Discrete Comput Geom 12:151–175 (1994)

Discrete & Computational
Geometry

Search Costs in Quadtrees and Singularity Perturbation Asymptotics*

P. Flajolet[1] and T. Lafforgue[2]

[1] Algorithms Project, INRIA, Rocquencourt,
F-78153 Le Chesnay, France
Philippe.Flajolet@inria.fr

[2] Laboratoire de Recherche en Informatique, Université Paris Sud,
F-91405 Orsay, France
lafforgue@lri.lri.fr

Abstract. Quadtrees constitute a classical data structure for storing and accessing collections of points in multidimensional space. It is proved that, in any dimension, the cost of a random search in a randomly grown quadtree has logarithmic mean and variance and is asymptotically distributed as a normal variable. The limit distribution property extends to quadtrees of all dimensions a result only known so far to hold for binary search trees.

The analysis is based on a technique of singularity perturbation that appears to be of some generality. For quadtrees, this technique is applied to linear differential equations satisfied by intervening bivariate generating functions

1. Introduction

This work concerns itself with an analysis in distribution of the cost of retrieving data from a randomly grown quadtree structure based on a combination of complex asymptotic and analytic probabilistic methods.

Quadtrees are a well-known data structure for multidimensional retrieval problems discovered by Finkel and Bentley [9]. They are discussed in classical treatises on algorithms [18], [31] and examined in great detail in Samet's reference books [29], [20]. Their analysis has made tangible progress over recent years [7], [10], [12], [20], [23], [27].

Given a list of points $\mathscr{P} = (P_1, P_2, \ldots, P_n)$ in two-dimensional space, the

* This work was partly supported by the ESPRIT Basic Research Action No. 7141 (ALCOM II).

standard quadtree process associates to it a tree defined by the rules:

—If $\mathscr{P} = \emptyset$, the tree is the empty tree. Otherwise, $n \geq 1$, and the first point P_1 of $\mathscr{P}$ is made the root of the tree.
—The four root subtrees are made recursively from the four disjoint sublists of points

$$\mathscr{P}_{\mathrm{NW}}, \mathscr{P}_{\mathrm{NE}}, \mathscr{P}_{\mathrm{SW}}, \mathscr{P}_{\mathrm{SE}},$$

defined by restricting $\mathscr{P} \backslash \{P_1\}$ to the four quadrants (NW, NE, SW, SE, respectively) determined by the root node P_1.

This definition is readily generalized to an arbitrary dimension d, the corresponding trees then having the branching factor 2^d.

The searching algorithm for a point P_0 in a quadtree constructed from a collection of data $\mathscr{P}$ starts with a comparison with the root; based on the outcome, it then recursively descends into one of the four subtrees. For any given $\mathscr{P}$ and P_0, this defines an access path whose length is characteristic of the search cost.

Throughout this paper, we let $d \geq 1$ be the dimension of the data space, and we liberally assume that data are from the d-dimensional hypercube $\mathscr{Q} = [0, 1]^d$. The probabilistic model considered takes all such data uniformly and independently from $\mathscr{Q}$. Having built a quadtree from $n - 1$ points under this model we consider the cost of searching an nth item in it, the search cost being measured as always by the number of internal nodes traversed. This search cost D_n (also called insertion depth) is then a random variable defined on the space $\mathscr{Q}^{n-1} \times \mathscr{Q} \cong \mathscr{Q}^n$. The outcome of the search is unsuccessful with probability 1 so that we are analysing with D_n a *random unsuccessful search*. Our main result is that D_n *converges in distribution to a Gaussian law* when the size n of the structure becomes large. Figure 1 illustrates the clear occurrence of this phenomenon already for low values of n.

More precisely, let μ_n and σ_n denote the mean and the standard deviation of the random variable D_n. We prove that, for all real α, β,

$$\Pr\left\{\alpha < \frac{D_n - \mu_n}{\sigma_n} \leq \beta\right\} \to \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-x^2/2}\, dx \qquad (n \to \infty), \tag{1}$$

where mean and standard deviation satisfy

$$\mu_n \sim \frac{2}{d} \log n \quad \text{and} \quad \sigma_n \sim \sqrt{\frac{2}{d^2} \log n}. \tag{2}$$

Similar results hold for the cost C_n of a *random successful search* where a random search is performed for one of the n records already present in the tree, the underlying probability space then being $\mathscr{Q}^n \times [1..n]$.

The type of analysis involved is perceptible when looking at the equation

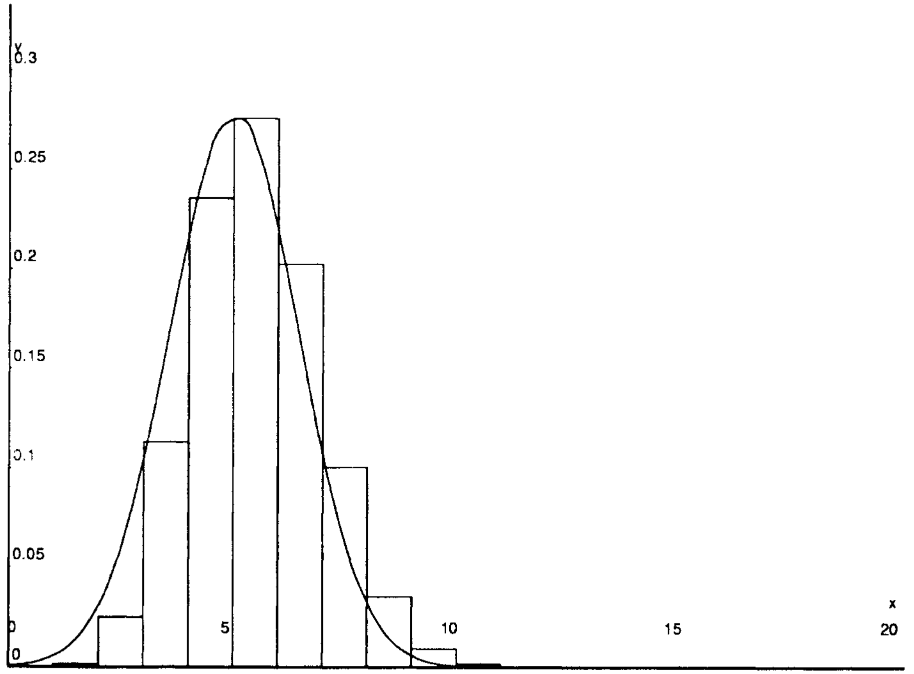

Fig. 1. The histogram of the probability distribution of D_n (for size $n = 100$ and dimension $d = 2$) plotted against a Gaussian density function of the same mean and variance.

satisfied by a modified form $\Phi(u, z)$ of the bivariate probability generating function $\sum_{n,k} \Pr\{D_n = k\} u^k z^n$, which, for dimension $d = 3$, reads

$$\Phi(u, z) = 1 + 2^3 u \int_0^z \frac{dx}{x(1-x)} \int_0^x \frac{dy}{y(1-y)} \int_0^y \Phi(u, t) \frac{dt}{1-t}. \tag{3}$$

The triple integral is a reflection of the combinatorices of the growth process of three-dimensional quadtrees.

Our results (1), (2) characterize the profile of a search in a quadtree of any dimension. The results already known are discussed in Mahmoud's book [27] that we adopt as our basic reference for analysis of search trees.

When $d = 1$, the quadtree reduces to the binary search tree [22]. In that case the distribution of D_n involves in a simple way the Stirling "cycle" numbers $\left[n \atop k \right]$ defined by

$$\sum_{k=0}^{n} \left[n \atop k \right] u^k = u(u+1)\cdots(u+n-1), \tag{4}$$

a fact known since the 1960s [15], [16], [26] and rediscovered by several authors. The distribution is Gaussian in the limit, in both the unsuccessful case [5] and the successful case [24]. This property is itself closely related to Gončarov's result of 1943 establishing the asymptotic normality of the Stirling cycle numbers.

