INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the Analysis of Linear Probing Hashing

Philippe Flajolet, Patricio Poblete, Alfredo Viola

N 3265
Septembre 1997

THÈME 2

Rapport de recherche

ISSN 0249-6399

Thème 2 — Génie logiciel et calcul symbolique
Projet Algorithmes

Rapport de recherche— Septembre 1997 — 24 pages

On the Analysis of Linear Probing Hashing

Philippe Flajolet, Patricio Poblete, Alfredo Viola

Abstract: This paper presents moment analyses and characterizations of limit distributions for the construction cost of hash table under the linear probing strategy. Two models are considered, that of full tables and that of sparse tables with a filling ratio strictly smaller than 1. For full tables, the construction cost has expectation $O(n^{3/2})$, the standard deviation is of the same order, and a limit law of the Airy type holds. (The Airy distribution is a semi-classical distribution that is defined in terms of the usual Airy functions or equivalently in terms of Bessel functions of indices $-\frac{1}{3}, \frac{2}{3}$.) For sparse tables, the construction cost has expectation $O(n)$, standard deviation $O(\sqrt{n})$, and a limit law of the Gaussian type. Combinatorial relations with other problems leading to Airy phenomena (like graph connectivity, tree inversions, tree path length, or area under excursions) are also briefly discussed.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

Sur l'analyse du hachage avec essais linéaires

Résumé : Cet article présente analyses de moments et caractérisations de lois limites pour le coût de construction de tables de hachage selon la stratégie dite d'adressage ouvert et essais linéaires. Deux modèles sont considérés, celui des tables pleines et celui des tables "éparses" dont le taux de remplisaage est strictement inférieur à 1. En ce qui concerne les tables pleines, le coût de construction présente une espérance en $O(n^{3/2})$, l'écart type est du même ordre de grandeur et une loi limite de type Airy prévaut. (La distribution d'Airy est une distribution semi-classique qui se caractérise en termes de fonctions d'Airy, c'est-à-dire en terme de fonctions de Bessel d'indices $-\frac{1}{3}, \frac{2}{3}$.) Quant aux tables éparses, leur coût de construction est $O(n)$ en moyenne, avec un écart type en $O(\sqrt{n})$, et il existe une loi limite de type Gaussien. Les relations avec d'autres problèmes combinatoires conduisant à des lois d'Airy (connectivité dans les graphes, longueur de cheminement dans les arbres, aire sous les excursions) sont aussi brièvement examinées dans cet article.

ON THE ANALYSIS OF LINEAR PROBING HASHING

PHILIPPE FLAJOLET, PATRICIO V. POBLETE, AND ALFREDO VIOLA

Dedicated to Don Knuth on the occasion of the 35th anniversary of his first analysis of an algorithm in 1962–1963.

Abstract. This paper presents moment analyses and characterizations of limit distributions for the construction cost of hash table under the linear probing strategy. Two models are considered, that of full tables and that of sparse tables with a filling ratio strictly smaller than 1. For full tables, the construction cost has expectation $O(n^{3/2})$, the standard deviation is of the same order, and a limit law of the Airy type holds. (The Airy distribution is a semi-classical distribution that is defined in terms of the usual Airy functions or equivalently in terms of Bessel functions of indices $-\frac{1}{3}, \frac{2}{3}$.) For sparse tables, the construction cost has expectation $O(n)$, standard deviation $O(\sqrt{n})$, and a limit law of the Gaussian type. Combinatorial relations with other problems leading to Airy phenomena (like graph connectivity, tree inversions, tree path length, or area under excursions) are also briefly discussed.

Introduction

Linear probing hashing, defined below, is certainly the simplest "in place" hashing algorithm [11, 20, 35].

> *A table of length m, $T[1\,..\,m]$ is set up, as well as a hash function h that maps keys from some domain to the interval $[1\,..\,m]$ of table addresses. A collection of n elements with $n \le m$ are entered sequentially into the table according to the following rule: Each element x is placed at the first unoccupied location starting from $h(x)$ in cyclic order, namely the first of $h(x), h(x)+1, \ldots, h(1), h(2), \ldots, h(x)-1$.*

For each element x that gets placed at some location y, the circular distance between y and $h(x)$ (that is, $y - h(x)$ if $h(x) \le y$, and $m + h(x) - y$ otherwise) is called its *displacement*. Displacement is both a measure of the cost of inserting x and of the cost of searching x in the table. *Total displacement* corresponding to a sequence of hashed values is the sum of the invidual displacements of elements. As it determines the *construction cost* of the table, we use both terms interchangeably.

We analyse here the total displacement $d_{m,n}$ of a table of length m (the number of table locations) and size n (the number of keys), under the assumption that all m^n hash sequences are equally likely. The problem has an equivalent formulation in terms of the classical *parking problem*, where the total displacement of cars from their intended base has exactly the same distribution as the construction cost of linear probing hashed tables as seen by a "cycle lemma" originally due to Knuth and presented in [20]. The basic version of linear probing hashing, as described above, is based on a first-come-first-serve (FCFS) policy; alternative priority rules

Date: September 28, 1997.

The work of Philippe Flajolet was supported in part by the Long Term Research Project *Alcom-IT* (# 20244) of the European Union. The work of Patricio Poblete was supported in part by FONDECYT(Chile) under grant 1960881. The work of Alfredo Viola was supported in part by proyecto BID-CONICYT 140/94 and proyecto CONICYT fondo Clemente Estable 2078/96.

exist (like last-come-first-serve or "Robin Hood"), but total displacement remains unchanged. Thus, our analysis also applies directly to such variants of the basic algorithm.

Linear probing hashing has been the object of intense study; see the table on results and the bibliography in [11, pp. 51-54]. The simplicity of the algorithms goes well with efficiency, at least when tables are not too much filled. However, despite the simplicity of the algorithm, some of the probabilistic phenomena involved are not quite easy to capture. In addition, there is also special value for these problems since the first analysis of algorithms ever performed by Knuth [17] in 1962–1963 was that of linear probing hashing. As Knuth indicates in many of his writings, the problem has had a strong influence on his scientific carreer[1].

Sparse tables, by which we mean tables with a filling ratio $\alpha = n/m$ strictly less than 1, tend to behave reasonably well. This has been known, in the average case at least, since Knuth's first analysis. We establish here that the construction cost of a sparse table has an average that is $O(n)$, a standard deviation that is $O(\sqrt{n})$, and we provide very precise estimates for these quantities. The expectation estimate agrees naturally with the known fact that a random search or insertion in a sparse table has expected cost $O(1)$. In addition, we precisely characterize the distribution of construction costs and prove that it is Gaussian in the asymptotic limit. Thus, for sparse tables, observed values of costs are highly likely to be extremely close to what the average case analysis predicts.

In contrast, *full* ($m = n$) or *almost full* ($m = n - 1$) tables are much less well-behaved. The construction cost is $O(n^{3/2})$ on average, a fact also consistent with Knuth's early analyses demonstrating that late insertions in a table that fills up tend to have a superlinear cost. We provide here precise estimates for the standard deviation which turns out to be of the same order as the mean, namely $O(n^{3/2})$, an indication of the fairly high dispersion of costs. In fact, the construction cost admits a limit distribution that is of the "Airy type", involving Airy functions, or equivalently Bessel functions of orders that are multiple of $\frac{1}{3}$.

The analysis starts with almost full tables (Section 2) that are the basic combinatorial objects. The combinatorial principle on which this paper rests is a binary tree-like decomposition of full tables. From this, a difference-differential equation is derived that is the key to the analysis (Lemma 2). Moments, in either exact or asymptotic form, are obtained by a "pumping" process akin to the analysis of other cumulative parameters of combinatorial structures. For instance, similar methods have been used in the investigation of limit distributions for path length in trees (Takacs [42, 41]), the comparison cost of quicksort (Hennequin [14]), the area under random walks (Louchard [26, 27]), as well as in moment analysis of other combinatorial structures [16].

Sparse tables (Section 3) are then treated as labelled products of (almost) full tables, so that the corresponding generating functions involve large powers. For moments, especially for the mean and variance, the analysis results rather directly from that of full tables. However, for the limiting distribution, a somewhat delicate perturbative analysis of saddle point integrals is needed in order to derive a Gaussian law by means of characteristic function estimates.

Globally, these results reinforce our confidence that linear probing represents an excellent tradeoff between algorithmic simplicity and efficiency, as long as the filling ratio is not too large, say less than 2/3 or 3/4. These conclusions also apply to linear probing sort [11, pp. 168–170] whose analysis is almost isomorphic to that of linear probing hashing.

From the methodological standpoint, linear probing connects to a wealth of interesting combinatorial and analytic problems. A primary rôle is played by the tree function first studied by Eisenstein and by the Ramanujan-Knuth Q-function whose major properties we briefly recall in Section 1. Regarding limit laws, the Airy distribution that surfaces in the case of full tables is also present in random trees (inversions and path length), in random graphs (the complexity

[1] See the footnote in [20, p. 529].

or excess parameter), and in random walks (area); we discuss briefly in Section 4 some of the "reasons" for this fact. The Gaussian law of sparse tables is an instance of a general combinatorial scheme of some generality: our methods actually demonstrate that it should be expected in most cases where one deals with an additive parameter on a random assembly of a large number of random components.

1. The tree function and the Q-functions

The main character in this paper is the tree function that is defined implicitly by $T(z) = ze^{T(z)}$ and appears originally in problems related with the counting of rooted labeled trees [8, 12, 29, 36, 44]. The Lagrange inversion theorem provides a number of related series expansion like

$$T(z) = \sum_{n\geq 1} \frac{n^{n-1}}{n!} z^n, \qquad T(z)^m = m \sum_{n\geq m} \frac{n^{n-m-1}}{n!} n^{\underline{m}} z^n, \tag{1}$$

where $a^{\underline{k}} = a(a-1)\cdots(a-k+1)$. Most generating functions in this paper involve rational fractions in $T(z)$ with denominators that are powers of $(1-T)^{-1}$. Lagrange inversion also provides

$$\frac{1}{1-T(z)} = 1 + \sum_{n=1}^{\infty} n^n \frac{z^n}{n!}. \tag{2}$$

The asymptotic form of coefficients of any rational function of T is also directly recovered by singularity analysis [7, 30]. Application of the method requires the singular expansion of $T(z)$, itself obtained from the implicit function theorem.

