{i,k} = 0$ if $k < 1$.

Using (7), we find

$$F(z) = \sum_{k \ge 1} F_k(z) = \sum_{i=1}^{C} F_i(z) \left[\sum_{k \ge 1} P_{i,k}(z) \right].$$

According to (7), $\sum_{k \ge 1} P_{i,k}(z) t^k$ is a rational function in z and t, of denominator $1 - z \sum_\ell \lambda_\ell t^{a_\ell}$. At $t = 1$, it is rational in z. Hence, to prove the rationality of $F(z)$, it suffices to prove the rationality of the $F_i(z)$, for $i \le C$.

Let us go back to the C first equations of our system; using again (7), we find, for $k \le C$:

$$F_k = [k = s_0] + z \sum_{i=1}^{C} F_i(z) \left[\sum_{j \ge 1} P_{i,j}(z) \pi_{j,k} \right].$$

Again, we can prove that $\sum_{j \ge 1} P_{i,j}(z) \pi_{j,k}$ is a rational function of z (the Hadamard product of two rational series is rational). Thus the series $F_k(z)$, for $k \le C$, satisfy a linear system with rational coefficients: they are rational themselves, as well as $F(z)$. $\square$

Examples (a), (c), (d) and (e) of Example 2 have the form (5). The proposition above implies that the first three have a rational generating function. System (e) will be discussed in Section 4.

3 The algebraic case

In this section, we consider systems that are of a "factorial" form. By this, we mean informally that the rules giving the successors of (k) are a finite modification of the integer interval $\{1, 2, \ldots, k\}$. As was detailed in the introduction, generating tree rules can be rephrased in terms of walks over the integer half-line. We thus consider the marginally more general problem of enumerating walks over the integer half-line such that the allowed moves from point k is a finite modification of the integer interval $[0, k]$. Precisely, a *factorial walk* is defined by its moves from point $k \ge 0$ that are of the form

$$(k) \rightsquigarrow (0)(1) \cdots (k - c - 1)(k + d_1)(k + d_2) \cdots (k + d_m), \tag{8}$$

with $c \ge 0$ and $-c < d_1 \le d_2 \le \cdots \le d_m > 0$. In other words, a finite number of forward jumps are allowed and all backward jumps of length at least $c + 1$ are possible when moving from point k.

The collection of *factorial generating trees* is then defined as those systems that, up to a possible shift of indices, correspond to factorial walks. The rules are then

$$(k + r_0) \rightsquigarrow (r_0)(r_0 + 1) \cdots (k + r_0 - c - 1)(k + r_0 + d_1)(k + r_0 + d_2) \cdots (k + r_0 + d_m),$$

that is,

$$(k) \rightsquigarrow (r_0)(r_0 + 1) \cdots (k - c - 1)(k + d_1)(k + d_2) \cdots (k + d_m), \qquad \text{for } k \ge r_0 \ge 1.$$

Such systems must also obey the consistency principle of generating trees, viz., a node labelled k has exactly k successors; here this implies the further restriction $r_0 + c = m$. For instance, Systems (f) and (g) of Example 2 are factorial.

We prove here that any system of walks of type (8) has an algebraic generating function. The result thus applies to generating trees given by factorial rules. We consider again the generating function $F(z,u) = \sum_{n,k\geq 0} f_{n,k} z^n u^k$, where $f_{n,k}$ is the number of walks of length n ending at point k. We also let $f_n(u)$ be the coefficient of z^n in this series. The first idea is based on introducing a linear operator M such that

$$f_{n+1}(u) = M f_n(u).$$

This operator is constructed in stages by means of an operator L that records symbolically all possible moves, and then, by modifying L in order to take into account the boundary conditions that forces the walk to be always nonnegative. Let $\{b_1, b_2, \ldots\} = \{d_j : d_j \geq 0\}$ be the set of allowed forward jumps. Similarly, let $\{a_1, a_2, \ldots\} = [1, c] \setminus \{-d_j : d_j < 0\}$ be the set of irregular backward jumps.