When $d = 2$, the mean μ_n and the variance σ_n^2 have explicit forms [7], [10] that involve the harmonic numbers,

$$H_n := \sum_{k=1}^{n} \frac{1}{k} \quad \text{and} \quad H_n^{(2)} := \sum_{k=1}^{n} \frac{1}{k^2}. \tag{5}$$

We push the analysis further and derive a closed form for the generating functions of D_n and C_n using the hypergeometric equation known to play a crucial role in similar analyses [10], [20]. In this way the distribution of search costs becomes expressible as a complicated convolution of Stirling numbers and asymptotic normality results.

When $d \geq 3$, the asymptotic form of the mean, see (2), was determined by Devroye and Laforest using probabilistic arguments [7] and independently by Flajolet *et al.* [10] using singularity analysis of solutions to linear ordinary differential equations. We establish here the asymptotic form (2) for the variance (σ_n^2) which was previously unknown and which furnishes a quantitative refinement of the convergence-in-probability result of Devroye and Laforest. Furthermore —and this constitutes the main result of the paper—we obtain the asymptotic normality of the distribution of search cost (1) in all dimensions. Observe that already no closed forms are currently available (even for the mean) from the integral equation (3) for $d = 3$.

2. Basic Equations

The random tree problem described in the Introduction is readily recast as a purely analytic problem, as shown by Lemma 1 below. This section is devoted to the reduction, and the reader unfamiliar with (or uninterested by) search trees can take it as a starting point. In effect this lemma rephrases our problem as an instance of a general question which is of independent interest: "*Estimate the coefficients of a bivariate series that satisfies a linear ordinary differential equation with polynomial coefficients.*"

Two integral operators play an essential role here:

$$\mathbf{I}f(z) = \int_0^z f(t)\,\frac{dt}{1-t}, \qquad \mathbf{J}f(z) = \int_0^z f(t)\,\frac{dt}{t(1-t)}.$$

(When applied to a bivariate function $f(u, z)$, we always assume that the first variable u is an auxiliary parameter. See (3) for an example when $d = 3$.)

Lemma 1. *The generating functions of the costs of a random search, successful and unsuccessful, in a quadtree of size n are given by*

$$\begin{cases} \gamma_n(u) := \sum_k \Pr\{C_n = k\} u^k = \dfrac{1}{n} \dfrac{u}{2^d u - 1} (\varphi_n(u) - 1), \\ \delta_n(u) := \sum_k \Pr\{D_n = k\} u^k = \dfrac{1}{2^d u - 1} (\varphi_n(u) - \varphi_{n-1}(u)), \end{cases} \tag{6}$$

where the bivariate generating function

$$\Phi(u, z) = \sum_n \varphi_n(u) z^n$$

of the polynomials $\varphi_n(u)$ is characterized by the integral equation

$$\Phi(u, z) = 1 + 2^d u \mathbf{J}^{d-1} \mathbf{I} \Phi(u, z). \tag{7}$$

Proof. The central quantities here are the level polynomials $\varphi_n(u)$ that record the distribution of levels of external (empty) nodes in trees, and to which the distributions of C_n, D_n are then attached.

Consider arbitrary "regular" r-ary trees where each internal node has outdegree exactly r (for quadtrees, $r = 2^d$). For such a tree T, we define the (external) *level polynomial* $\varphi(u; T) = \sum_e u^{d(e)}$, where the sum extends to the external nodes of e and $d(e)$ is the depth of e measured in the number of internal nodes from the root of T to e. The level polynomial of the empty tree is 1 and inductively

$$\varphi(u; T) = u \sum_{j=1}^{r} \varphi(u; T_j), \tag{8}$$

with T_j the root subtrees of T.

The internal level polynomial is similarly $\psi(u; T) = \sum_i u^{d(i)}$ where the sum extends now to the internal nodes i of T, depth being still measured in the number of internal nodes on the branch of i. Since an internal node at depth k connects to r nodes, either internal with depth $k + 1$ or external with depth k, a balance relation holds,

$$t\psi(u; T) = \frac{1}{u} (\psi(u; T) - u) + \varphi(u; T). \tag{9}$$

Note that $\psi(u; T)/|T|$ describes the probability distribution of the cost of searching a random internal node conditioned upon the fact that the shape of the tree is T.

Next turn to the *quadtree growth process*. A tree of size n gives rise to a designated root subtree (for instance the NW subtree when $d = 2$) *having size* k with probability

$$\pi_{n,k} = \frac{1}{n} \sum_{\mathscr{L}} \frac{1}{(l_1 + 1)(l_2 + 1) \cdots (l_{d-1} + 1)}, \tag{10}$$

where the summation is over all sequences $(l_1, l_2, \ldots, l_d)$, the condition $\mathscr{L}$ being $n > l_1 \geq l_2 \geq \cdots \geq l_{d-1} \geq l_d = k$. These splitting probabilities are consequences of the quadtree growth process which they fully characterize. See Lemma 8 of [10] for a simple computation via Eulerian integrals.

Now define $\varphi_n(u)$ to be the *expectation* of the polynomial $\varphi(u; T)$ when T is a randomly grown tree of size n according to the quadtree process. (We also call $\varphi_n(u)$ a level polynomial.) Then, from (8) and (10), we get the recurrence

$$\varphi_0(u) = 1, \qquad \varphi_n(u) = 2^d u \sum_{k=0}^{n-1} \pi_{n,k} \varphi_k(u).$$

Taking generating functions, this is equivalent to (7).

The cost generating function $\gamma_n(u)$ of a random successful search C_n derives from $\varphi_n(u)$ by translating relation (9) into expectations, which gives the first part of (6). For an unsuccessful search, by a classical argument [22, p. 427], D_n measures the difference between the shapes of the tree at stages n and $n - 1$, so that the second part of (6) relative to $\delta_n(u)$ follows. □

We note that $[u^k]\varphi_n(u)$ is the expected number of external nodes at depth k in a randomly grown quadtree of n nodes. Except in the case of $d = 1$, it is not true that all external nodes get accessed with equal likelihood for randomly grown quadtrees.

3. The Binary Search Tree ($d = 1$)

When $d = 1$, the integral equation satisfied by $\Phi(u, z)$ is homogeneous of order 1, and thus solvable by quadratures:

$$\Phi(u, z) = \frac{1}{(1 - z)^{2u}} \quad \text{and} \quad \varphi_n(u) = \frac{(2u) \cdot (2u + 1) \cdot (2u + n - 1)}{n!}. \tag{11}$$

Thus, comparing with (4), we see that $[u^k]\varphi_n(u) = 2^k \left[{n \atop k} \right] \Big/ n!$, which involves the Stirling numbers. Proceeding in this vein, mean, variance, and distribution of D_n are found directly from Lemma 1 and (11).

Theorem 1 (Hibbard, Lynch). *The cost D_n of a random unsuccessful search in a binary search tree of size $n - 1$ has mean and variance given by*

$$\mu_n = 2(H_n - 1), \qquad \sigma_n^2 = 2H_n - 4H_n^{(2)} + 2,$$

and probability distribution

$$\Pr\{D_n = k\} = \frac{2^k}{n!}\begin{bmatrix} n-1 \\ k \end{bmatrix}.$$

Analogous results hold for C_n. They are originally due to Hibbard for the mean and Lynch for the whole distribution. See [22] and [27].

4. The Standard Quadtree ($d = 2$)

In the case of dimension $d = 2$, the analytic model of quadtrees can be solved explicitly in terms of hypergeometric functions. The corresponding easy background in analysis may be found in [1], [19], and [33].