Lemma 1. *The function $T(z)$ has a dominant singularity at $z = 1/e$, and its singular expansion there is*

$$T(z) = 1 - \delta(z) + \frac{1}{3}\delta(z)^2 - \frac{11}{72}\delta(z)^3 + \frac{43}{540}\delta(z)^4 + O\left(\delta(z)^5\right). \tag{3}$$

where $\delta(z) = \sqrt{2}\sqrt{1-ez}$.

The Q-functions. In close association with the tree function is what Knuth has popularized under the name of the "Ramanujan Q-function". This function [1, 18, 19, 20, 36] and its close relatives play a central rôle in the analysis of many algorithms and data structures —hashing with linear probing [17, 20], union-find algorithms [24], interleaved memory [23], optimal caching [21], and random mappings [2, 6, 19], most notably. The Q-function is defined by

$$Q(n) = 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots,$$

or, in a way that is equivalent thanks to (1),

$$\log \frac{1}{1-T(z)} = \sum_{n\geq 0} Q(n) n^{n-1} \frac{z^n}{n!}. \tag{4}$$

Singularity analysis of the generating function yields immediately

$$Q(n) \sim \sqrt{\frac{\pi n}{2}} - \frac{1}{3} + \frac{1}{12}\sqrt{\frac{\pi}{2n}} - \frac{4}{135n} + \cdots. \tag{5}$$

An asymptotic series for $Q(n)$ was first derived by Ramanujan [1], and tight estimates are obtained in [4].

For the purpose of expressing the average-case analysis of sparse tables, Knuth has extended the Ramanujan Q-function as

$$Q_0(m,n) = \sum_{i\geq 0} \frac{n^{\underline{i}}}{m^i}, \tag{6}$$

so that $Q(n) = Q_0(n, n-1)$. From the definition, one has

$$\sum_{n=0}^{\infty} Q_0(m,n) m^n \frac{t^n}{n!} = \frac{e^{mt}}{1-t}. \tag{7}$$

Basic asymptotic approximations entail

$$\begin{aligned} Q_0(m, \alpha m - 1) \quad &= \quad \frac{1}{1-\alpha} - \frac{1}{(1-\alpha)^3} m^{-1} + \frac{2+\alpha}{(1-\alpha)^5} m^{-2} - \frac{\alpha^2 + 8\alpha + 6}{(1-\alpha)^7} m^{-3} \\ &\quad + \frac{\alpha^3 + 22\alpha^2 + 58\alpha + 24}{(1-\alpha)^9} m^{-4} + O(m^{-5}). \end{aligned} \tag{8}$$

See [32] for a general framework.

2. Full tables

Throughout this paper, we consider tables that have m locations (m is called the "length" of the table) and we let n denote the number of keys (the "size"). Clearly, the number of tables (the number of hash sequences) with length m and size n is m^n, and such a table has $m-n$ empty locations. By circular symmetry [20], for tables such that $m > n$, we may freely assume that one of the empty locations is the rightmost one,. *This assumption of a last empty location is made from now onwards.* When $n = m-1$, we say that such a table table is *almost full*. Since there are $m-n$ empty locations, then the probability of the rightmost cell being empty is $(m-n)/m$, and therefore there are $m^{n-1}(m-n)$ ways of creating such tables. In particular, the number of almost full tables is $m^{m-2} = (n+1)^{n-1}$.

Inserting the last element into an almost full table yields a *completely full* table. Since this last element may hash to any of the m locations of the table, there are $m^{m-1} = n^{n-1}$ ways of creating a full table in this way. Thus, by our convention almost full and completely full tables don't "wrap around.". Clearly, the distributions of total displacements $d_{n,n-1}$ and $d_{n,n}$ are not affected by such a restriction.

Notations. The analysis is carried out by means of bivariate generating functions and moments are then recovered via a family of operators defined as follows. For any function $G(z,q)$,

$$\left\{ \begin{array}{lcllcl} \mathrm{U}\, G(z,q) & = & G(z,1) & \partial_q G(z,q) & = & \dfrac{\partial G(z,q)}{\partial q} \\ \mathrm{Z} G(z,q) & = & z G(z,q) & \partial_z G(z,q) & = & \dfrac{\partial G(z,q)}{\partial z} \\ \mathrm{H}\, G(z,q) & = & \dfrac{G(z,q) - qG(qz,q)}{1-q}. & & & \end{array} \right. \tag{9}$$

These operators act in the usual way on formal power series $G(z,q) = \sum_n g_n(q) \frac{z^n}{n!}$, with each $g_n(q)$ a polynomial; in particular,

$$\mathrm{H}\, G(z,q) = \sum_n g_n(q)(1 + q + q^2 + \cdots + q^n) \frac{z^n}{n!}.$$

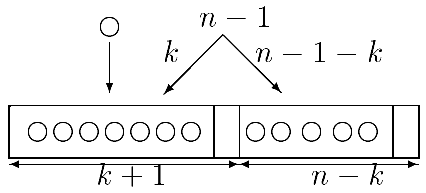

FIGURE 1. The binary tree decomposition of almost full tables.

Mike Paterson has designed an ingenious operator framework for the "local" analysis of displacements; see the account of the "cookie monster" in [13]. The problem of total displacement being fully history–dependent is however not clearly amenable to Paterson's techniques.

2.1. **Combinatorial analysis.** We define $F_{n,k}$ as the number of ways of creating an almost full table with n elements and total displacement k. The corresponding bivariate generating function is then

$$F(z,q) = \sum_{n,\,k\geq 0} F_{n,k}\; q^k \frac{z^n}{n!}.$$

and it starts like

$$F(z,q) = 1 + \frac{z}{1!} + (2+q)\frac{z^2}{2!} + (6+6q+3q^2+q^3)\frac{z^3}{3!} + \cdots,$$

Consider an almost full table of size n (and length $n+1$). Right before the last element is inserted, the table has two empty cells: one at some position $k+1$, the other at position $n+1$ (see Figure 1). Then, the element that is last to be inserted has an address that is any number of the interval $[1\,..\,k+1]$, which corresponds to a displacement that assumes any value in $[0\,..\,k]$. The counting of possibilities gives rise to a recurrence on the $F_n(q) = n![z^n]F(z,q)$:

$$F_n(q) = \sum_{k=0}^{n-1} \binom{n-1}{k} F_k(q)(1+q+\cdots+q^k)F_{n-1-k}(q). \tag{10}$$

This fundamental recurrence reflects a recursive binary decomposition of full tables. We recognize here a product of exponential generating functions modified by the occurrence of the H-operator defined in (9).

Lemma 2 (Basic functional equation).

$$\frac{\partial}{\partial z}F(z,q) = F(z,q)\cdot\frac{F(z,q)-qF(qz,q)}{1-q}. \tag{11}$$

In operator notation, this reads simply as $\partial_z F = F\cdot \mathrm{H}\,F$.

Let similarly $C_{n,k}$ be the number of completely full tables of size n, with $C(z,q) = \sum_{n,k} C_{n,k} q^k z^n/n!$ the corresponding bivariate generating function. Since a completely full table of size $n+1$ is created by inserting the last element in an almost full table of size n, we have from the definition of the H-operator ,

$$\partial_z C(z,q) = \mathrm{H}\,F(z,q).$$

Note that the basic functional equation together with this last relation implies the additional relations

$$F(z,q) = e^{C(z,q)} \quad \text{or} \quad C(z,q) = \log F(z,q). \tag{12}$$

Not surprisingly, the analyses of total displacement in full and in almost full tables are closely related.

2.2. **Moments.** For total displacement in almost full tables, what we call the generating function of rth factorial moments is by definition,

$$f_r(z) := \mathrm{U}\,\partial_q^r F(z,q) = \left.\frac{\partial^r}{\partial q^r} F(z,q)\right|_{q=1}. \tag{13}$$

This name is justified by the fact that the rth factorial moment of total displacement is given by

$$\mathbf{E}[d\,(d-1)\cdots(d-r+1)] = \frac{[z^n]f_r(z)}{[z^n]f_0(z)} \qquad \text{where} \quad d \equiv d_{n,n-1}.$$

The basic functional equation (11) implicitly contains all the information about moments. We now develop properties of the family of operators introduced in (9) designed to extract such moments explicitly.

First, let us rederive the enumeration of full tables. What is needed is $f_0(z) := \mathrm{U}\,F(z,q)$, where F is determined by (11). Now, from the action of H on power series, one has

$$\mathrm{U\,H} = \partial_z \mathrm{Z\,U} \quad \text{ or equivalently } \quad \mathrm{U\,H}\,F(z,q) = \frac{\partial}{\partial z}(zG(z,1)).$$

Thus, $f_0(z)$ satisfies the nonlinear differential equation obtained by applying U to (11):

$$Y'(z) = Y(z)(zY(z))'.$$

This equation is equivalent to $(\log Y(z))' = (zY(z))'$, and so $Y(z) = e^{zY(z)}$. In other words,

$$f_0(z) \equiv F(z,1) = \frac{1}{z}T(z) = e^{T(z)},$$

where $T(z)$ is the classical tree function. Therefore, by (1), the number of almost full tables is $(n+1)^{n-1}$. Similarly, by (12), $\mathrm{U}\,C(z,q) = \log(f_0(z)) = T(z)$ so that the number of completely full tables is n^{n-1}. These values in accordance with what we know already from direct combinatorial reasoning.