— The set of moves from k to all the positions $0, 1, \ldots, k-1$ is described by an operator L_0 that maps u^k to $u^0 + u^1 + \cdots + u^{k-1} = (1-u^k)/(1-u)$. Consequently, let

$$L_0[f](u) = \frac{f(1) - f(u)}{1-u}.$$

— The fact that transitions near k are modified, with those of type $k + b_j$ (with $b_j \geq 0$) allowed and those of type $k - a_j$ (with $0 < a_j \leq c$) disallowed is expressed by a Laurent polynomial,

$$P(u) = B(u) - A(u) \quad \text{with} \quad B(u) = \sum_j u^{b_j}, \; A(u) = \sum_j u^{-a_j}. \tag{9}$$

Then, the operator

$$L[f](u) := L_0[f](u) + P(u) f(u)$$

plays the rôle of a generating operator for a single step of the walk.

— The modified operator M is given by

$$M[f](u) = L[f](u) - \{u^{<0}\} L[f](u),$$

where $\{u^{<0}\} f$ is the sum of all the monomials in f that involve negative exponents. This is nothing but L stripped of negative exponent monomials that correspond to noncombinatorial situations.

Assume for simplicity that the initial point of the walk is 0; other cases follow by the same argument. The linear relation $f_{n+1}(u) = M[f_n](u)$, together with $f_0(u) = 1$ yields

$$F(z,u) = \sum_{n\geq 0} f_n(u) z^n = 1 + z \left(\frac{F(z,1)}{1-u} - \frac{F(z,u)}{1-u} + P(u) F(z,u) - \{u^{<0}\} \sum_{n\geq 0} z^n L[f_n](u) \right). \tag{10}$$

One has $\{u^{<0}\} L f_n(u) = \sum_{j=0}^{c-1} c_j(u) \partial_u^j f_n(0)$, where $c_j(u)$ is a Laurent polynomial with monomials whose degrees belong to $[j-c, \ldots, -1]$. Thus, equation (10) implies our main equation,

$$F(z,u) \left(1 + \frac{z}{1-u} - z P(u) \right) = 1 + \frac{z}{1-u} F(z,1) - z \sum_{j=0}^{c-1} c_j(u) \partial_u^j F(z,0). \tag{11}$$

Therefore, the bivariate generating function $F(z,u)$ satisfies a *functional differential equation.*

The quantities that appear in the functional equation are all explicit. For instance, the moves

$$(k) \rightsquigarrow (0)(1) \cdots (k-5)(k-3)(k-1)(k)(k+7)(k+9),$$

lead to $A(u) = u^{-4} + u^{-2}$ and $B(u) = u^0 + u^7 + u^9$, with $P(u) = B(u) - A(u)$. In general, the degree of P is $d := d_m$, the size of the largest forward jump; the smallest degree occurring in P is c, the size of the largest disallowed backward jump.

The second ingredient is sometimes known as the kernel method[1]. This consists in forcing the left hand-side of the fundamental functional equation (11) to be zero by coupling z and u so that the coefficient of the (unknown) quantity $F(z,u)$ is zero. This constraint defines u as one of the branches of an algebraic function of z. If enough branches can be substituted analytically, then enough relations will be generated so that one can solve for the (unknown) quantities appearing on the right, namely, $F(z,1)$ and the $\partial_u^j F(z,0)$ that are then obtained as algebraic functions. From there, an expression for $F(z,u)$ also results in the form of a bivariate algebraic function.