Theorem 2. *The cost C_n of a random successful search in a standard quadtree of size $n - 1$ has a generating function $\gamma_n(u)$ given by*

$$\gamma_n(u^2) \equiv \mathbf{E}\{u^{2C_n}\} = \frac{1}{n}\frac{u^2}{4u^2 - 1}\left[-1 + \sum_{j=0}^{n}\binom{2u}{j}\binom{2u-1}{j}\binom{2u-1+n-j}{n-j}\right].$$

Equivalently, the distribution of C_n is expressible as a convolution of Stirling cycle numbers,

$$\Pr\{C_n = k\} = \frac{2^{2k-2}}{n}\left[1 - \sum_{j=0}^{n}\frac{1}{(j!)^2(n-j)!}\sum_{\mathcal{K}}(-1)^{k_1+k_2}\begin{bmatrix} j \\ k_1 \end{bmatrix}\begin{bmatrix} j+1 \\ k_2+1 \end{bmatrix}\begin{bmatrix} n-j \\ k_3 \end{bmatrix}\right],$$

where the sum $\sum_{\mathcal{K}}$ is to be taken over all triples (k_1, k_2, k_3) such that

$$(\mathcal{K}) \qquad k_1 + k_2 + k_3 \equiv 0 \pmod 2 \qquad \text{and} \qquad k_1 + k_2 + k_3 \le 2k - 2.$$

This also entails an explicit expression for the probability distribution of D_n since, from Lemma 1,

$$\Pr\{D_n = k\} = n[\Pr\{C_n = k+1\} - \Pr\{C_{n-1} = k+1\}].$$

Proof. From Lemma 1, the generating function of the level polynomials

$$\Phi(u, z) = 1 + 4uz + (4u^2 + 3u)z^2 + (22u + 52u^2 + 16u^3)\frac{z^3}{9} + \cdots$$

satisfies

$$\Phi(u, z) = 1 + 2^2u \int_0^z \frac{dx}{x(1-x)} \int_0^x \Phi(u, t)\frac{dt}{1-t}\ .$$

It is thus the solution of the linear differential equation of order 2,

$$z(1-z)^2y'' + (1-2z)(1-z)y' - 4u = 0. \tag{12}$$

This equation has singularities at the three points $z = 0, 1, \infty$ so that it is natural to compare it with the hypergeometric type.

In order to determine the local behavior at some point z_0 of possible solutions to a linear equation like (12), we simply consider a form $(z - z_0)^\alpha$, then identify the possible values of α by substituting into the equation and cancelling the dominant terms. This produces an *indicial equation* that is necessarily satisfied by α, and is here of degree 2. At $z_0 = 0$, where $\alpha^2 = 0$, two fundamental solutions are found in this way to grow like 1 and $\log z$. At $z_0 = 1$, where $\alpha^2 = 2^2u$, solutions are locally of the form $(1-z)^{-2\sqrt{u}}$ and $(1-z)^{+2\sqrt{u}}$. At $z_0 = \infty$, solutions behave like 1 and $1/z$.

This suggests setting

$$y = \frac{Y}{(1-z)^\alpha},$$

where $\alpha^2 = 2^2u$, and we choose the principal determination $\alpha = 2\sqrt{u}$, when u is near 1. Then Y satisfies an equation where one of the solutions is $O(1)$ as $z \to 1$, a property shared by the standard hypergeometric equation. The transformed equation makes possible a precise simplification by a factor of $1 - z$:

$$z(1-z)Y'' + (1 - z(2-2\alpha))Y' - \alpha(\alpha - 1)Y = 0. \tag{13}$$

The hypergeometric equation is

$$z(1-z)F'' + (c - (a + b + 1)z)F' - abF = 0, \tag{14}$$

which, under $F(0) = 1$, $F'(0) = ab$, admits the *hypergeometric* solution

$$F \equiv F[a, b; c; z] = 1 + \frac{a \cdot b}{c}\frac{z}{1!} + \frac{a(a+1) \cdot b(b+1)}{c(c+1)}\frac{z^2}{2!} + \cdots. \tag{15}$$

The two equations (13) and (14) are matched by the substitution

$$a = -\alpha, \quad b = 1 - \alpha, \quad c = 1 \qquad \text{with} \quad \alpha = 2\sqrt{u}.$$

Thus the generating function of the level polynomials $\Phi(u, z)$ admits the explicit form

$$\begin{aligned}\Phi(u^2, z) &= \frac{1}{(1-z)^{2u}} F[-2u, 1-2u; 1; z]\\ &= \left(\sum_{v=0}^{\infty} \binom{2u+v-1}{v} z^v\right) \cdot \left(\sum_{j=0}^{\infty} \binom{2u}{j}\binom{2u-1}{j} z^j\right).\end{aligned} \tag{16}$$

A convolution formula for $\varphi_n(u)$ then derives and the relations of Lemma 1 provide for $\delta_n(u)$. □

As an immediate corollary, we obtain the known values of the mean [7], [10] and of the variance [7] of a random search.

Theorem 3 (Devroye–Laforest, Flajolet *et al.*). *The mean and variance of a random search D_n in a standard quadtree of size $n - 1$ are given by*

$$\mu_n = H_n - \frac{1}{6} - \frac{2}{3n}, \qquad \sigma_n^2 = \tfrac{1}{2}H_n + H_n^{(2)} - \frac{13}{6} + \frac{5}{4n} - \frac{4}{9n^2}.$$

Proof. Compute $\partial\Phi/\partial u$ and $\partial^2\Phi/\partial u^2$, evaluate at $u = 1$, and expand. □

In preparation for our subsequent discussion, we note that the generating function Φ admits the expansion at $z = 1$:

$$\begin{aligned}\Phi(u^2, z) &= \frac{\Gamma(4u)}{\Gamma(2u)\Gamma(1+2u)}(1-z)^{-2u}F[-2u, 1-2u; 1-4u; 1-z]\\ &\quad + \frac{\Gamma(-4u)}{\Gamma(-2u)\Gamma(1-2u)}(1-z)^{2u}F[+2u, 1+2u; 1+4u; 1-z].\end{aligned} \tag{17}$$

Such a form is available since the connection formulae for hypergeometrics are fully explicit due to the existence of integral representations. In what follows we see that expressions *qualitatively* similar to (17), although much less explicit, hold in higher dimensions.

Asymptotic normality for C_n and D_n would result from these developments using the main theorem of Flajolet and Soria [14]. A derivation is, however, not given here as it is subsumed by the more general treatment valid for all dimensions that we now expose.

5. The Singularity Perturbation Method

The architecture of the proof of the main theorem asserting asymptotic normality of the distribution of search costs in all dimensions is transparent; implementation of it requires quite some care, though. We offer here a brief outline.

The starting point is the integral equation (14) furnished by Lemma 1, which we recall:

$$\Phi(u, z) = 1 + 2^d u \mathbf{J}^{d-1} \mathbf{I} \Phi(u, z). \tag{18}$$

That equation is itself equivalent to a linear differential equation (see (12) for dimension $d = 2$) with coefficients that are polynomial in the main variable z and the parameter u. The order of the equation is equal to the dimension of the data space, d. The standard theory is more conveniently developed from *differential systems* rather than equations, and the associated system is also of dimension d. (Systems are notationally simpler because of the more transparent form afforded by matrix exponentials as well as the simpler expression of the variation-of-constant formulae [19].)

The main idea consists in relating perturbations of the differential system corresponding to (18) which is singular at $z = 1$ when u is near 1 to the asymptotic properties of the coefficients of $\Phi(u, z)$.

The most common case for linear differential equations and systems is the one called *regular singularity* or *singularity of the first kind*. In such a case, a basis of solutions can be found that, in essence, are locally of the form

$$\frac{c}{(1-z)^\alpha}.$$

The possible exponents α are determined by substituting into the equation and expressing cancellation of the dominant terms. They thus appear as roots of a polynomial called the *indicial polynomial.*

In a parametrized case like (18), we thus expect solutions to involve linear combinations of terms of the form

$$\frac{c(u)}{(1-z)^{\alpha(u)}}, \tag{19}$$

as $z \to 1$. In the case of (18), it is found that the possible exponents are algebraic functions that are roots of the indicial equation

$$(\alpha(u))^d - 2^d u = 0.$$

Forms belonging to the general type (19) were already encountered when $d = 1$, see (11), and when $d = 2$, see (17). Asymptotic normality of coefficients is known to hold for a closely related class of bivariate functions exhibiting a similar singular behavior [13].