A similar device produces moments upon applying $\mathrm{U}\,\partial_q^r$ to the fundamental equation (11). What is needed is a "commutation rule" for $\mathrm{U}\,\partial_q^r$ and H. This is readily found for $r=1$ since

$$\mathrm{U}\,\partial_q\,\mathrm{H}(z^nq^k) = \mathrm{U}\,\partial_q((1+q+\cdots+q^n)z^nq^k) = ((1+2+\cdots+n)+(n+1)k)z^n.$$

Thus, symbolically

$$\mathrm{U}\,\partial_q\,\mathrm{H} = \frac{1}{2}\mathrm{Z}\partial_z^2 Z + \mathrm{U}\,\partial_z \mathrm{Z}\partial_q, \tag{14}$$

and by similar devices,

$$\mathrm{U}\,\partial_q^2\,\mathrm{H} = \partial_z \mathrm{Z\,U}\,\partial_q^2 + \mathrm{Z}\partial_z^2\mathrm{Z\,U}\,\partial_q + \frac{1}{3}\mathrm{Z}^2\partial_z^3\mathrm{Z\,U}\,. \tag{15}$$

As a consequence, $f_1(z)$ and $f_2(z)$ satisfy the following linear ordinary differential equations,

$$\mathcal{L}Y = \frac{1}{2}zf_0(zf_0)'' \tag{16}$$

$$\mathcal{L}Y = zf_1(zf_0)'' + 2f_1(zf_1)' + \frac{1}{3}z^2f_0(zf_0)''' + zf_0(zf_1)'', \tag{17}$$

where $\mathcal{L}$ is the differential operator

$$\begin{aligned}\mathcal{L}Y &= Y' \cdot (1 - zf_0) - Y \cdot ((zf_0)' + f_0) \\ &= Y' \cdot (1-T) - Y \cdot \frac{T(2-T)}{z(1-T)}. \end{aligned} \tag{18}$$

The corresponding homogeneous ordinary differential equation,

$$\mathcal{L}Y = 0$$

admits the solution,

$$Y(z) = \frac{e^{T(z)}}{1-T(z)}. \tag{19}$$

The variation–of–constant method then applies to the inhomogeneous differential equations (16) and (17) that are of the both form

$$\mathcal{L}Y(z) = R(z),$$

and yields the solution

$$Y(z) = \frac{e^{T(z)}}{1-T(z)} \int_0^z R(u) e^{-T(u)}\, du. \tag{20}$$

The quantities appearing in these differential equations can be expressed as functions of $T(z)$ alone since $z = Te^{-T}$ and $dz = (1-T)e^{-T}dT$. Thus the integrations needed in the variation-of-constant method all eventually reduce to integration of elementary functions for which decision procedures exist. We then obtain mechanically the generating functions of the first two moments for an almost full table. (This is for instance well within the capabilities of the computer algebra system Maple.)

Lemma 3 (Almost full tables, generating functions for the moments)**.**

$$\begin{aligned} zf_1(z) &= \frac{1}{2}\frac{T^3(z)}{(1-T(z))^2}, \\ zf_2(z) &= \frac{1}{12}\frac{T(z)^4(24 - 11T(z) + 2T(z)^2)}{(1-T(z))^5} \end{aligned}$$

For a completely full table, the corresponding generating functions result from Lemma 3 and equation (12):

$$\mathrm{U}\,\partial_q C(z,q) = \frac{f_1}{f_0} = \frac{1}{2}\frac{T^2}{(1-T)^2}, \tag{21}$$

$$\mathrm{U}\,\partial_q^2 C(z,q) = \frac{f_2 f_0 - f_1^2}{f_0^2} = \frac{1}{12}\frac{T^3(24 - 14T + 5T^2)}{(1-T)^5}. \tag{22}$$

Explicit expressions for the coefficients of functions appearing in (21) and (22) are then obtained from the expansions (2) and (4). Since $T(z)$ satisfies the differential relation

$$(\mathrm{Z}\partial_z)T(z) = \frac{T(z)}{1-T(z)},$$

the class of functions

$$\left\{ (\mathrm{Z}\partial_z)^r \frac{1}{1-T} \right\}_{r=0}^{\infty}, \qquad \left\{ (\mathrm{Z}\partial_z)^r \log\frac{1}{1-T} \right\}_{r=1}^{\infty},$$

spans a linear space that contains all the rational functions of the form $A(T)/(1-T)^r$, with A a polynomial of degree $< r$. As a consequence, for any such rational function of T, there exists an expansion

$$[z^n]\frac{A(T(z))}{(1-T(z))^r} = \frac{n^{n-1}}{n!}\left(U(n) + V(n)\,Q(n)\right), \tag{23}$$

for some polynomials U and V that can be mechanically determined.

Theorem 1 (Full tables, exact form of moments).

$$\begin{aligned}\mathbf{E}[d_{n,n}] &= \frac{n}{2}(Q(n)-1)\\ \mathbf{E}[d_{n,n}^2] &= \frac{n}{12}(5n^2+4n-1-8n\ Q(n)).\end{aligned}$$

Thanks to (3), singularity analysis applies directly to the solutions (21) and (22). (Alternatively, the explicit forms of Theorem 1 can be used in conjunction with (5).)

Theorem 2 (Full tables, asymptotic form of moments).

$$\begin{aligned}\mathbf{E}[d_{n,n}] &= \frac{\sqrt{2\pi}}{4}n^{3/2} - \frac{2}{3}n + \frac{\sqrt{2\pi}}{48}n^{1/2} - \frac{2}{135} + O(n^{-1}),\\ \mathbf{Var}[d_{n,n}] &= \frac{10-3\pi}{24}n^3 + \frac{16-3\pi}{144}n^2 + \frac{\sqrt{2\pi}}{135}n^{3/2} - \frac{\pi+48}{576}n + O(n^{1/2}).\end{aligned}$$

2.3. **Limit law.** Our goal in this subsection is to establish the existence of a limit distribution for the construction cost of almost full tables. As this limit distribution turns out not to be part of the set of classical continuous distributions, we first precisely specify it.

Definition 1. *The* Airy distribution *is the probability distribution of a random variable X with support on $[0,+\infty)$ that is uniquely determined by its moments,*

$$\mathbf{E}[X^r] = -\frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{3r-1}{2})}\,\Omega_r,$$

where the basic constants Ω_r are defined by the formal series expansion

$$\sum_{r\geq 0}\Omega_r\frac{w^r}{r!} = -\frac{\Phi_{2/3}(w)}{\Phi_{-1/3}(w)},$$

with

$$\begin{aligned}\Phi_\nu(w) &= 1-(4\nu^2-1)\left(\frac{w}{24}\right) + \frac{(4\nu^2-1)(4\nu^2-9)}{2!}\left(\frac{w}{24}\right)^2\\ &\quad - \frac{(4\nu^2-1)(4\nu^2-9)(4\nu^2-25)}{3!}\left(\frac{w}{24}\right)^3 + \cdots.\end{aligned}$$

Under various guises, the Airy distribution arises as a limit distribution in quite diverse contexts. Examples include the area under nonnegative random walks (Louchard [26, 27]) or path length in random trees (Takacs [42, 41]); this limit law also relates to asymptotic estimates of connectivity in random graphs (Wright [46], Janson *et al.* [15]). At the end of this paper, we comment briefly on the combinatorics that underlies some of these connections. Our derivation here follows in spirit the approach of Louchard and Takacs [26, 27, 42, 41] who also justified that the Airy distribution as defined here is indeed uniquely characterized by its moments.

We now examine in detail the process that yields the moments asymptotically and show how the Airy distribution arises from a recurrent determination of moments.

First, a process similar to the one employed for the first two moments yields a general commutation rule for the H and ∂_q operators

Lemma 4.

$$\mathrm{U}\,\partial_q^j\,\mathrm{H} = \sum_{s=0}^{j} \binom{j}{s} \frac{1}{s+1} \mathrm{Z}^s \partial_z^{s+1} \mathrm{Z}\,\mathrm{U}\,\partial_q^{j-s}. \tag{24}$$

Proof. The left hand side applied to $z^n q^k$ gives

$$\mathrm{U}\,\partial_q^j\,\mathrm{H}(z^n q^k) = z^n\,\mathrm{U}\,\partial_q^j \left(\sum_{i=0}^{n} q^i q^k \right).$$

Then, the Leibniz rule applied to the differentiation of products $q^i q^k$ yields

$$\begin{aligned} z^n\,\mathrm{U}\,\partial_q^j \left(\sum_{i=0}^{n} q^i q^k \right) &= z^n \sum_{i=0}^{n} \sum_{s=0}^{j} \binom{j}{s} i^{\underline{s}}\, k^{\underline{j-s}} = z^n \sum_{s=0}^{j} \binom{j}{s} \frac{1}{s+1}\, (n+1)^{\underline{s+1}}\, k^{\underline{j-s}} \\ &= \sum_{s=0}^{j} \binom{j}{s} \frac{1}{s+1} \mathrm{Z}^s \partial_z^{s+1} \mathrm{Z}\,\mathrm{U}\,\partial_q^{j-s} (z^n q^k). \end{aligned}$$

□

Then, a differential equation for the rth factorial moment generating function f_r is directly obtained by a combination of the Leibniz's rule and of the commutation relation (24) applied to the fundamental equation (11):

$$\partial_z f_r(z) = \sum_{j=0}^{r} \sum_{s=0}^{j} \binom{r}{j} \binom{j}{s} \frac{1}{s+1} f_{r-j}(z) \cdot \left(\mathrm{Z}^s \partial_z^{s+1} \mathrm{Z} f_{j-s}(z) \right). \tag{25}$$