One defines here the kernel K as

$$K(u,z) := -u^c(1-u)\left(1 + \frac{z}{1-u} - zP(u)\right), \tag{12}$$

which is nothing but the numerator of the coefficient of $F(z,u)$ in (11). There are $c+d+1$ solutions in u of this equation, which are algebraic functions of z. The classical theory of algebraic functions and the Newton polygon construction enable us to expand the solutions near any point as Puiseux series (that is, series involving fractional exponents). The $c+d+1$ solutions around 0 can be classified as follows:

- the "unit" branch, denoted by u_0, which tends to 1 as $z \to 0$;
- c "small" branches, denoted $u_1, \ldots, u_c$, which grow like $z^{1/c}$ at $z = 0$;
- d "large" branches, denoted by $v_1, \ldots, v_d$, which grow like $z^{-1/d}$ at $z = 0$;

In particular, there are exactly $c+1$ finite branches: the unit branch u_0 and the c small branches $u_1, \ldots, u_c$. An elementary argument shows that $F(z,1)$ is an analytic function of z at the origin, so that there are in total $c+1$ branches that can be substituted. Luckily, $c+1$ is the number of unkown quantities, $F(z,1)$ and $\partial_u^j F(z,0)$ on the right hand-side of (11).

Define the entire form of the right hand-side of (11),

$$Q(u,z) := -u^c(1-u)\left(1 + \frac{z}{1-u}F(z,1) - z\sum_{j=0}^{c-1} c_j(u)\partial_u^j F(z,0)\right).$$

The quantities K and Q are by construction polynomials in u. The roots $u_0, u_1, \ldots, u_c$ of K are also roots of Q which is monic with u-degree equal to $c+1$, so that Q admits the factorization:

$$Q(u,z) = \prod_{i=0}^{c}(u - u_i).$$

Let $l_d := [u^d]P(u)$ be the the multiplicity of the largest forward jump. One has similarly:

$$K(u,z) = -zl_d \prod_{i=0}^{c+d}(u - u_i).$$

Finally, the equation defining $F(z,u)$ is $K \cdot F(z,u) = Q$ and so that the factorizations above give

$$F(z,u) = \frac{Q(u,z)}{K(u,z)} = \frac{\prod_{i=0}^{c}(u-u_i)}{-zl_d\prod_{j=0}^{c+d}(u-u_j)} = \frac{1}{-zl_d\prod_{i=1}^{d}(u-v_i)}. \tag{13}$$

This specializes to give $F(z,1)$ which is the generating function of all walks taken irrespective of the value of their end point.

[1]The kernel method belongs to mathematical folklore since the 1970's; *e.g.*, it has been used by combinatorialists [8, 14] and probabilists [11]. There is also some recent work which makes a deep use of it [6, 7, 15].

Proposition 5 *A factorial walk, hence also a factorial system of generating trees, has an algebraic generating function. In particular, the generating function for all walks is*

$$F(z,1) = -\frac{1}{z}\prod_{i=0}^{c}(1-u_i),$$

where the product is over all branches $u_0, \ldots, u_c$ finite at $z=0$ of the algebraic function given by the equation $K(u,z)=0$, the kernel K being defined by (12).

The kernel method can also be applied (with some subtleties) to slightly more general systems, where backward steps leading to a fixed finite subset C of points near the origin are forbidden. The system is then $(k) \rightsquigarrow \{0,\ldots,k-1\} \setminus [C \cup k-B] \ \cup \ k+A$ and the generating function is still algebraic. An example is the system $(k) \rightsquigarrow (0)(2)(4)(5)(6)\cdots(k-1)(k)(k+2)$.

Classically, one defines excursions by the constraint that their end point is 0. The excursion generating function is then found directly from (13). With $l_c = [u^c]P(u)$, one has:

$$F(z,0) = \frac{(-1)^{c+1}}{l_c z}\prod_{i=0}^{c} u_i.$$

Proposition 5 was first obtained in March 1998 (see [1]), independently of [7, 15] to which the present treatment is closely related.