As $z \to 1$, the dominant term in the expansion of $\Phi(u, z)$ is the one corresponding to the root $2u^{1/d}$ which has maximal real part. In particular when the parameter u is close to 1, this is the principal determination of $2\sqrt[d]{u}$. From the shape (19) of singular elements, we thus expect the singular form of Φ to be

$$\Phi(u, z) \approx \frac{c(u)}{(1-z)^{2u^{1/d}}} \qquad (z \to 1), \tag{20}$$

at least for u near 1.

According to the usual principles of singularity analysis [11], the *dominant singular behavior* of Φ provides the dominant asymptotic term in its coefficients $\varphi_n(u) = [z^n]\Phi(u, z)$. Translating (20) to coefficients, we expect, as an approximation of $\varphi_n(u)$,

$$\varphi_n(u) \approx c(u) \frac{n^{2u^{1/d}-1}}{\Gamma(2u^{1/d})}. \tag{21}$$

Given the approximation (21), values of the polynomial $\varphi_n(u)$ are asymptotically known at least for u in a neighborhood of 1. An inversion problem—the second one after the phase of singularity analysis ensuring the transition from (20) to (21)—is then to be solved. The approximation (21) permits estimating $\varphi_n(e^{i\theta})$, suitably normalized, when θ is near 0. The Fourier transform of the distribution defined by the coefficients of $\varphi_n(u)$ is found to tend to $e^{-\theta^2/2}$, the characteristic function of the Gaussian distribution,

$$\lim_{n \to +\infty} e^{-i\theta a_n/b_n} \frac{f_n(e^{i\theta/b_n})}{f_n(1)} = e^{-\theta^2/2}, \tag{22}$$

for some suitably chosen a_n, b_n.

Since $\varphi_n(u)$ has positive coefficients, the continuity theorem for characteristic functions (or equivalently Fourier transforms of measures) of analytic probability theory applies. This leads to the end result, namely the convergence in distribution to a normal distribution for the coefficients of $\varphi_n(u)$ which in turn carries to the distribution of D_n as expressed by (1).

The technical difficulty of the actual proof, compared with this rough outline, is due to the strict necessity of deriving singular expansions that are *uniform* with respect to u. This requires a detailed investigation of the way such expansions are "perturbed" when u lies near 1, hence the term *singularity perturbation* in our title.

The necessary background on singularities of linear differential systems may be found in [6], [19], and [32].

6. The Higher-Dimensional Quadtree ($d \geq 3$)

The main result to be established in this section is that the distribution of a random unsuccessful search in a d-dimensional quadtree is asymptotically normally distributed, the proof following the outline of the previous section. We then prove an exponential tail result for the distribution. From there, the asymptotic forms of the mean and variance of the distributions follow. The same properties also hold for a random successful search, by a direct adaptation of the arguments.

Theorem 4. *The cost of a random unsucessful search in a randomly grown quadtree converges in distribution to a normal variable, i.e., for all real* α, β,

$$\Pr\left\{\alpha < \frac{D_n - a_n}{b_n} \le \beta\right\} \to \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-x^2/2}\, dx \qquad (n \to \infty), \tag{23}$$

where the centering constants are

$$a_n = \frac{2}{d} \log n \quad \textit{and} \quad b_n = \sqrt{\frac{2}{d^2} \log n}. \tag{24}$$

The proof of Theorem 4 starts with general analytic conditions for normality (Lemma 2) followed by a detailed analysis of the differential equation expressing the physics of quadtrees (Lemma 3). The analytic lemma, Lemma 2, is closely related to bivariate schemas considered by Flajolet and Soria [13], [14], and recently extended by Gao and Richmond [17].

Lemma 2. *Let* $F(u, z) = \sum_{n,k} f_{n,k} u^k z^n$ *be a bivariate function with positive coefficients. Assume that*:

C1. $F(u, z)$ *is analytic in* $\mathbb{V} \times \mathbb{C}\backslash[1, +\infty[$ *with* $\mathbb{V}$ *some neighborhood of* 1.

C2. *In the intersection of a neighborhood of* (1, 1) *and* $\mathbb{V} \times \mathbb{C}\backslash[1, +\infty[$,

$$F(u, z) = \frac{1}{(1 - z)^{\alpha(u)}} (c(u) + \eta(u, z)),$$

where

(i) $\alpha(u)$ *is analytic at* $u = 1$ *and* $\alpha(1) > 0$.

(ii) $c(u)$ *is analytic at* $u = 1$ *with* $c(1) \neq 0$.

(iii) $\eta(u, z) = o(1)$ *as* $z \to 1$ *uniformly in* u.

Then the coefficients $f_{n,k}$ *are asymptotically normal with centering constants*

$$a_n = \alpha'(1) \log n \quad \textit{and} \quad b_n^2 = (\alpha'(1) + \alpha''(1)) \log n.$$

In other words, for all real β,

$$\frac{\sum_{k \le a_n + \beta b_n} f_{n,k}}{\sum_k f_{n,k}} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\beta e^{-x^2/2}\, dx. \tag{25}$$

Proof. (Sketch; see [13], [14], and [17] for details.) Integration along a Hankel contour according to the principles of singularity analysis [11] yields the approximation valid in a neighborhood of (1, 1),

$$f_n(u) = \frac{n^{\alpha(u)-1}}{\Gamma(\alpha(u))}(c(u) + o_u(1)), \tag{26}$$

where $o_u(1)$ indicates uniformity with respect to u in a neighborhood of 1, as $n \to \infty$. In other words, we have transferred termwise a uniform expansion of $F(u, z)$ onto its coefficient. This is permissible because of the constructive character of error terms afforded by the singularity analysis method [11].

With the stated values of a_n and b_n, a direct computation from (26) shows that, for all fixed θ,

$$\lim_{n\to+\infty} e^{-i\theta a_n/b_n}\frac{f_n(e^{i\theta/b_n})}{f_n(1)} = e^{-\theta^2/2}, \tag{27}$$

the proof requiring the continuity of $c(u)$ at 1, the condition $c(1) \neq 0$ and the uniformity of the error term $o_u(1)$ in the expansion of $f_n(u)$.

Thus, the characteristic function of the distribution $\{f_{n,k}\}$ (varying k) tends to that of a standard normal variable as $n \to \infty$. By the continuity theorem for characteristic functions (see Section 26 of [4] or [25]), this implies pointwise convergence of the corrresponding distribution functions, which is what (25) precisely expresses.

Note finally that the one-sided relation of (25) with $\int_{-\infty}^{\beta}$ trivially entails a two-sided version $\int_{\alpha}^{\beta}$ as stated in Theorem 4. This concludes the proof of Lemma 2. □

The next lemma constitutes the core of the argument of the proof of Theorem 4. It establishes that the bivariate series Φ satisfies the conditions of Lemma 2 with $\alpha(u) = 2u^{1/d}$.

Lemma 3. *In any dimension $d \geq 1$, the generating function $\Phi(u, z)$ of the level polynomials of quadtrees (defined by (7)) and the generating function of quadtree search costs*

$$\Delta(u, z) \equiv \sum_n \delta_n(u)z^n = \sum_{k,n} \Pr\{D_n = k\}u^k z^n$$

both satisfy the conditions of Lemma 2 ensuring asymptotic normality.

Proof. 1. *Positivity.* From the combinatorics of the problem, we find

$$\Phi(1, z) = \frac{1 + (2^d - 2)z}{(1 - z)^2} \tag{28}$$

or $\varphi_n(1) = 1 + (2^d - 1)n$. Given the positivity of coefficients, the function Φ is thus *a priori* analytic in $|z| < 1$, $|u| < 1$.