The differential equation (25) that gives access to the rth moment is of the form

$$\mathcal{L} Y(z) = R_r(z),$$

where $\mathcal{L}$ is the linear differential operator of (18). There, $R_r(z)$ is exactly the right hand side of (25) stripped of its terms than contain $f_r(z), f_r'(z)$, namely the terms corresponding to $(j,s) = (0,0)$ and $(j,s) = (r,0)$. By the variation-of-constant method, moments can then be pumped *ad libidinem*, and we have from (20)

$$f_r(z) = \frac{e^{T(z)}}{1 - T(z)} \int_0^z R_r(u) e^{-T(u)}\, du. \tag{26}$$

For instance, we obtain automatically

$$\begin{aligned} z f_1 &= \frac{T^3}{2} \frac{1}{(1-T)^2}, \qquad z f_2 = \frac{T^4}{12} \frac{24 - 11\,T + 2\,T^2}{(1-T)^5} \\ z f_3 &= \frac{T^4}{8} \frac{8 + 144\,T - 110\,T^2 + 63\,T^3 - 17\,T^4 + 2\,T^5}{(1-T)^8} \\ z f_4 &= \frac{T^5}{240} \frac{10800 + 64560\,T - 60072\,T^2 + 53760\,T^3 - 26865\,T^4 + 9140\,T^5 - 1750\,T^6 + 152\,T^7}{(1-T)^{11}} \end{aligned} \tag{27}$$

The success of the pumping method is obvious as regards asymptotic forms at least since conditions of singularity analysis are preserved under multiplication by rational functions of T and under integration. (In fact, there always exist exact rational forms in T, as shown by a more sophisticated argument, but this is immaterial here.) An asymptotic pattern clearly emerges,

$$z f_1 \sim \frac{1}{2} \frac{1}{(1-T)^2},\ z f_2 \sim \frac{5}{4} \frac{1}{(1-T)^5},\ z f_3 \sim \frac{45}{4} \frac{1}{(1-T)^8},\ z f_4 \sim \frac{3315}{16} \frac{1}{(1-T)^{11}},$$

where the approximations hold when $z \to e^{-1}$, that is to say $T \to 1$. The following lemma precisely characterizes the dominant terms of $f_r(z)$.

Lemma 5. *The factorial moment generating functions satisfy, for* $r \geq 1$,

$$zf_r(z) \sim \frac{C_r}{(1-T(z))^{3r-1}} \sim \frac{C_r}{(2(1-ez))^{3r/2-1/2}} \qquad (z \to e^{-1}), \tag{28}$$

where the constants C_r *are determined by the nonlinear recurrence*

$$(3r-4)rC_{r-1} + \sum_{j=0}^{r} \binom{r}{j} C_j C_{r-j} - \delta_{r,0} = 0, \qquad C_0 = -1. \tag{29}$$

Proof. The property holds for $r = 1, 2$ by Lemma 3. For general r, the variation-of-constant formula (26) entails (by induction) that the singular behaviour f_r is of the form $zf_r \sim C_r(1-T)^{-3r+1}$ as $z \to e^{-1}$. In other words, the ∂_z operator shifts a singular expansion by a factor of $(1-T)^{-2}$ while the ∂_q operator shifts such an expansion by a factor of $(1-T)^{-3}$.

Thus, the dominant contribution from (25) arises from the terms corresponding to $s = 0$ and $(j, s) = (r, 1)$. As a consequence we have, as $z \to e^{-1}$,

$$\partial_z f_r(1 - zf_0) - f_r \partial_z(zf_0) = \frac{r}{2}\partial_z^2(zf_{r-1}) + \sum_{j=1}^{r-1} \binom{r}{j} f_{r-j}\partial_z(zf_j) + O((1-T)^{-3r+1}). \tag{30}$$

Integration and multiplication of both sides by $2z$ yields the asymptotic relation

$$2zf_r(1-zf_0) = rz\partial_z(zf_{r-1}) + \sum_{j=1}^{r-1} \binom{r}{j} (zf_{r-j})(zf_j) + O((1-T)^{-3r+3}). \tag{31}$$

The coefficients of the dominant terms involving $(1-T)^{-3r+2}$ can then be identified, and this provides a recursive determination of the coefficients C_r:

$$2C_r = (3r-4)rC_{r-1} + \sum_{j=1}^{r-1} \binom{r}{j} C_j C_{r-j} \quad r \geq 1.$$

There, by a natural convention, we take $C_0 = -1$ since $f_0 = 1 - (1-T)$ and it is singular components that count. This recurrence is equivalent to the one stated in (29). ☐

The constants C_r determine the dominant asymptotic form of the *moments* of the law of total displacement. Clearly, factorial moments and power moments are asymptotically equivalent, and, by singularity analysis, one has

$$\mu_n^{(r)} \equiv \mathbf{E}[(d_{n,n-1})^r] = \frac{2\sqrt{\pi}}{\Gamma(\frac{3r-1}{2})} C_r \left(\frac{n}{2}\right)^{3r/2} \left(1 + O(n^{-1/2})\right). \tag{32}$$

In order to establish the Airy limit distribution property, it is then necessary to identify the coefficients in (32). We are going to show that in fact $C_r = \Omega_r$, with the Ω_r the fundamental constants of Definition 1.

From (29), the quantities $\gamma_r := C_r/r!$ satisfy a nonlinear recurrence

$$(3r-4)\gamma_{r-1} + \sum_{j=0}^{r} \gamma_j\gamma_{r-j} - \delta_{r,0} = 0, \tag{33}$$

so that the exponential generating function of the C_r, $\gamma(z) := \sum_{r \geq 0} C_r z^r/r!$, itself satisfies a nonlinear first order ODE of the Riccati type.

$$3z^2\gamma'(z) - z\gamma(z) + \gamma(z)^2 - 1 = 0. \tag{34}$$

r	0	1	2	3	4	5	6	7	8
Ω_r	-1	$\frac{1}{2}$	$\frac{5}{4}$	$\frac{45}{4}$	$\frac{3315}{16}$	$\frac{25425}{4}$	$\frac{18635625}{64}$	$\frac{18592875}{1}$	$\frac{1282031525}{32768}$
ω_r	-1	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{15}{8}$	$\frac{1105}{128}$	$\frac{1695}{32}$	$\frac{414125}{1024}$	$\frac{59025}{16}$	$\frac{1282031525}{32768}$
$\omega_r^\star$	$-\frac{1}{2}$	1	5	60	1105	27120	828250	30220800	1282031525
$\mu^{(r)}$	1	$\sqrt{\pi}$	$\frac{10}{3}$	$\frac{15}{4}\sqrt{\pi}$	$\frac{884}{63}$	$\frac{565}{32}\sqrt{\pi}$	$\frac{662600}{9009}$	$\frac{19675}{192}\sqrt{\pi}$	$\frac{4102500880}{8729721}$

TABLE 1. The Airy constants Ω_r and their various normalizations: $\omega_r = \Omega_r/r!$, $\omega_r^\star = 2^{2r-1}\Omega_r/r!$, $\mu^{(r)} = -\Omega_r\Gamma(-1/2)/\Gamma((3r-1)/2)$. (The $\mu^{(r)}$ are the moments of the Airy distribution.)

In a way, this basic equation is a "reduced image" of the fundamental difference-differential equation when only dominant singular parts are retained. Now, it is known that Riccati equation are reducible to linear second order ODE's: set $\gamma(z) = 3z^2 g'(z)/g(z)$, so that

$$9z^4 g''(z) + 15z^3 g'(z) - g(z) = 0. \tag{35}$$

From there, the connection with Bessel functions is easy to establish and a computer algebra like Maple provides valuable hints. Some care is however needed due to the multivalued character of Bessel functions of nonintegral order, so that we provide some detail.

The "modified" Bessel functions are defined by

$$\left\{ \begin{array}{lcl} I_\nu(z) & = & \left(\frac{z}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!\Gamma(\nu+k+1)} \\ K_\nu(z) & = & \frac{\pi}{2\sin\nu\pi}\left(I_{-\nu}(z) - I_\nu(z)\right), \end{array} \right. \tag{36}$$

and for nonintegral ν, they form of basis of solutions to the Bessel equation

$$z^2 \frac{d^2 w}{dz^2} + z\frac{dw}{dz} - (z^2+\nu^2)w = 0. \tag{37}$$

One can then simply match Equations (35) with (37) and verify that the general solution to (35) is

$$g(z) = z^{-1/3}\left(\lambda_1 K_{-1/3}(\frac{1}{3z}) + \lambda_2 I_{-1/3}(\frac{1}{3z})\right).$$

A simple computation then shows that the general solution of the original Riccati equation is

$$\gamma^{\langle\lambda\rangle}(z) = -\frac{I_{2/3}(\frac{1}{3z}) - \lambda K_{2/3}(\frac{1}{3z})}{I_{-1/3}(\frac{1}{3z}) + \lambda K_{-1/3}(\frac{1}{3z})} \tag{38}$$

For determinacy, we restrict (38) to the complex z-plane slit along $(-\infty, 0)$.

We note at this stage that Bessel functions of order a multiple of $1/3$ are related to the classical Airy functions that are defined as solutions to the linear differential equation $w'' - zw = 0$. In particular, one has

$$\begin{array}{rcl} \mathrm{Ai}(z) & = & \frac{1}{\pi}\int_0^\infty \cos\left(\frac{1}{3}t^3 + zt\right) dt \\ & = & \frac{1}{\pi}\left(\frac{z}{3}\right)^{1/2} K_{1/3}\left(\frac{2z^{3/2}}{3}\right). \end{array}$$

This (and other connections) justify our choice of naming the distribution of Definition 1 the Airy distribution.