Example 5. *Catalan numbers.* This is the simplest factorial walk, $(k) \rightsquigarrow (0)(1)\ldots(k)(k+1)$, which corresponds to System (f) of Example 2. The characteristic operator is:

$$L[f](u) = \frac{f(1)-f(u)}{1-u} + (1+u)f(u).$$

The kernel is $K(u,z) = -(1-u) - z + z(1-u)(1+u) = u - 1 - u^2 z$, hence $u_0(z) = \frac{1-\sqrt{1-4z}}{2z}$, so that

$$F(z,1) = \frac{1-u_0}{-z} = 1 + 2z + 5z^2 + 14z^3 + 42z^4 + 132z^5 + O(z^6),$$

the generating function of the Catalan numbers (sequence **M1459**[2]). This result could be expected, given the well-known relation between these walks and Łukasiewicz codes. □

Example 6. *Motzkin numbers.* This example, due to Pinzani *et al.*, is derived from the previous one by forbidding "forward" steps of size zero. The rule is then

$$(k) \rightsquigarrow (0)\cdots(k-1)(k+1).$$

The characteristic operator is

$$L[f](u) = \frac{f(1)-f(u)}{1-u} + uf(u);$$

The kernel is $K(u,z) = -(1-u) - z + z(1-u)u$, leading to

$$F(z,1) \quad = \quad \frac{1-z-\sqrt{1-2z-3z^2}}{2z^2} = 1 + z + 2z^2 + 4z^3 + 9z^4 + 21z^5 + O(z^6),$$

the generating function for Motzkin numbers (sequence **M1184**). □

Example 7. *Schröder numbers.* This example, presented by Pinzani *et al.*, corresponds combinatorially to $(k) \rightsquigarrow (0)\ldots(k-1)(k)(k+1)^2$. One finds from Proposition 5 that

$$F(z,1) = \frac{1-3z-\sqrt{1-6z+z^2}}{4z^2} = 1 + 3z + 11z^2 + 45z^3 + 197z^4 + \cdots.$$

The coefficients are the Schröder numbers (**M2898**: Schröder's second problem). A higher order generalization that appears to be new is presented in the table of at the end of this paper (Fig 2). □

[2]The numbers **Mxxxx** are identifiers of the sequences in *The Encyclopedia of Integer Sequences* [16].

The examples obtained so far are all quadratic. It is however clear from our treatment that algebraic functions of arbitrary degree can be obtained: it suffices that the set of "exceptions" around k have a span greater than 1. We list here a few more examples. Verification is easy given a computer algebra system that handles algebraic functions and elimination.

EXAMPLE 8. *Ternary trees, dissections of a polygon, and t-ary trees.* The system with axiom $(s_0) = (2)$ and rules

$$(k) \rightsquigarrow (3)(4)\cdots(k)(k+1)(k+2)$$

is equivalent to the walk

$$(k) \rightsquigarrow (0)(1)\cdots(k)(k+1)(k+2),$$

and leads to

$$F(z,1) = 1 + 2z + 7z^2 + 30z^3 + 143z^4 + 728z^8 + \cdots,$$

that is, ternary plane rooted trees where the root has exceptional degree 2. This corresponds to sequence **M1782**. If the axiom is taken to be $(s_0) = (3)$, we get the "tricatalan" numbers $\binom{3n}{n}/(2n+1)$, that is, sequence **M2926**, that counts ternary trees.

The "tetracatalan" numbers $\binom{4n}{n}/(3n+1)$ are obtained by the rule

$$(k) \rightsquigarrow (4)\cdots(k)(k+1)(k+2)(k+3),$$

and axiom (4). This is sequence **M3587** that starts as $1, 4, 22, 140, 969$ and is described as "dissections of a polygon".

More generally, the system with axiom (t) and production rules

$$(k) \rightsquigarrow (t)\cdots(k)(k+1)(k+2)\cdots(k+t-1)$$

yields the t-Catalan numbers, $\binom{tn}{n}/((t-1)n+1)$ that count t-ary trees. The basic generating function derived from the kernel method is defined by the familiar equation $y = 1 + zy^t$. □

4 The transcendental case

One possible reason for a system to give a transcendental series is the fact that its coefficients grow too fast, so that its radius of convergence is zero. This is the case for the last system of Example 2.