2. *The differential system.* The integral equation (7) satisfied by Φ gives rise to a differential equation of order d,

$$\mathbf{I}^{-1}\mathbf{J}^{1-d}\Phi(u, z) = 2^d u\Phi(u, z).$$

By standard reduction techniques, that equation transforms into a differential *system.* In effect, the vector $(\Phi(u, z), \mathbf{I}\Phi(u, z), \dots, \mathbf{J}^{d-2}\mathbf{I}\Phi(u, z))$ is a solution to

$$\frac{d}{dz} Y(u, z) = \frac{1}{1-z}\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 2u/z \\ 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2/z & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2/z & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 2/z & 0 \end{pmatrix} Y(u, z). \tag{29}$$

3. *Analyticity.* Under the form (29), it is recognized that the system has two *regular singularities* at the *fixed points* $z = 0$ and $z = 1$; here "fixed" means "nonmovable." (A more general discussion of fixed versus movable singularities is given in the last section of the paper.) The general setting of the problem as we saw in step 1 guarantees that Φ is analytic at $z = 0$, so that this point needs no further attention.

The fundamental theorem of regular perturbation guarantees that the solution Φ remains analytic in both the parameter u and the main variable z as long as the dependency on parameters is analytic and singularities corresponding to the main variable are avoided.[1] The dependency on u is entire, so that $\Phi(u, z)$ is indeed an analytic function of the two complex variables (u, z) for $(u, z) \in \mathbb{C} \times \mathbb{C}\setminus[1, +\infty[$.

4. *Approximate Euler system at $z = 1$.* Singling out the singular part at $z = 1$, the differential system (29) writes

$$\frac{d}{dz} Y(u, z) = \left(\frac{M(u)}{1-z} + E(u, z)\right) Y(u, z) \tag{30}$$

with

$$M(u) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 2u \\ 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 2 & 0 \end{pmatrix}, \tag{31}$$

where $E(u, z)$ is analytic on $\mathbb{C} \times (\mathbb{C}\setminus\{0\})$.

---

[1] From now on, we globally refer to Section 24 of Wasow's book [32] or Sections 7 and 8 in Chapter 1 of the book by Coddington and Levinson [6].

A first-order approximation of (30) is

$$\frac{d}{dz}\Upsilon(u, z) = \frac{M(u)}{1-z}\Upsilon(u, z),$$

which constitutes a system of the *Euler type* that admits explicit solutions. The characteristic polynomial of $M(u)$ is $\alpha^d - 2^d u$, so that the *eigenvalues* of $M(u)$ are simply the numbers

$$\lambda_j(u) = \lambda(u)\omega^j \qquad \text{where} \quad \lambda(u) = 2u^{1/d} \quad \text{and} \quad \omega = e^{2i\pi/d} \tag{32}$$

for $j = 0, \ldots, d-1$. No problem with branch determinations arises as long as u stays in a neighborhood of 1 that avoids 0.

In particular, $M(1)$ has d distinct eigenvalues and is therefore diagonalizable, this property remaining true as long as $u = 0$ is avoided. For instance, $M(u)$ is diagonalizable in the open ball $B(1, 1)$ of center 1 and radius 1. Furthermore, $M(u)$ is analytic at $u = 1$. It results from a general observation of Sibuya that the diagonalization of an analytic matrix is itself an analytic process. Thus, see Section 25 of Wasow's book [32], an analytic matrix $Q(u)$, invertible over $B(1, 1)$, exists such that

$$M(u) = Q(u)^{-1}D(u)Q(u) \qquad \text{with} \quad D(u) = \mathrm{Diag}(\lambda_0(u), \ldots, \lambda_{d-1}(u)). \tag{33}$$

5. *Approximate singularity analysis at* $z = 1$. We return to the full differential system (29) and set $V(u, z) = Q(u)^{-1}Y(u, z)$ (with Y our particular solution vector involving Φ). The goal is to build up a solution to the full system from the solution to the Euler system. The particular solution vector $V(u, z)$ is, by construction, analytic on $B(1, 1) \times \mathbb{C}\backslash[1, +\infty[$. Thus, functions $a_j(u)$ analytic on $B(1, 1)$ exist such that

$$\Phi(u, z) = \sum_{j=0}^{d-1} a_j(u)V_j(u, z). \tag{34}$$

Furthermore, $V(u, z)$ satisfies the transformed differential system

$$\frac{d}{dz}V(u, z) = \left(\frac{D(u)}{1-z} + F(u, z)\right)V(u, z), \tag{35}$$

where $F(u, z)$ is analytic on $B(1, 1) \times \mathbb{C}\backslash\{0\}$.

For the simplified form of the system (35) in which $F(z, u)$ is set to 0 (this is now, by construction, a diagonal Euler system), a vector of solutions is given by

$$V_j^*(u, z) = \frac{c_j(u)}{(1-z)^{\lambda_j(u)}}. \tag{36}$$

Near $z = 1$, each $|V_j^*(u, z)|$ behaves like

$$O(|1 - z|^{-\Re(\lambda_j(u))}).$$

The dominant singular behavior is thus $(1 - z)^{-\lambda(u)}$ since $\lambda(u)$ $(=\lambda_0(u))$ is the determination with the largest real part when u is near 1.

At this stage, our problem is reduced to showing that the presence of the correction term $F(u, z)$ in (35) does not radically affect the solutions so that the V_j are approximated by the V_j^*, themselves satisfying (36).

6. *Singularity analysis at $z = 1$, odd dimension.* We proceed to prove that the exact solutions (36) of the approximate (diagonal) Euler system do represent asymptotically the exact solutions of the full system. In the univariate case this is a well-known fact in the theory of regular singularities, though complications arise in certain confluence situations—when two λ_j are congruent modulo 1—which may induce logarithmic terms [6], [19], [32].

For lower dimensions ($d = 1, 2$), a direct computation from (11) and (17) confirms that $V_j \sim V_j^*$ and permits us to establish the statement of the lemma directly.

For an arbitrary *odd*-valued d, the eigenvalues $\lambda_j(u)$ are distinct and no two of them are congruent modulo 1, since their imaginary parts are all distinct for $u \in B(1, 1)$. From the general theorem of regular singularity, a fundamental solution to the main system (29) of the form

$$P(u, z)(1 - z)^{-D(u)},$$

with P analytic in z, when $z \in B(1, 1)$, for each fixed choice of $u \in B(1, 1)$, exists. The global dependency of P with respect to u, especially analyticity, is however to be ascertained.

The general theorem of regular singularity relies on recurrence relations that the differential equation induces for the coefficients of the P matrix, and analyticity then readily follows from direct majorizations. In effect, the proof of Theorem 4.1, p. 119, of [6] adapts to our parametrized problem and $P(u, z)$ turns out to be analytic in both variables u and z, for (u, z) in a neighborhood of (1, 1). To see it, set

$$P(u, z) = \sum_{n=0}^{\infty} P_n(u) z^n.$$

First, by the recurrence translating the differential equation, the $P_n(u)$ are each analytic for $u \in B(1, 1)$. Next, from the Cauchy inequalities applied to $F(u, z)$ and from the recurrence, a uniform upper bound of the form

$$|P_n(u)| \le C \cdot (n + 1)^2$$

follows for $u \in B(1, 1)$, with C some positive constant. As a result,

$$\Phi(u, z) = \sum_{j=0}^{d-1} b_j(u, z)(1 - z)^{-\lambda_j(u)}, \tag{37}$$

where the $b_j(u, z)$ are analytic in $B(1, 1) \times B(1, 1)$.

Equation (37) constitutes the main analytic stage of the proof. It shows hat Φ satisfies the conditions of Lemma 2, for odd values of d.

7. *Singularity analysis at $z = 1$, even dimension.* For an arbitrary *even*-valued d, the eigenvalues $\lambda_j(u)$ are all distinct, for $u \in B(1, 1)$. However, some of them become congruent modulo 1, when $u = 1$. This is always the case for the pair $\{-2, +2\}$, and it may happen for other roots, in the hexagonal configuration corresponding to $d = 6$ for instance.