Obviously, the $\gamma^{\langle\lambda\rangle}(z)$ as obtained in (38) are nonanalytic at 0. Then, Eq. (38) is to be taken in the sense that the divergent (formal) series $\gamma^{\langle\lambda\rangle}(z)$ represents asymptotically the right hand side as $z \to 0^+$. But, asymptotic expansions of Bessel functions are well-known: with $\mu = 4\nu^2$, we have as the variable z tends to $+\infty$,

$$I_\nu(z) \sim \frac{e^z}{\sqrt{2\pi z}}\left(1 - \frac{\mu-1}{8z} + \frac{(\mu-1)(\mu-9)}{2!(8z)^2} - \frac{(\mu-1)(\mu-9)(\mu-25)}{3!(8z)^3} + \cdots\right), \tag{39}$$

while each $K_\nu(z) = O(e^{-z})$ is exponentially small. Thus, the asymptotic expansions of all the $\gamma^{\langle\lambda\rangle}(z)$ in the scale $\{z^{-m}\}$ coincide, and we may as well take $\gamma(z) = \gamma^{\langle 0\rangle}(z)$. In other words, the C_r are generated as coefficients in the asymptotic expansion,

$$-\frac{I_{2/3}(\frac{1}{3z})}{I_{-1/3}(\frac{1}{3z})} \sim \sum_{r=0}^{\infty} C_r \frac{z^r}{r!} \qquad (z \to 0^+). \tag{40}$$

From (39) and (40), we thus obtain a purely algebraic and explicit specification of $\gamma(z)$ as a quotient of two divergent hypergeometric series (of the ${}_2F_0$ type) that matches exactly the definition of the Airy distribution, with $C_r = \Omega_r$. This characterizes the distribution of construction cost in almost full hash tables.

Theorem 3 (Limit law for full tables). *For almost full tables, the distribution of the random variable* $\frac{d_{n,n-1}}{(n/2)^{3/2}}$ *converges to the Airy distribution, in the sense that, pointwise for each* x,

$$\Pr\{\frac{d_{n,n-1}}{(n/2)^{3/2}} \le x\} \to \Pr\{X \le x\} \qquad (n \to \infty),$$

where X *is Airy distributed in the sense of Definition 1. The same property holds for completely full tables and the random variable* $\frac{d_{n,n}}{(n/2)^{3/2}}$.

(The property for full tables results from the fact that $d_{n,n}$ has the same distribution as $d_{n,n-1} + \mathcal{U}_n$, where $\mathcal{U}_n$ is uniform over $[0\,..\,n-1]$.)

Initial values of the Airy constants are given in Table 1. The normalized constants $\omega_r^\star = 2^{2r-1}\Omega_r/r!$ turn out to be integers for $r \ge 1$. This interesting sequence starts like

1, 5, 60, 1105, 27120, 828250, 30220800, 1282031525, 61999046400, 3366961243750, 202903221120000, 13437880555850250, 970217083619328000, 75849500508999712500, 6383483988812390400000, 575440151532675686278125, 55318762960656722780160000,

and we propose to call it the Wright–Louchard-Takacs sequence (see the remarks above and our conclusion). It is however *not* to be found in Sloane and Plouffe's *Encyclopedia of Integer Sequences* [38]. The variance of the Airy distribution is

$$\frac{10}{3} - \pi = 0.19174\,06797\,43540\ldots$$

and the appearance of this magic value in a variance expression may be taken as a good indication of the possible occurrence of the Airy distribution.

3. Sparse Tables

In this section, we analyse sparse tables, where the filling ratio defined as $\alpha = n/m$ is bounded away from 1. The behaviour of such tables turns out to be much more tame than that of full tables discussed in the previous section.

3.1. **Combinatorial analysis.** As seen at the beginning of Section 2, a simple circular symmetry argument enables us to restrict attention to tables whose last location is empty. Such a table then decomposes as a labelled product of $m-n$ clusters (sometimes also figuratively called "islands") that are, up to relabelling, almost full tables. Note that a cluster may well have size 0, in which case it comprises only one unoccupied cell. For instance, the table

	3	9	4	7			5	2		8		1	6				

is, up to relabelling, a sequence of six almost full tables of respective sizes 0,4,0,2,1,2.

Define the generating function $H_{m,n}(q)$ that counts the number of ways of creating a non-full table of lenght m and size n (the rightmost location is empty) with q marking the total displacement. The construction cost (or total displacement) of partial tables is inherited additively from component clusters. Therefore, the total displacement in partial tables of parameter (m,n) has generating function

$$H_{m,n}(q) = n![z^n]F(z,q)^{m-n}.$$

The number of tables of length m, size n, with the last location empty is then

$$H_{m,n}(1) = n![z^n]f_0(z)^{m-n} = n![z^m]T(z)^{m-n},$$

a quantity that, by virtue of (1), equals $(m-n)m^{n-1}$, in agreement with the circular symmetry argument. The probability generating function of the total displacement $d_{m,n}$ is then

$$\frac{H_{m,n}(q)}{H_{m,n}(1)} = \frac{n!}{(m-n)m^{n-1}}[z^n]F(z,q)^{m-n}.$$

3.2. **Moments.** The generating functions for sparse tables admit power forms that lend themselves nicely to differentiation. In this way, moment generating functions are obtained immediately from the corresponding computation for full tables.

The analysis still relies on the functions $f_r = \mathrm{U}\,{\partial_q}^r F$ introduced in (13). We have

$$\begin{aligned} \mathrm{U}\,\partial_q F(z,q)^{m-n} &= (m-n)f_0^{m-n-1}f_1 \\ \mathrm{U}\,{\partial_q}^2 F(z,q)^{m-n} &= (m-n)(m-n-1)f_0^{m-n-2}f_1^2 + (m-n)f_0^{m-n-1}f_2. \end{aligned}$$

The values of zf_0, zf_1, zf_2 are known from Section 2, and are expressible in terms of $T = T(z)$ alone; this gives for instance,

$$\frac{n!}{(m-n)m^{n-1}}[z^n]\,\mathrm{U}\,\partial_q F(z,q)^{m-n} = \frac{n!}{m^{n-1}}[z^m]\frac{1}{2}\,\frac{T(z)^{m-n+2}}{(1-T(z))^2},$$

What is required at this point in order to obtain explicit forms is a method for coefficient extraction,

$$[z^m]\,T(z)^{m-n}\,\frac{A(T(z))}{(1-T(z))^r}, \tag{41}$$

where A is a polynomial of degree $< r$. For computational purposes, it is convenient to introduce the change of variables in Cauchy coefficient integrals that underlies Lagrange inversion.

$$\begin{aligned} [z^m]\lambda(T(z)) &= \frac{1}{2i\pi}\int \lambda(T(z))\frac{dz}{z^{m+1}} \\ &= \frac{1}{2i\pi}\int \lambda(T)(1-T)e^{-T}\,\frac{dT}{(Te^{-T})^{m+1}} \\ &= [t^n]e^{mt}(1-t)\lambda(t). \end{aligned} \tag{42}$$

(Small contours around 0 are understood in this derivation, and this shortcut is of course logically equivalent to Lagrange-Bürmann inversion.)

Then, the application of (42) to (41) yields

$$[z^m]T^{m-n}\,\frac{A(T)}{(1-T)^r} = [t^n]e^{mt}\,\frac{A(t)}{(1-t)^{r-1}}.$$

This is close to the form (7) of the generating function of $Q_0(m,n)$. Now, an argument similar to the one used in (23) for full tables applies. The linear space spanned by

$$\left\{\left(t\frac{d}{dt}\right)^r \frac{e^{mt}}{1-t}\right\}_{r=0}^{\infty} \cup \left\{e^{mt}\right\}$$

contains all the rational functions of the form $e^{mt}A(t)/(1-t)^{r-1}$. Thus, there exist polynomials U and V such that

$$[z^m]T^{m-n}\,\frac{A(T)}{(1-T)^r} = \frac{m^n}{n!}\left(U(m,n) + V(m,n)\,Q_0(m,n)\right).$$

The computation is again purely mechanical. It can be recast in terms of $Q_0(m,n-1)$ since $Q_0(m,n) = 1 + \frac{m}{n}Q_0(m,n-1)$.

Theorem 4 (Sparse tables, exact form of moments)**.**

$$\begin{aligned}
\mathbf{E}[d_{m,n}] &= \frac{n}{2}(Q_0(m,n-1)-1),\\
\mathbf{E}[d_{m,n}^2] &= \frac{n}{12}\big((m-n)^3 + (n+3)(m-n)^2 + (8n+1)(m-n) + 5n^2 + 4n - 1\\
&\qquad -((m-n)^3 + 4(m-n)^2 + (6n+3)(m-n) + 8n)Q_0(m,n-1)\big).
\end{aligned}$$

The approximation formula (8) then produces the asymptotic form of the first moments and the variance of an α-sparse table.

Theorem 5 (Sparse tables, asymptotic form of moments)**.**

$$\begin{aligned}
\mathbf{E}[d_{m,n}] &= \frac{\alpha}{2(1-\alpha)}n - \frac{\alpha}{2(1-\alpha)^3} + O(n^{-1}),\\
\mathbf{Var}[d_{m,n}] &= \frac{6\alpha - 6\alpha^2 + 4\alpha^3 - \alpha^4}{12(1-\alpha)^4}n - \frac{6\alpha^3 + 24\alpha^2 + 6\alpha}{12(1-\alpha)^6} + O(n^{-1}).
\end{aligned}$$

3.3. **Limit law.** In this subsection, we estimate the distribution of total displacement in sparse tables, when m, n tend to infinity in such a way that the filling ration $\alpha = n/m$ remains constant. We thus fix throughout α and assume $0 < \alpha < 1$ The mean μ_n and the variance σ_n^2 of the distribution are in this case both $O(n)$ and their precise form has been given by the last two theorems.

The limit law is approached here by characteristic functions rather than by moments as was done in the case of full tables. Indeed, cancellations already present in the variance preclude a moment approach. On the other hand, the power form of the involved generating functions suggests an appeal to the saddle point method applied to Cauchy coefficient integrals, this in order to estimate characteristic functions. Some care is however needed since $F(z,q)$ is sharply nonanalytic at $q = 1$. The analysis proceeds by a (delicate) perturbation of the (easy) saddle point estimates of the univariate problem of counting sparse tables, namely $[z^n]F(z,1)^{m-n}$.

Theorem 6 (Limit law for sparse tables)**.** *The limit law of total displacement $d_{m,n}$ in tables with filling ratio $\alpha = \frac{n}{m}$ that satisfies $\alpha < 1$ is asymptotically Gaussian, as $n \to \infty$,*

$$\Pr\left\{\frac{d_{m,n} - \mu_n}{\sigma_n} \le x\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-s^2/2}\,ds,$$

where $\mu_n = \mathbf{E}[d_{m,n}]$ is the mean of the distribution and σ_n defined by $\sigma_n^2 = \mathbf{Var}[d_{m,n}]$ is the standard deviation, as given by Theorems 4 and 5.