Proposition 6 *Consider a system such that:*

1. *only a finite number of the functions e_i's are bounded;*
2. *for all k, there exists a forward jump from k (i.e., $e_i(k) > k$ for some i).*

Then the (ordinary) generating function $F(z)$ has radius of convergence zero.

Sketch of Proof. It is easy to prove that the coefficients of $F(z)$ grow like a factorial. □

EXAMPLE 9. *Arrangements.* The system $(k) \rightsquigarrow (k)(k+1)^{k-1}$ with axiom $(s_0) = (2)$ generates the sequence that starts with $1, 2, 5, 16, 65, 326$ (**M1497**). It is not hard to see that the triangular array $f_{n,k}$ is given by the arrangement numbers $k!\binom{n}{k}$, so that the exponential generating function of the sequence is $e^z/(1-z)$. This system satisfies the conditions of Proposition 6; accordingly, one has $f_n \sim e\,n!$, so that the ordinary generating function has radius of convergence 0 and cannot be algebraic. □

Algebraic generating functions are strongly constrained in their algebraic structure (by a polynomial equation) as well as in their analytic structure (in terms of singularities and asymptotic behaviour). In particular, algebraic functions have a finite number of isolated singularities that are algebraic numbers with local asymptotic expansions that may involve only rational exponents. *A contrario*, a generating function that has infinitely many singularities (*e.g.*, a natural boundary) or that involves a transcendental element (*e.g.*, a logarithm) in a local asymptotic expansion is by necessity transcendental; see [12] for a discussion of such transcendence criteria. In the case of generating trees, this means that the presence of a condition involving a transcendental element is expected to lead to a transcendental generating function. An instance that we examine now is system (e) of Example 2 where the rules are modified at powers of 2.

EXAMPLE 10. *The Fredholm case.* Case (e) of Example (2) involves the "Fredholm series" $h(z) := \sum_{m\geq 1} z^{2^m}$, which is well-known to admit the unit circle as a natural boundary. (This can be seen by way of the functional equation $h(z) = z^2 + h(z^2)$, from which there results that $h(z)$ is infinite at all iterated square-roots of unity.) Then, the F_k's satisfy the following equations:

$$z + (z-1)F_2(z) + \left(\frac{z}{(1-z)^2} + \frac{h(z)}{z^2} - 1\right) F_3(z) = 0, \qquad zF_2(z) + \left(\frac{z}{1-z} - \frac{h(z)}{z^2}\right) F_3(z) = 0,$$
$$F_k(z) = z^{k-3} F_3(z) \qquad \text{for } k \geq 4.$$

Solving for F_2 and F_3, then summing ($F = F_2 + F_3/(1-z)$), we get

$$F(z) = \frac{z(1-z)^2 h(z)}{(1-2z)(1-z)^2 h(z) - z^4}.$$

Now, the functions $h(z)$ and $F(z)$ are rationally related, so that $F(z)$ is itself transcendental. Its radius of convergence is determined by the cancellation of the denominator: it is finite and nonzero; its value is easily determined numerically and found to be about 0.360102. □

In the transcendental case, one can also discuss the *holonomic* character of the generating function $F(z)$. (A series is said to be *holonomic*, or *D-finite* [17], if it satisfies a linear differential equation with polynomial coefficients in z.) Holonomic functions include algebraic functions, and have a finite number of singularities. Example 9 is holonomic, while Example 10 is not, as it has infinitely many singularities.

Amongst the simplest systems are those that involve moves from k of the form $k \pm 1$ and k. Such systems are naturally associated to continued fractions. Many of them lead to holonomic functions (of the Hermite, Laguerre, or arrangement type; see also Figure 2). However, despite their simplicity, the following two systems lead to nonholonomic generating functions.