At $u = 1$, at most two distinct eigenvalues may be congruent modulo 1 (examine the imaginary parts). Thus, from the general theory of linear differential systems,

$$\Phi(1, z) = \sum_{j=0}^{d} b_j(z) \frac{1}{(1 - z)^{\lambda_j(1)}} + \log \frac{1}{1 - z} \sum_{j=0}^{d} \hat{b}_j(z) \frac{1}{(1 - z)^{\lambda_j(1)}}, \tag{38}$$

where the $b_j(z)$ and $\hat{b}_j(z)$ are analytic at 1 (some possibly equal to 0). In particular, the special solution (28) is of this form with all the $\hat{b}_j(z) = 0$ and with the $b_j(z) = 0$ for $j \neq 0$.

In contrast, for u close enough to 1 but $u \neq 1$, a simple computation shows that the $\lambda_j(u)$ are all distinct modulo 1, so that

$$\Phi(u, z) = \sum_{j=0}^{d-1} b_j(u, z) \frac{1}{(1 - z)^{\lambda_j(u)}}, \tag{39}$$

where (for each u separately) the $b_j(u, z)$ are analytic in z. We let $\mathbb{V}$ denote a sufficiently small neighborhood of 1 in which the $\lambda_j(u)$ remain distinct modulo 1, except possibly at $u = 1$ itself.

Comparison of (38) and (39) precludes a matching of the two expansions at $u = 1$, and $b_j(u, z)$ cannot depend analytically on u at $u = 1$ whenever logarithmic terms occur. A solution is obtained by using an idea due to Frobenius, and changing the base functions in which solutions are to be expressed. Instead of the base functions underlying (38) that are of the form

$$\frac{1}{(1 - z)^{\lambda}}, \frac{1}{(1 - z)^{\lambda}} \log \frac{1}{1 - z},$$

we introduce

$$\varepsilon_{\alpha,\beta}(z) = \begin{cases} \dfrac{1}{\alpha-\beta}\left[\dfrac{1}{(1-z)^{\alpha}} - \dfrac{1}{(1-z)^{\beta}}\right] & \text{if} \quad \alpha \neq \beta, \\ \dfrac{1}{(1-z)^{\alpha}} \log \dfrac{1}{1-z} & \text{if} \quad \alpha = \beta. \end{cases} \tag{40}$$

The base function $\varepsilon_{\alpha,\beta}(z)$ is now analytic in α, $\beta \in \mathbb{C}$ and $z \in \mathbb{C}\backslash[1, +\infty[$.

Let us group the λ_j into equivalence classes according to the values of the $\lambda_j(1)$ modulo 1. Let J denote the collection of the equivalence classes that comprise two eigenvalues. For instance, when $d = 4$, the $\lambda_j(u)$ are in order $2u^{1/4}$, $2iu^{1/4}$, $-2u^{1/4}$, $-2iu^{1/4}$, the equivalence classes of the $\lambda_j(1)$ modulo 1 are $\{2, -2\}$, $\{2i\}$, and $\{-2i\}$, and J contains one equivalence class, namely, $\{\lambda_0, \lambda_2\}$ corresponding to $\{2, -2\}$. Each class of J is thus associated to two eigenvalues $\lambda_{k(j)}$ and $\lambda_{l(j)} = \lambda_{k(j)} + m(j)$, where $m(j)$ is a nonnegative integer.

The classical argument of Frobenius (see Section 4.8 of [6]) leads to the existence of an expansion replacing (39),

$$\Phi(u, z) = \sum_{j=0}^{d-1} c_j(u, z)\left(\frac{1}{1-z}\right)^{\lambda_j(u)} + \sum_{j \in J} d_j(u, z)\varepsilon_{\lambda_{k(j)}(u),\, \lambda_{l(j)}(u) - m(j)}. \tag{41}$$

An adaptation of the treatment of [6, pp. 120–121] to the parametrized case shows that we may take the $c_j(u, z)$ and $d_j(u, z)$ to be analytic in $\mathbb{V} \times B(1, 1)$. Details are relegated to an appendix.

The expansion (41) of Φ now has the uniform behavior encompassing both cases, $u \neq 1$ and $u = 1$, sought. It satisfies the conditions of Lemma 2. In particular, the $c(u)$ of that lemma is continuous since the terms involving the ε's only contribute negligibly, as

$$\varepsilon_{\alpha,\beta}(z) = o\left(\log \frac{1}{|1-z|} \cdot \frac{1}{|1-z|^{\max \Re(\alpha), \Re(\beta)}}\right) \qquad (z \to 1).$$

8. *Conclusion.* The bivariate generating function of search costs is readily computed from Lemma 1, and it equals

$$\frac{1-z}{2^d u - 1}\,\Phi(u, z).$$

Thus from (37)—for odd dimensions—and (41)—for even dimensions—the bivariate generating function $\Phi(u, z)$ and its variant $\Delta(u, z)$ satisfy the conditions of Lemma 2, with $\alpha(u) = 2u^{1/d}$ for Φ and with $\alpha(u) = 2^{1/d} - 1$ for Δ that is related to Φ by (6) of Lemma 1. This completes the proof of Lemma 3. $\square$

Theorem 4 is now established by a direct combination of Lemmas 2 and 3.

Note. Some of the intricacies of the proof arose from the confluence of eigenvalues modulo 1 in the case when d is even. Confluences of order higher than 2 could also be coped with using the base functions

$$\varepsilon_{\alpha_1,\ldots,\alpha_r} = \frac{1}{r}\sum_{j=1}^{r}\prod_{k\neq j}\frac{1}{(\alpha_j-\alpha_k)}\left[\frac{1}{(1-z)^{\alpha_j}}-\frac{1}{(1-z)^{\alpha_k}}\right].$$

The next theorem provides *uniform exponential tails* for the probability of large deviations of D_n which improves on the convergence-in-probability result of [7].

Theorem 5. *Two positive constants C and $\alpha < 1$ exist such that, for all n and k,*

$$\Pr\left\{\left|\frac{D_n-a_n}{b_n}\right| > k\right\} < C\cdot\alpha^k.$$

Proof. (Sketch, see [14] and [17] for details.) The proof is a simple adaptation of the argument giving the characteristic function, with u now taken in a *real* neighborhood of 1.

From the conditions of the Lemma 2, and by the same singularity analysis argument as (26) and (27), a fixed real neighborhood of 1 exists such that, for θ in that neighborhood,

$$\lim_{n\to+\infty} e^{-\theta a_n/b_n}\frac{f_n(e^{\theta/b_n})}{f_n(1)} = e^{\theta^2/2}.$$

In other words, the Laplace transform of $\Omega_n = (D_n - a_n)/b_n$ converges to the Laplace transform of a normal variable. The uniformity conditions of Lemma 2 further ensure that

$$e^{-\theta a_n/b_n}\frac{f_n(e^{\theta/b_n})}{f_n(1)}$$

stays uniformly bounded for θ in some interval $[-\theta_1, \theta_2]$ containing 0.

It is well known (see [4], and [14] for the uniform version) that existence of Laplace transforms in an interval surrounding 0 implies exponential tails. The statement of Theorem 5 simply expresses this fact. □

The next theorem gives the asymptotic form of the mean and variance of a search. It is obtained here as a by-product of the limit distribution property and its centering constants (Theorem 4) complemented by effective tail estimates (Theorem 5). An analytic derivation along the lines of [10] should also be feasible.

Theorem 6. *The mean μ_n and standard deviation σ_n of a random search in a random quadtree of size $n-1$ in some arbitrary dimension $d \geq 1$ satisfy asymptotically*

$$\mu_n \sim \frac{2}{d} \log n \quad \textit{and} \quad \sigma_n \sim \sqrt{\frac{2}{d^2} \log n}. \tag{42}$$

The mean value estimate was already obtained by Flajolet *et al.* using analytic methods, and independently by Devroye and Laforest using a probabilistic geometric argument.