Proof. By Lévy's continuity theorem, it is sufficient to consider the characteristic function of the standardized distribution (centred around it mean and scaled by its standard deviation), that is,

$$\phi_n^\star(t) = \frac{1}{[z^n]F(z,1)^{m-n}} \left(e^{it\mu_n/\sigma_n} [z^n] F(z, e^{it/\sigma_n})^{m-n} \right), \tag{43}$$

and prove that it converges pointwise for any fixed t to the characteristic function of a standard normal variate,

$$\phi_n^\star(t) \to e^{-t^2/2}. \tag{44}$$

Since $\sigma_n = O(\sqrt{n})$, we analyse instead the closely related quantity

$$h_n(t) := [z^n]F(z, e^{it/\sqrt{n}})^{m-n} \quad \text{so that} \quad \frac{h_n(t)}{h_n(0)} = e^{-it\mu_n/\sqrt{n}} \phi^\star\left(\frac{\sigma_n t}{\sqrt{n}}\right). \tag{45}$$

The occurrence of large coefficients of large powers is known in the univariate case to be amenable to the saddle point method [3], and we start by briefly reviewing the case $t = 0$ that corresponds to the "unperturbed" integral,

$$[z^n]F(z,1) = \frac{1}{2i\pi} \int F(z,1)^{m-n} \frac{dz}{z^{n+1}}. \tag{46}$$

By a standard argument, such an integral (46) involving large powers is precisely of the type amenable to saddle point analysis. Here, we have $F(z,1) = f_0(z) = T(z)/z$, and the saddle point equation is

$$\frac{d}{dz}\left((m-n)f_0(z) - n\log z\right) = 0,$$

which has a unique positive root between 0 and e^{-1} at $\zeta = \alpha e^{-\alpha}$. At that point, one has additionally $T(\zeta) = \alpha$ and $f_0(\zeta) = e^\alpha$.

The classical saddle point analysis is based on integration on the circle $|z| = \zeta$ together with the fact that only a small sector of amplitude δ around ζ dictates the asymptotic contribution of the integral in (46). One should take $n\delta^2 \to \infty$ and $n\delta^3 \to 0$, for instance $\delta = n^{-0.4}$ is suitable, a choice that we fix here. Then, a local expansion reduces asymptotically and up to normalization the integral to be evaluated to a complete integral of $e^{-w^2/2}$.

Now, the strategy for evaluating the integral in (45) consists in adopting the same integration contour $|z| = \zeta$ as in the unperturbed case (46). The perturbation introduced in (45) by $q = e^{it/\sqrt{n}}$ must then be quantified precisely. It turns out that concentration in a sector of amplitude $\delta = n^{-0.4}$ still holds as the maximum of the integrand's modulus on the contour only gets displaced by a much smaller amount, namely $O(n^{-0.5})$. Local expansions near the real axis then provide the asymptotic form of $h_n(t)$, from which the Gaussian law eventually results.

First, we establish globally that the geometry of $F(z,q)$ on $|z| = \zeta$ does not differ much from that of $F(z,1)$ when $q = e^{i\theta}$ and θ lies in a suitably restricted interval around 0. The derivatives

$$f_r(z) = \left.\frac{\partial^r}{\partial q^r} F(z,q)\right|_{q=1},$$

exist as formal power series in z that are furthermore analytic in $|z| < e^{-1}$. Also, since the total displacement parameter on an object of size n is always at most n^2, we have

$$\left(q\frac{\partial}{\partial q}\right)^r F(z,q)\bigg|_{q=1} \ll \left(z\frac{\partial}{\partial z}\right)^{2r} F(z,1),$$

where $\ll$ indicates here coefficientwise dominance between power series with nonnegative coefficients. There results that $F(z,e^{i\theta})$ is in fact an infinitely differentiable function of θ for all fixed z inside the disc $|z| < e^{-1}$. (Construct formal derivatives whose analytic existence is guaranteed by the domination property and then recover $F(z,e^{i\theta})$ by repeated integration.) In particular, Taylor's formula with remainder, when aplied to $F(z,e^{i\theta})$, with z treated as a parameter, yields

$$F(z,e^{i\theta}) = f_0(z) + i\theta f_1(z) - \frac{\theta^2}{2}(f_2(z)+f_1(z)) + \frac{1}{3!}\int_0^\theta (\theta-u)^2 \frac{\partial^3}{\partial u^3}F(z,e^{iu})\,du. \tag{47}$$

The last term is $O(\theta^3)$ and this estimate holds uniformly with respect to z, for z in any subdisc of $|z| < e^{-1}$, since, by coefficient dominance again, the third partial derivative is dominated coefficientwise by $(z\frac{d}{dz})^6 f_0(z)$. The *uniform* estimate (47) precisely quantifies the way $F(z,e^{i\theta})$ approximates $F(z,1)$.

Next, along the circle $|z| = \zeta$, the quantity $|F(z,1)|$ has a unique maximum on the real axis at the saddle point $z = \zeta$. Also, $|F(\zeta e^{i\phi},1)|$ is an upward concave function of the argument ϕ in a fixed neighbourhood of $\phi = 0$. By the uniform approximation property (47) and the continuity that it implies, upward ϕ-concavity, that is expressed by a sign condition on second derivatives, must persist for $F(\zeta e^{i\phi}, e^{i\theta})$ provided θ stays in a sufficiently small neighbourhood of 0. Also, for values of ϕ outside of the guaranteed concavity interval and again θ suitably small, the approximation relation (47) entails that $|F(\zeta e^{i\phi},e^{i\theta})| < F(\zeta,1) - \epsilon$, for some fixed $\epsilon > 0$.

The preceding discussion thus provides a clear picture of $|F(z,e^{i\theta})|$ on the circle $|z| = \zeta$. When θ, now a parameter, is such that $|\theta|$ remains less than a small fixed nonzero threshold θ_0, the quantity $|F(\zeta e^{i\phi},e^{i\theta})|$ is upward concave near $\phi = 0$ (that is for z near the real axis) while its values at least remain boundedly smaller than the absolute maximum $f_0(\zeta)$, outside the concavity interval.

Take now $\theta = t/\sqrt{n}$, which is needed for estimating $h_n(t)$. The value of t is fixed and n is assumed to be large enough so that the local concavity and majorization properties hold. A local expansion shows that the maximum of $|F(\zeta e^{i\phi}, e^{it/\sqrt{n}})|$ occurs at $\phi = \phi_0(n)$, where

$$\phi_0(n) = -c_1\frac{t}{\sqrt{n}}(1+O(n^{-1})), \qquad c_1 = \frac{f_1(\zeta)}{\zeta f_0'(\zeta)}.$$

This is well within the range of the unperturbed saddle point integral which is given by the boundary points $\zeta e^{\pm i\delta}$, where $\delta = n^{-0.4}$. Therefore, we can conclude in the usual way that

$$h_n(t) = \frac{1}{2\pi}\int_{-\delta}^{+\delta} F(\zeta e^{i\phi}, e^{it/\sqrt{n}})^{m-n} e^{-ni\phi}\,d\phi\,(1+O(n^{-1/2})), \tag{48}$$

where the error term is in fact exponentially small.

Now, the analysis can be performed in the small interval $[-\delta,+\delta]$ by means of local expansions of the integrand, themselves attainable from the main approximation (47). For estimates up to relative $O(n^{-1/2})$ error terms, it suffices to use the quadratic approximation part of (47), so that

$$h_n(t) = \frac{1}{2\pi}\int_{-\delta}^{+\delta} f_0(\zeta e^{i\phi})^{m-n} A(\zeta e^{i\phi})^{m-n} e^{-ni\phi}\, d\phi\, (1 + O(n^{-1/2})), \tag{49}$$

where

$$A(z) = 1 + i\frac{t}{\sqrt{n}}\frac{f_1(z)}{f_0(z)} - \frac{t^2}{2}\frac{f_2(z)+f_1(z)}{f_0(z)}.$$

From this point on, the computations are routine but particularly tedious, so that we only sketch them. It suffices to expand $(m-n)\log A(\zeta e^{i\phi})$ with respect to ϕ up to quadratic terms again, then set $\phi = w/\sqrt{n}$, and extend the integration bounds to $(-\infty,+\infty)$. The integral is thereby reduced asymptotically to a form

$$\int_{-\infty}^{+\infty} \exp(a_0 + ia_1 w - a_2 w^2/2)\, dw = \sqrt{\frac{2\pi}{a_2}}\exp\left(a_0 - \frac{a_1^2}{2a_2}\right), \tag{50}$$

that is evaluated by completing the square. Once more the support of a computer algebra system like Maple is especially welcome, and one finds (some details omitted),

$$\begin{aligned}
a_0 &= \beta\log\frac{f_0(\zeta)}{\zeta}n + \frac{i\beta t f_1(\zeta)}{f_0(\zeta)}n^{1/2} - \frac{\beta t^2}{2}\left(\frac{f_2(\zeta)+f_1(\zeta)}{f_0(\zeta)} - \frac{f_1(\zeta)^2}{f_0(\zeta)^2}\right) + O(n^{-1/2})\\
a_1 &= \left(\frac{\beta f_0'(\zeta)}{f_0(\zeta)} - 1\right)n + \frac{it\beta\zeta}{f_0(\zeta)^2}(f_0(\zeta)f_1'(\zeta) - f_0'(\zeta)f_1(\zeta))n^{1/2} + O(1)\\
a_2 &= \frac{\beta\zeta}{2}\frac{\zeta f_0'(\zeta)^2 - \zeta f_0''(\zeta)f_0(\zeta) - f_0(\zeta)f_0'(\zeta)}{f_0(\zeta)^2}n + O(n^{1/2})
\end{aligned} \tag{51}$$

with $\beta = \alpha^{-1} - 1$.