EXAMPLE 11. *Stirling polynomials.* The system $[2, (k) \rightsquigarrow (k)^{k-1}(k+1)]$ gives rise to the Stirling numbers of the second kind $\left\{{n \atop k}\right\}$ (the number of ways one can group n objects into k nonempty subsets). The recursion $\left\{{n+1 \atop k}\right\} = \left\{{n \atop k-1}\right\} + k\left\{{n \atop k}\right\}$ entails that

$$\widetilde{F}(z,u) = \sum_{n\geq 0} \left(\sum_{k=1}^{n} \left\{{n \atop k}\right\} u^k\right) \frac{z^n}{n!} = \exp(u(\exp(z)-1)).$$

At $u = 1$, the exponential generating function $\sum f_n z^n/n!$ specializes to

$$\widetilde{F}(z,1) = \exp(\exp(z)-1)) = 1 + 2z + 5\frac{z^2}{2!} + 15\frac{z^3}{3!} + 52\frac{z^4}{4!} + 203\frac{z^5}{5!} + \ldots$$

the exponential generating function of the Bell numbers. This function is an entire function that is nonholonomic since its growth (a tower of two exponentials) is too large to be compatible with that at an irregular singular point of the solution to a differential equation with polynomial coefficients. Hence, $\widetilde{F}(z,1)$ as well as $F(z,1)$ are nonholonomic. □

EXAMPLE 12. *Bessel histories.* This is given by the system with axiom (2) and productions $(k) \rightsquigarrow (k-1)(k)^{k-2}(k+1)$, with the first rule $(1) \rightsquigarrow (2)$ adjusted for consistency of degrees in ecosystems. Consider the corresponding paths $[(0), (k) \rightsquigarrow (k-1)(k)^k(k+1)]$, with bivariate generating function $F(z,u)$. This generating function satisfies the functional differential equation

$$F(z,u)\left(1 - z - z(u+u^{-1})\right) - zu\frac{\partial}{\partial u}F(z,u) = 1 + z(1-u^{-1})F(z,0),$$

whose processing is not obvious. Instead, the classical combinatorial theory of continued fractions provides for a direct representation,

$$F(z,0) = \cfrac{1}{1 - z - \cfrac{z^2}{1 - z - \cfrac{z^2}{1 - 2z - \cfrac{z^2}{1 - 3z - \ddots}}}} = 1 + z + 2z^2 + 4z^3 + 9z^4 + \cdots,$$