Proof. It need not be true in every generality that the centering constants a_n, b_n be equal, or even asymptotically equal, to the mean μ_n and standard deviation σ_n of the distribution of index n. In other words, convergence in distribution (weak convergence) is not sufficient to ensure convergence of moments, the latter being affordable here by the uniform exponential tail estimates of Theorem 5.

Let X_n denote the normalized variable $(D_n - a_n)/b_n$. We need to show that X_n has mean $o(1)$ and variance $1 + o(1)$. The expectation of X_n is

$$\mathbf{E}\{X_n\} = \int_{-\infty}^{0} -F_n(x)\,dx + \int_{0}^{+\infty} (1 - F_n(x))\,dx, \tag{43}$$

where F_n is the distribution function of X_n. The function $F_n(x)$ converges pointwise to $F_\infty(x)$, the distribution function of a standard Gaussian variable that satisfies $\mathbf{E}\{X_\infty\} \equiv 0$.

By Theorem 5, $F_n(x)$ has uniform exponential tails:

$$0 \leq F_n(x) \leq C\alpha^{-x} \quad (x < 0) \qquad \text{and} \qquad 0 \leq 1 - F_n(x) \leq C\alpha^{x} \quad (x < 0).$$

As $\int_{-\infty}^{0} C\alpha^x\,dx$ and $\int_0^{+\infty} C\alpha^{-x}\,dx$ both converge, Lebesgue's dominated convergence theorem applies. Thus,

$$\lim_{n \to +\infty} \mathbf{E}\{X_n\} = \mathbf{E}\{X_\infty\} = 0.$$

In other words, we have $(\mu_n - a_n)/b_n = o(1)$, so that $\mu_n = a_n + o(b_n)$.

The proof that $\sigma_n = b_n + o(b_n)$ results from a similar consideration of the variance of X_n, starting from

$$\mathbf{E}\{X_n^2\} = \int_{-\infty}^{0} -2xF_n(x)\,dx + \int_{0}^{+\infty} 2x(1 - F_n(x))\,dx. \qquad \square$$

From this last theorem, the centering constants a_n, b_n of the limit distribution (Theorem 4) may be replaced by the mean and standard deviation μ_n, σ_n, as was expressed in (1).

Table 1. The values of $\Pr\{D_n \geq k\}$ against values of k, for $n = 100$ and for standard quadtrees, $d = 2$.

k	10	20	30	40	50	60	70	80	90
$\Pr\{D_n \geq k\}$	$2 \cdot 10^{-3}$	$2 \cdot 10^{-14}$	$3 \cdot 10^{-33}$	$2 \cdot 10^{-54}$	$3 \cdot 10^{-80}$	$3 \cdot 10^{-110}$	$4 \cdot 10^{-141}$	$3 \cdot 10^{-176}$	$4 \cdot 10^{-215}$

Also, it results from an observation of Gao and Richmond [17] that a *local limit* theorem holds, with a direct convergence of the probabilities $\Pr\{D_n = k\}$ to the Gaussian *density*. This what the histogram of Figure 1 actually depicts.

Table 1 displays a sample of the probability distribution of D_n determined exactly using computer algebra, in the case of dimension $d = 2$ and $n = 100$. The low figures confirm that probabilities of large deviations soon become exceedingly small.

7. Conclusion

The method of singularity perturbation developed here is of a generality that transcends the particular situation of quadtrees. Retaining the essentials of the argument, we obtain in effect a result valid for large classes of differential equations.

Theorem 7. *Let $f_n(u)$ be a sequence of polynomials with positive coefficients satisfying the following conditions.*

C1. [Fixed regular singularity] *The generating function $F(u, z) = \sum_n f_n(u)z^n$ satisfies a linear differential equation of the form*

$$a_0(u, z)\frac{\partial^r F}{\partial z^r} + \frac{a_1(u, z)}{(1 - z)}\frac{\partial^{r-1} F}{\partial z^{r-1}} + \cdots + \frac{a_r(u, z)}{(1 - z)^r} F = 0,$$

where the $a_j(u, z)$ are polynomials and $a_0(u, z) \neq 0$ for $|z| \leq 1$, $|u| \leq 1$.

C2. [Nonconfluence] *The indicial equation*

$$a_0(1, 1)\alpha(\alpha - 1)\cdots(\alpha + r - 1) + \cdots + a_r(1, 1) = 0$$

has a root $\sigma > 0$ which is simple and such that all other roots $\alpha \neq \sigma$ satisfy $\Re(\alpha) < \sigma$.

C3. [Dominant growth] $f_n(1) \sim C \cdot b^{\sigma - 1}$ *for some* $C > 0$.

Then the coefficients of the polynomial $f_n(u)$ are asymptotically normal.

The conditions of Theorem 7 may seem outrageous. However, the spirit of the theorem is simple. If a bivariate generating function satisfies a linear differential equation with analytic coefficients, then a normal approximation derives from the existence of a *fixed regular singularity* (condition C1) provided there is no confluence of dominant singular solutions (condition C2) and the generating function exhibits the dominant growth regime (conditions C3). These conditions can be relaxed in various ways and, already, a particular case of confluence of roots modulo 1 had to be coped with in the case of even dimensions. The ε functions have a fundamental role in such developments.

Some supplementary conditions are certainly necessary in order to ensure asymptotic normality. For instance, the generating function

$$\frac{z}{(1-z)(1-uz)} = \sum_{n=1}^{\infty} \left(\sum_{k=0}^{n-1} u^k \right) z^n$$

corresponds to a uniform distribution while there is confluence of singularities at $z = 1$ and $z = 1/u$ as $u \to 1$, so that condition C1 is already violated.

Suitably general analytic schemas like

$$\frac{c(u)}{(1 - z/\rho(u))^{\alpha(u)}}$$

are otherwise known to lead to normal laws, see [14] and [17] generalizing an early work of Bender [2]. Such limit laws are thus likely to occur also in many cases where a *movable singularity* is encountered. This happens for node types and levels in varieties of increasing trees, in the context of a nonlinear differential equation [3]. Mahmoud and Pittel also derived normality results for the size of search trees with higher branching factors by considering a nonlinear equation of a different type, see [28] and Chapter 3 of [27].

In a related area, Drmota has introduced in [8] an interesting class of bivariate algebraic functions related to tree enumerations and independent sets that conduce to asymptotic normality. Jacquet and Régnier have obtained asymptotic normality for the size of digital "tries" from a nonlinear difference equation [21], [27] treated via Mellin transforms.

As a final word, we should thus expect many ordinary differential equations and functional equations arising from bivariate generating functions of combinatorics or the analysis of algorithms to lead to normal laws. General theorems in this area are certainly much desired.

Appendix

We briefly elaborate here on the main step of the proof of Lemma 3 in the case of an even dimension d, specializing the discussion to $d = 4$. Our aim is to justify (41).

A standard approach to regular singularity, in the case where confluence of eigenvalues modulo 1 occur, consists in reducing the system so that such confluent eigenvalues become multiple roots to which the general treatment (based on Jordan normal forms) applies.

The algebraic reduction lemma on p. 120 of [6] makes it possible to shift a designated eigenvalue by 1. In the case of $d = 4$, where $\lambda_0(1) - \lambda_2(1) = 4$, repeated application of the lemma yields a fundamental system of solutions of the form

$$P(u, z)\begin{pmatrix} 1 & \cdots & \cdots & \cdots \\ \vdots & 1 & \vdots & \vdots \\ \vdots & \vdots & 1 & \vdots \\ \cdots & \cdots & \cdots & (1-z)^4 \end{pmatrix}$$
$$\times \exp\left[\begin{pmatrix} \lambda_1(u) & \cdots & \cdots & \cdots \\ \vdots & \lambda_3(u) & \vdots & \vdots \\ \vdots & \vdots & \lambda_0(u) & 0 \\ \cdots & \cdots & b(u) & \lambda_2(u)+4 \end{pmatrix}\log\frac{1}{1-z}\right], \tag{44}$$

where $P(u, z)$ is analytic on $\mathbb{V} \times B(1, 1)$ and $b(u)$ is an analytic function for $u \in \mathbb{V}$. In a block decomposition the product

$$\begin{pmatrix} 1 & 0 \\ 0 & (1-z)^4 \end{pmatrix}\cdot\exp\left[\begin{pmatrix} \lambda_0(u) & 0 \\ b(u) & \lambda_2(u)+4 \end{pmatrix}\log\frac{1}{1-z}\right] \tag{45}$$

appears.