All reductions done (!), we obtain from Equations (50) and (51) the asymptotic estimate

$$\frac{h_n(t)}{h_n(0)} = \exp\left(i\mu_n\frac{t\sigma_n}{\sqrt{n}} - \frac{t^2\sigma_n^2}{2n}\right)(1 + O(n^{-1/2})), \tag{52}$$

where use is made of the asymptotic forms of μ_n and σ_n.

We observe in passing (see also the comments below) that the asymptotic form of moments derives systematically from the basic saddle point method and that the expressions can be all obtained directly in terms of f_r and their derivatives evaluated at ζ. For instance,

$$\mu_n = (m-n)\frac{\int f_0(z)^{m-n-1}f_1(z)z^{-n-1}\,dz}{\int f_0(z)^{m-n}z^{-n-1}\,dz} \sim \beta n\frac{f_1(\zeta)}{f_0(\zeta)},$$

and so on.

The final estimate (52) after renormalization according to (45) then yields the convergence of characteristic functions (44). This completes the proof of the Gaussian limit law. □

The saddle point method has been used in a technically different context by Pittel [31] who showed that the size of the largest cluster (hence, also the maximum displacement) in a sparse linear probing table only grows logarithmically, on average and in probability. The process used in the proof of the last theorem is in fact very general and we encapsulate it into a statement.

Corollary 1. *A Gaussian limit law holds for the coefficients of any "large power",*

$$[z^n]G(z,q)^{\beta n}, \quad \beta > 0,$$

provided the following conditions hold:

(C_1) $G(z,q) = \sum_n g_n(q)z^n$ *has nonnegative coefficients and* $\deg g_n(q) = O(n^\kappa)$ *for some integer* κ*;*

(C_2) $G(z,1)$ *is analytic in* $|z| < r$, $G(0,1) \neq 0$, $G_z'(0,1) \neq 0$*;*

(C_3) $G_z'(r,1)/G(r,1) = +\infty$

Proof. (Sketch) Condition (C_1) ensures analyticity of partial derivatives and smooth perturbation; (C_3) ensures existence of the basic saddle point; (C_2) ensures unicity of this saddle point. It can be recognized that these are the only conditions used in the proof of Theorem 6, when one defines abstractly the functions f_r by $\mathrm{U}\,\partial_q^r G$ and the saddle point ζ by the equation $\beta\zeta f_0'(\zeta) - f_0(\zeta) = 0$. □

Given its mild analytic conditions, Corollary 1 applies to a diversity of situations where large random assemblies of labelled or unlabelled combinatorial objects are involved. In the case of linear probing hashing, it immediately implies that the number of clusters of some fixed size p has a distribution that is asymptotically Gaussian with mean and variance that are both $O(n)$.

4. Conclusion

The analysis of sparse tables (Section 3) is a by-product of the treatment of full tables (Section 2) that do constitute the primary combinatorial objects, so that we discuss them in more depth here. The Airy distribution and its companion moment formulæ turn out to be part of a ring of problems treated often independently by a variety of methods and authors. A brief census of "Airy phenomena" in combinatorial applications then reveals five main ranges of problems that we now list.

- (P_1) *Construction cost in linear probing hashing.* This is the context of Section 2 and the analysis applies almost *verbatim* to total displacement in parking sequences.
- (P_2) *Number of inversions in trees.* An inversion in a rooted labelled tree is a pair (i, j) such that i is on the path from the root to j and $i > j$. Exact generating functions have been first found by Mallows and Riordan [28] in the case of "Cayley" trees and other families of trees are considered in [9].
- (Q) *Connectivity in graphs.* A major problem in graphical enumeration and random graph theory [5, 15] is the determination of the number $\gamma(n, k)$ of *connected* graphs with n vertices and k edges. The basic problem was first solved by Wright in a famous series of papers [46, 47, 48]. Wright's solution involves a quadratic recurrent sequence that, after normalization, is the same as that of Section 2, so that the Airy constants make an appearance.
- (R_1) *Area of excursions.* By an excursion is meant a random walk that is never negative, and has initial and final altitudes both equal to 0; area is defined as the sum of altitudes of all nodes. The simplest type is the Bernoulli excursion defined by ± 1 steps (also called gambler's ruin sequence); Louchard [26, 27] established that the area of the Bernoulli excursion is asymptotically Airy distributed. Louchard's results are also related to other contemporary works from the early 1980's dealing with Brownian motion [34, 37] where Airy functions also crop up explicitly.
- (R_2) *Path length in trees.* The path length of a tree is the sum of the distances of all nodes to the root of the tree. In a series of papers, Takács [41, 43] has established limit Airy distribution results for various families of trees including Cayley trees and Catalan trees as special cases, while rederiving independently in [40, 42] some of the results of Louchard.

Regarding methods, our Theorem 3 establishes directly the Airy law for (P_1) by a recursive determination of moments. A similar process has been employed by Louchard and Takacs for (R_1) and (R_2). The underlying combinatorial decompositions are however quite different. For (P_1), one may regard the Airy law —via the quadratic recurrence or the Riccati equation— as a reflection of the basic binary tree-like decomposition of linear probing tables, while other decompositions prevail for (R_1) and (R_2). Problem (P_2), as far as we know, has not been

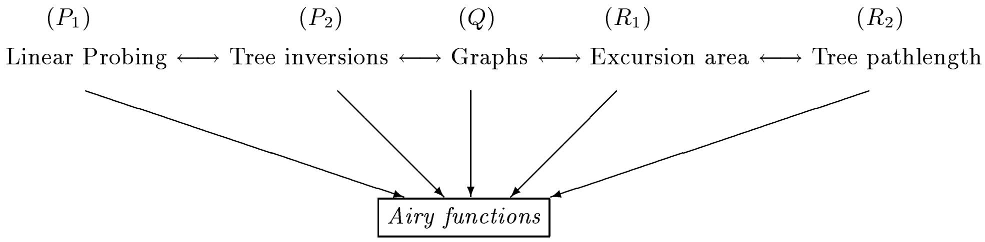

FIGURE 2. Five problems resorting to the "Airy phenomenon": (P_1, P_2, R_1, R_2) lead to the Airy distribution while (Q) involves the Airy coefficients.

previously considered under the angle of asymptotics and limit distributions, but it appears eventually to be an essential element in the combinatorial picture (see below).

The enumeration of connected graphs (Q) has a different status, since combinatorial enumerations rather than limit distributions are involved. Wright's major result states that

$$\gamma(n, n+k) \sim \frac{\sqrt{\pi}2^{(1-3k)/2}\sigma_k}{\Gamma((3k/2)+1)}\, n^{n+(3k-1)/2},$$

for a family of constants σ_k. A direct comparison between our basic recurrence of (29) and [46] shows that

$$\sigma_k = \frac{1}{2(k+1)!}\,\Omega_{k+1},$$

where the Ω_r are the Airy coefficients of Section 2.

Thus, three limit distribution problems, namely (P_1, R_1, R_2), have been found previously to lead to Airy laws, while structural constants of the Airy type are seen to appear in the graph connectivity problem (Q). This suggests a closer look at combinatorial relations between these problems.

$(P_1 - P_2)$: The fundamental recurrence of Section 2 in relation to (P_1) is indeed *identical* to the recurrence of [28] that models (P_2). Thus, the two problems must be combinatorially isomorphic. This fact has been noted by Knuth [22], following Kreweras [25]. Knuth also evokes in [22] the alternative of a direct and exact combinatorial correspondence based on [20, Ex. 6.4–31]. An immediate consequence of the present work in conjunction with the observations of Kreweras and Knuth is then: *the number of inversions in a random Cayley trees is asymptotically Airy distributed.*

$(P_2 - Q)$: An *exact* correspondence between inversions in trees and connectivity in graphs seems to have been first detected around the turn of the 1980's by several authors. Gessel and Wang [10] have an especially elegant formulation in terms of *depth* first search, where this mode of graph traversal leads precisely to a tree augmented by return edges that form inversions.

$(Q - R_1)$: This thread is due to Spencer [39] who noted that *breadth*–first search of a random connected graph has a "trace" that is a random excursion of the so-called Poisson type. (Depth–first search in the style of [10] is an alternative for this correspondence.) Under Spencer's correspondence, area under the Poisson excursion relates in *exact* terms to "excess" of the original graph.

$(R_1 - R_2)$: There are many known similarities between random walks and random trees. One of the most classical combinatorial correspondences relates bijectively Bernoulli excursions and general Catalan trees (see, *e.g.*, [18, 36]). Under this correspondence, area of an excursion transforms into path length of the associated tree.

These various relations[2] in a way "explain" the common occurrence of the Airy distribution in (P_1, P_2, R_1, R_2) as well as the rôle of the Airy coefficients in problem (Q). They also point to an alternative and more combinatorial deduction of the Airy law based on the following steps: (i) the exact equivalence between linear probing to inversions in trees by $(P_1 - P_2)$; (ii) the exact equivalence between tree inversions and graph connectivity $(P_2 - Q)$ by depth first search; (iii) the exact reduction to Poisson walks $(Q - R_1)$ by Spencer's principle; (iv) the reduction to Louchard's derivation of the Airy law through an appeal to universality of Brownian motion. Our derivation of Section 2 offers instead a self-contained approach to the problem.

Postscript. Apart from this conclusion section, the technical developments of our paper are otherwise independent of the recent preprint of Knuth [22]. Knuth has been kind enough to share numerous informations regarding [22], and this sheds additional light on the structure of the generating functions that appear in Sections 2 and 3. A major consequence of [22] is that the fundamental difference-differential equation admits of a closed form solution. This can be checked by direct comparison with [28] or [22]. (Alternatively, the combinatorial correspondences mentioned above could be used.) As a result, the bivariate generating function $F(z, q)$ happens to have an explicit expression,

$$
\begin{aligned}
F(z,q) &= (q-1)\frac{\partial}{\partial z}\log\left(1+\sum_{n=1}^{\infty} q^{n(n-1)/2}\frac{z^n(q-1)^{-n}}{n!}\right) \\
&= \frac{\displaystyle\sum_{n=0}^{\infty} q^{n(n+1)/2}\frac{z^n(q-1)^{-n}}{n!}}{\displaystyle\sum_{n=0}^{\infty} q^{n(n-1)/2}\frac{z^n(q-1)^{-n}}{n!}}
\end{aligned}
\tag{53}
$$

These forms are recognizable variants of the bivariate generating functions of graphs,

$$
F(z, q+1) = \sum_{n,t} \gamma(n, n+t-1) q^t \frac{z^{n-1}}{(n-1)!},
$$

and Knuth's analysis precisely starts from this relation.