System	Name	Id.	Generating Function
	Rational GF's		(ogf)
$(1),\ (k) \rightsquigarrow (k)^{k-1}((k \bmod 2)+1)$	Fibonacci	M0692	$\frac{1}{1-z-z^2}$
$(2),\ (k) \rightsquigarrow (2)^{k-1}(k+1)$	odd Fibonacci	M1439	$\frac{1-z}{1-3z+z^2}$
$(3),\ (k) \rightsquigarrow (2)^{k-1}(k+1)$	even Fibonacci	M2741	$\frac{1}{1-3z+z^2}$
	Algebraic GF's		
$(1),\ (k) \rightsquigarrow (1)\cdots(k-1)(k+1)$	Motzkin numbers	M1184	$\frac{1-z-\sqrt{1-2z-3z^2}}{2z}$
$(2),\ (k) \rightsquigarrow (2)\cdots(k)(k+1)$	Catalan numbers	M1459	$\frac{1-2z-\sqrt{1-4z}}{2z}$
$(3),\ (k) \rightsquigarrow (3)\cdots(k)(k+1)^2$	Schröder numbers	M2898	$\frac{1-5z+z^2-\sqrt{1-6z+z^2}}{4z}$
$(4),\ (k) \rightsquigarrow (4)\cdots(k)(k+1)^3$		M3556	$\frac{1-7z+z^3-\sqrt{1-8z+4z^2}}{6z}$
$(t+1),\ (k) \rightsquigarrow (t+1)\cdots(k)(k+1)^t$			$\frac{1-2tz-z+z^t-\sqrt{(1+z-tz)^2-4z}}{2tz}$
$(3),\ (k) \rightsquigarrow (3)\cdots(k+2)$	Ternary trees	M2926	equation: $F=1+zF^3$
$(4),\ (k) \rightsquigarrow (4)\cdots(k+3)$	Dissection of a polygon	M3587	equation: $F=1+zF^4$
$(t),\ (k) \rightsquigarrow (t)\cdots(k+t-1)$	t-ary trees		equation: $F=1+zF^t$
	Transcendental GF's		(egf)
$(2),\ (k) \rightsquigarrow (k-1)^{k-1}(k+1)$	Involutions	M1221	$e^{z+\frac{1}{2}z^2}$
$(2),\ (k) \rightsquigarrow (k-1)^{k-2}(k)(k+1)$	Switchboard problem	M1461	$e^{2z+\frac{1}{2}z^2}$
$(2),\ (k) \rightsquigarrow (k)(k+1)^{k-1}$	Arrangements	M1497	$e^z/(1-z)$
$(2),\ (k) \rightsquigarrow (k-1)^{k-2}(k+1)^2$	Bicolored involutions	M1648	e^{2z+z^2}
$(2),\ (k) \rightsquigarrow (k+1)^k$	Factorial numbers	M1675	$1/(1-z)$
$(2),\ (k) \rightsquigarrow (k+1)^{k-1}(k+2)$	Increasing subsequences	M1795	$e^{z/(1-z)}/(1-z)$
	Nonholonomic GF's		
$(2),\ (k) \rightsquigarrow (k)^{k-1}(k+1)$	Bell numbers	M1484	e^{e^z-1}
$(2),\ (k) \rightsquigarrow (k)^{k-2}(k+1)^2$	Values of Bell poly.	M1662	$e^{2(e^z-1)}$
$(2),\ (k) \rightsquigarrow (k-1)(k)^{k-2}(k+1)$	Bessel numbers	M1462	—

Figure 2: A catalog of some ecosystems of combinatorial interest.

in which only the first level is anomalous. Comparison with [13] shows that

$$F(z,0) = \frac{1}{1-z-z^2B(z)} \quad \text{where} \quad B(z) = 1+z+2z^2+5z^3+14z^4+43z^5+143z^6+\cdots$$

is the generating function of "Bessel numbers", that is, sequence **M1462**. From [13], we know that

$$1-z^2B(-z) \sim z\frac{J_{1/z-1}(2)}{J_1/z(2)},$$

with J_ν the Bessel J-function of order ν. It remains to check that $F(z,u)$ is nonholonomic. The fast increase of $[z^n]B(z)$ entails

$$[z^n]F(z,0) \sim [z^{n-2}]B(z),$$

and the known asymptotic form [13] of $[z^n]B(z)$ that is recognizably of nonholonomic type (see [20] for admissible types) entails in turn that $F(z,0)$ is nonholonomic. □

Conclusion. To conclude, we present in Fig. 2 a small catalog of rules defining generation trees that lead to sequences of combinatorial interest. Several examples are detailed in the paper; others are due to West [18, 19] or Barcucci, Del Lungo, Pergola, Pinzani [2, 3, 4, 5], or are folklore. Each of them is an instance of application of our criteria; the generating function entries correspond to ordinary generating functions (ogf's) in the rational and algebraic cases, to exponential generating functions (egf's) in the "transcendental" case. (Note, however, that our terminology catalogs as "transcendental" the sequence $n!$, though its *exponential* generating function is rational.) The last three examples of the table are nonholonomic.

Acknowledgements. We thank Elisa Pergola and Renzo Pinzani who presented us the problem we deal with in this paper. We are also very grateful for helpful discussions with Jean-Paul Allouche.