Now for a general triangular matrix, a simple computation shows that

$$\exp\begin{pmatrix} \mu_1 & 0 \\ b & \mu_2 \end{pmatrix} = \begin{pmatrix} e^{\mu_1} & 0 \\ b\dfrac{e^{\mu_2} - e^{\mu_1}}{\mu_2 - \mu_1} & e^{\mu_2} \end{pmatrix}$$

with the convention

$$\frac{e^{\mu_2} - e^{\mu_1}}{\mu_2 - \mu_1} = e^{\mu_1},$$

whenever $\mu_1 = \mu_2$. Thus,

$$\exp\left[\begin{pmatrix} \lambda_1(u) & \cdots & \cdots & \cdots \\ \vdots & \lambda_3(u) & \vdots & \vdots \\ \vdots & \vdots & \lambda_0(u) & 0 \\ \cdots & \cdots & b(u) & \lambda_2(u)+4 \end{pmatrix}\log\frac{1}{1-z}\right]$$
$$= \begin{pmatrix} \left(\dfrac{1}{1-z}\right)^{\lambda_0(u)} & 0 \\ b(u)\varepsilon_{\lambda_2(u)+4,\,\lambda_0(u)}(z) & \left(\dfrac{1}{1-z}\right)^{\lambda_2(u)} \end{pmatrix}.$$

From there, by elementary properties of ε, the matrix product of (45) transforms into

$$\begin{pmatrix} \left(\dfrac{1}{1-z}\right)^{\lambda_0(u)} & 0 \\ b(u)\varepsilon_{\lambda_2(u),\,\lambda_0(u)-4} & \left(\dfrac{1}{1-z}\right)^{\lambda_2(u)} \end{pmatrix}.$$

The general solution (44) thus admits the form

$$\Phi(u, z) = \sum_{i=0}^{3} c_i(u, z)\left(\frac{1}{1-z}\right)^{\lambda_i(u)} + d(u, z)\varepsilon_{\lambda_2(u),\,\lambda 0(u)-4}.$$

The function $b(u)$ being analytic for $u \in \mathbb{V}$, the end result (41), specialized to $d = 4$, follows.

The approach extends to the general case of any even dimension d, leading to

$$\Phi(u, z) = \sum_{j=0}^{d-1} c_j(u, z)\left(\frac{1}{1-z}\right)^{\lambda_j(u)} + \sum_{j\in J} d_j(u, z)\varepsilon_{\lambda_{k(j)}(u),\,\lambda_{l(j)}(u)-m(j)},$$

with the $c_j(u, z)$ and $d_j(u, z)$ analytic on $\mathbb{V} \times B(1, 1)$ as was claimed in step 7 of the proof of Lemma 3.

References

1. M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover, 1973. A reprint of the tenth National Bureau of Standards edition, 1964.
2. E. A. Bender. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory*, 15:91–111, 1973.
3. F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In J.-C. Raoult, editor, *CAAP '92*, pages 24–48 (Proceedings of the 17th Colloquium on Trees in Algebra and Programming, Rennes, France, February 1992). Lecture Notes in Computer Science, Volume 581. Springer-Verlag, Berlin, 1992.
4. P. Billingsley. *Probability and Measure*, 2nd edition. Wiley, New York, 1986.
5. G. G. Brown and B. O. Shubert. On random binary trees. *Mathematics of Operations Research*, 9(1):43–65, 1984.
6. E. A. Coddington and M. Levinson. *Theory of Ordinary Differential Equations.* McGraw-Hill, New York, 1955.
7. L. Devroye and L. Laforest. An analysis of random d-dimensional quad trees. *SIAM Journal on Computing*, 19:821–832, 1990.
8. M. Drmota. Asymptotic distributions and a multivariate Darboux method in enumeration problems. Manuscript, 1990.
9. R. A. Finkel and J. L. Bentley. Quad trees, a data structure for retrieval on composite keys. *Acta Informatica*, 4:1–9, 1974.
10. P. Flajolet, G. Gonnet, C. Puech, and J. M. Robson. Analytic variations on quadtrees. *Algorithmica*, 10:473–500, 1993.

11. P. Flajolet and A. M. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216–240, 1990.
12. P. Flajolet and C. Puech. Partial match retrieval of multidimensional data. *Journal of the ACM*, 33(2):371–407, 1986.
13. P. Flajolet and M. Soria. Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory, Series A*, 53:165–182, 1990.
14. P. Flajolet and M. Soria. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics*, 114:159–180, 1993.
15. J. Françon. Arbres binaires de recherche: Propriétés combinatoires et applications. *RAIRO Informatique Théorique*, 10(12):35–50, 1976.
16. J. Françon. On the analysis of algorithms for trees. *Theoretical Computer Science*, 4:155–169, 1977.
17. Z. Gao and L. B. Richmond. Central and local limit theorems applied to asymptotic enumerations, IV: Multivariate generating functions. *Journal of Computational and Applied Mathematics*, 41:177–186, 1992.
18. G. H. Gonnet and R. Baeza-Yates. *Handbook of Algorithms and Data Structures: in Pascal and C*, 2nd edition. Addison-Wesley, Reading, MA, 1991.
19. P. Henrici. *Applied and Computational Complex Analysis*, 3 volumes. Wiley, New York, 1977.
20. M. Hoshi and P. Flajolet. Page usage in a quadtree index. *BIT*, 32:384–402, 1992.
21. P. Jacquet and M. Régnier. Trie partitioning process: Limiting distributions. In P. Franchi-Zanetacchi, editor, *CAAP* '86, pages 196–210 (Proceedings of the 11th Colloquium on Trees in Algebra and Programming, Nice France, March 1986). Lecture Notes in Computer Science, Volume 214. Springer-Verlag, Berlin, 1986.
22. D. E. Knuth. *The Art of Computer Programming*, Volume 3. Addison-Wesley, Reading, MA, 1973.
23. L. Laforest. Étude des arbres hyperquaternaires. Technical Report 3, LACIM, UQAM, Montreal, Nov. 1990. (Author's Ph.D. Thesis at McGill University.)
24. G. Louchard. Exact and asymptotic distributions in digital and binary search trees. *RAIRO Theoretical Informatics and Applications*, 21(4):479–495, 1987.
25. E. Lukacs. *Characteristic Functions*. Griffin, London, 1970.
26. W. C. Lynch. More combinatorial problems on certain trees. *Computer Journal*, 7:299–302, 1965.
27. H. Mahmoud. *Evolution of Random Search Trees*. Wiley, New York, 1992.
28. H. M. Mahmoud and B. Pittel. Analysis of the space of search trees under the random insertion algorithm. *Journal of Algorithms*, 10:52–75, 1989.
29. H. Samet. *Applications of Spatial Data Structures*. Addison-Wesley, Reading, MA, 1990.
30. H. Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, MA, 1990.
31. R. Sedgewick. *Algorithms*, 2nd edition. Addison-Wesley, Reading, MA, 1988.
32. W. Wasow. *Asymptotic Expansions for Ordinary Differential Equations*. Dover, 1987. A reprint of the Wiley edition, 1965.
33. E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*, 4th edition. Cambridge University Press, Cambridge, 1927. Reprinted 1973.

Received March 9, 1993, and in revised form January 10, 1994.