Notice that (53) does not trivialize our moment computations, since it is already far from clear, given (53), that the tree functions should be involved. The manipulation of q-series expansions like (53) is in fact particularly delicate, as attested by the "Giant component" paper [15] on which [22] relies.

The exact correspondence with graph connectivity is at any rate neatly exposed by the equation (53). Unpublished work by Flajolet and Salvy (1995) inspired by Prellberg [33] indicates that moments can in fact be extracted from such a q-series expansion by a method of coalescent saddle points [45] that is well known to lead to Airy functions. We thus find yet another "reason" for the Airy phenomena observed here: a law of the Airy type may be expected whenever there occurs a coalescence of two neighbouring saddle points (a so-called "monkey saddle") in a Cauchy coefficient integral.

---

[2]Exchanges with Joel Spencer in 1995, while [39] was being developed, have been at the origin of our interest in Airy phenomena. Private communications with Knuth have then greatly helped us to complete the picture offered in this Section. See [22] for an insightful perspective.

Acknowledgements. The authors are thankful to Don Knuth, to whom this paper is dedicated, for his constant support and his openness in sharing his thoughts on the subject.

References

[1] Berndt, B. C. *Ramanujan's Notebooks, Part II.* Springer Verlag, 1989.
[2] Broder, A. Two counting problems solved via string encodings. In *Combinatorial Algorithms on Words*, A. Apostolico and Z. Galil, Eds., vol. 12 of NATO Advance Science Institute Series. Series F: Computer and System Sciences. Springer Verlag, 1985, pp. 229–240.
[3] De Bruijn, N. G. *Asymptotic Methods in Analysis.* Dover, 1981. A reprint of the third North Holland edition, 1970 (first edition, 1958).
[4] Flajolet, P., Grabner, P., Kirschenhofer, P., and Prodinger, H. On Ramanujan's Q-function. *Journal of Computational and Applied Mathematics 58*, 1 (1995), 103–116.
[5] Flajolet, P., Knuth, D. E., and Pittel, B. The first cycles in an evolving graph. *Discrete Mathematics 75* (1989), 167–215.
[6] Flajolet, P., and Odlyzko, A. M. Random mapping statistics. In *Advances in Cryptology* (1990), J.-J. Quisquater and J. Vandewalle, Eds., vol. 434 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 329–354. Proceedings of Eurocrypt'89, Houtalen, Belgium, April 1989.
[7] Flajolet, P., and Odlyzko, A. M. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics 3*, 2 (1990), 216–240.
[8] Flajolet, P., and Sedgewick, R. Analytic combinatorics. Book in preparation, 1998. (Individual chapters are available as INRIA Research Reports 1888, 2026, 2376, 2956, 3162.).
[9] Gessel, I., Sagan, B. E., and Yeh, Y.-N. Enumeration of trees by inversions. *Journal of Graph Theory 19*, 4 (1995), 435–459.
[10] Gessel, I., and Wang, D. L. Depth-first search as a combinatorial correspondence. *Journal of Combinatorial Theory,* Series A *26*, 3 (1979), 308–313.
[11] Gonnet, G. H., and Baeza-Yates, R. *Handbook of Algorithms and Data Structures: in Pascal and C*, second ed. Addison–Wesley, 1991.
[12] Goulden, I. P., and Jackson, D. M. *Combinatorial Enumeration.* John Wiley, New York, 1983.
[13] Greene, D. H., and Knuth, D. E. *Mathematics for the analysis of algorithms*, second ed. Birkhauser, Boston, 1982.
[14] Hennequin, P. Combinatorial analysis of quicksort algorithm. *RAIRO Theoretical Informatics and Applications 23*, 3 (1989), 317–333.
[15] Janson, S., Knuth, D. E., Luczak, T., and Pittel, B. The birth of the giant component. *Random Structures and Algorithms 4*, 3 (1993), 233–358.
[16] Kirschenhofer, P., Prodinger, H., and Tichy, R. Über einige Funktionaldifferentialgleichingen aus des Analyse von Algorithmen. In *Zahlentheoretische Analysis II* (1987), E. Hlawka, Ed., no. 1262 in Lecture Notes in Mathematics, pp. 111–123.
[17] Knuth, D. E. Notes on "open" addressing. Unpublished memorandum, 1963. (Memo dated July 22, 1963. With annotation "*My first analysis of an algorithm, originally done during Summer 1962 in Madison*". Also conjectures the asymptotics of the Q-function, with annotation "*Proved May 24, 1965*".).
[18] Knuth, D. E. *The Art of Computer Programming*, vol. 1 Fundamental Algorithms. Addison-Wesley Publishing Company, 1968.
[19] Knuth, D. E. *The Art of Computer Programming*, vol. 2 Seminumerical Algorithms. Addison-Wesley Publishing Company, 1969.
[20] Knuth, D. E. *The Art of Computer Programming*, vol. 3 Sorting and Searching. Addison-Wesley Publishing Company, 1973.
[21] Knuth, D. E. Analysis of optimum caching. *Journal of Algorithms 6* (1985), 181–199.
[22] Knuth, D. E. Linear probing and graphs. Preprint, July 1997.
[23] Knuth, D. E., and Rao, G. S. Activity in an interleaved memory. *IEEE Transactions on Computers C-24* (1975), 943–944.
[24] Knuth, D. E., and Schönhage, A. The expected linearity of a simple equivalence algorithm. *Theoretical Computer Science 6* (1978), 281–315.
[25] Kreweras, G. Une famille de polynômes ayant plusieurs propriétés énumératives. *Periodica Mathematica Hungarica 11* (1980), 309–320.
[26] Louchard, G. The Brownian excursion: a numerical analysis. *Computers and Mathematics with Applications 10*, 6 (1984), 413–417.

[27] Louchard, G. Kac's formula, Lévy's local time and Brownian excursion. *Journal of Applied Probability 21* (1984), 479–499.
[28] Mallows, C. L., and Riordan, J. The inversion enumerator for labeled trees. *Bulletin of the American Mathematical Society 1968* (74), 92–94.
[29] Moon, J. W. Counting labelled trees. In *Canadian Mathematical Monographs*, vol. 1. Canadian Mathematical Congress, 1970.
[30] Odlyzko, A. M. Asymptotic enumeration methods. In *Handbook of Combinatorics*, M. G. R. Graham and L. Lovász, Eds., vol. II. Elsevier, Amsterdam, 1995, pp. 1063–1229.
[31] Pittel, B. Linear probing: The probable largest search time grows logarithmically with the number of records. *Journal of Algorithms 8* (1987), 236–249.
[32] Poblete, P. Approximating functions by their Poisson transform. *Information Processing Letters 23* (1986), 127–130.
[33] Prellberg, T. Uniform q-series asymptotics for staircase polygons. *Journal of Physics A: Math. Gen. 28* (1995), 1289–1304.
[34] Rice, S. O. The integral of the absolute value of the pinned Wiener process—calculation of its probability density by numerical integration. *Annals of Probability 10* (1982), 240–243.
[35] Sedgewick, R. *Algorithms*, second ed. Addison–Wesley, Reading, Mass., 1988.
[36] Sedgewick, R., and Flajolet, P. *An Introduction to the Analysis of Algorithms.* Addison-Wesley Publishing Company, 1996.
[37] Shepp, L. A. On the integral of the absolute value of the pinned Wiener process. *Annals of Probability 10* (1982), 234–239.
[38] Sloane, N. J. A., and Plouffe, S. *The Encyclopedia of Integer Sequences.* Academic Press, 1995.
[39] Spencer, J. Enumerating graphs and Brownian motion. *Communications on Pure and Applied Math. 50* (1997), 293–296.
[40] Takács, L. A Bernoulli excursion and its various applications. *Advances in Applied Probability 23* (1991), 557–585.
[41] Takács, L. Conditional limit theorems for branching processes. *Journal of Applied Mathematics and Stochastic Analysis 4*, 4 (1991), 263–292.
[42] Takács, L. On a probability problem connected with railway traffic. *Journal of Applied Mathematics and Stochastic Analysis 4*, 1 (1991), 1–27.
[43] Takács, L. The asymptotic distribution of the total heights of random rooted trees. *Acta Scientifica Mathematica (Szeged) 57* (1993), 613–625.
[44] Wilf, H. S. *generatingfunctionology.* Academic Press, 1994.
[45] Wong, R. *Asymptotic Approximations of Integrals.* Academic Press, 1989.
[46] Wright, E. M. The number of connected sparsely edged graphs. *Journal of Graph Theory 1* (1977), 317–330.
[47] Wright, E. M. The number of connected sparsely edged graphs. II. Smooth graphs. *Journal of Graph Theory 2* (1978), 299–305.
[48] Wright, E. M. The number of connected sparsely edged graphs. III. Asymptotic results. *Journal of Graph Theory 4* (1980), 393–407.

P.F.: Algorithms Project, INRIA, Rocquencourt, 78150 Le Chesnay (France).
E-mail address: `Philippe.Flajolet@inria.fr.`

P.P.: Department of Computer Science, University of Chile, Casilla 2777, Santiago, Chile.
E-mail address: `ppoblete@dcc.uchile.cl.`

A.V.: Pedeciba Informatica, Casilla de Correo 16120, Distrito 6, Montevideo, Uruguay.
E-mail address: `viola@fing.edu.uy`

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399