References

[1] C. Banderier. Combinatoire analytique : application aux marches aléatoires. *Mémoire de DEA, Université Paris VII*, 1998.

[2] E. Barcucci, A. Del Lungo, E. Pergola, and R. Pinzani. ECO: a methodology for the enumeration of combinatorial objects. *Preprint*, 1998.

[3] E. Barcucci, A. Del Lungo, E. Pergola, and R. Pinzani. From c_n to $n!$: permutations avoiding $s_j(j+1)(j+2)$. In *Proceedings of FPSAC'98*, pages 31–41, Toronto, 1998.

[4] E. Barcucci, A. Del Lungo, E. Pergola, and R. Pinzani. From Motzkin to Catalan permutations: a "discrete continuity". *Preprint*, 1998.

[5] E. Barcucci, A. Del Lungo, E. Pergola, and R. Pinzani. A methodology for plane tree enumeration. *Discrete Mathematics*, 180(1-3):45–64, 1998.

[6] M. Bousquet-Mélou. Multi-statistic enumeration of two-stack sortable permutations. *Electronic Journal of Combinatorics*, 5:R21, 1998.

[7] M. Bousquet-Mélou and M. Petkovšek. Linear recurrences with constant coefficients: the multivariate case. *Discrete Mathematics*, to appear.

[8] R. Cori and J. Richard. Énumération des graphes planaires à l'aide des séries formelles en variables non commutatives. *Discrete Mathematics*, 2:115–162, 1972.

[9] S. Dulucq, S. Gire, and O. Guibert. A combinatorial proof of J. West's conjecture. *Discrete Mathematics*, 187(1-3):71–96, 1998.

[10] S. Dulucq, S. Gire, and J. West. Permutations with forbidden subsequences and nonseparable planar maps. *Discrete Mathematics*, 153(1-3):85–103, 1996.

[11] G. Fayolle and R. Iasnogorodski. Solutions of functional equations arising in the analysis of two-server queueing models. In *Performance of computer systems*, pages 289–303. North-Holland, 1979.

[12] P. Flajolet. Analytic models and ambiguity of context–free languages. *Theoretical Computer Science*, 49:283–309, 1987.

[13] P. Flajolet and R. Schott. Non–overlapping partitions, continued fractions, Bessel functions and a divergent series. *European Journal of Combinatorics*, 11:421–432, 1990.

[14] D. E. Knuth. *The Art of Computer Programming. Vol. 1: Fundamental Algorithms.* Addison-Wesley, 1968. Exercises 4 and 11, section 2.2.1.

[15] M. Petkovšek. The irrational chess knight. In *Proceedings of FPSAC'98*, pages 513–522, Toronto, 1998.

[16] N. J. Sloane and S. Plouffe. *The Encyclopedia of Integer Sequences.* Academic Press Inc., New York, 1995.

[17] R. P. Stanley. Differentiably finite power series. *Electronic Journal of Combinatorics*, 1:175–188, 1980.

[18] J. West. Generating trees and the Catalan and Schröder numbers. *Discrete Mathematics*, 146:247–262, 1995.

[19] J. West. Generating trees and forbidden subsequences. *Discrete Mathematics*, 157:363–374, 1996.

[20] J. Wimp and D. Zeilberger. Resurrecting the asymptotics of linear recurrences. *Journal of Mathematical Analysis and Applications*, 111:162–176, 1985.

Cyril Banderier, Philippe Flajolet. Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay (France).

Mireille Bousquet-Mélou. LaBRI, Université Bordeaux 1, 351 cours de la Libération, F-33405 Talence Cedex (France).

Alain Denise, Dominique Gouyou-Beauchamps. LRI, Bat. 490, Université Paris-Sud XI, F-91405 Orsay Cedex (France).

Danièle Gardy. Université de Versailles/Saint-Quentin, Laboratoire PRISM, 45, avenue des Etats-Unis, F-78035 Versailles Cedex (France).