PATTERNS IN RANDOM BINARY SEARCH TREES

Philippe Flajolet[1], Xavier Gourdon[1], and Conrado Martínez[2]

[1] Algorithms Project, INRIA-Rocquencourt,
F-78153 Le Chesnay, France

[2] Departament de llenguatges i sistemes informàtics
Universitat Politècnica de Catalunya, Pau Gargallo, 5,
E-08028 Barcelona, Spain

October 6, 1996

Abstract

In a randomly grown binary search tree (BST) of size n, any fixed pattern occurs with a frequency that is on average proportional to n. Deviations from the average case are highly unlikely and well quantified by a Gaussian law. Trees with forbidden patterns occur with an exponentially small probability that is characterized in terms of Bessel functions. The results obtained extend to BSTs a type of property otherwise known for strings and combinatorial tree models. They apply to paged trees or to quicksort with halting on short subfiles. As a consequence, various pointer saving strategies for maintaining trees obeying the random BST model can be precisely quantified. The methods used are based on analytic models, especially bivariate generating functions subjected to singularity perturbation asymptotics.

1 Introduction

The model of randomly grown binary search trees, hereafter called the BST model, is of interest in the analysis of binary search trees, their randomized versions —like treaps [1, 36] or rBSTs [38]—, and k-d-trees for multidimensional search [30, 33, 40]. By a standard equivalence principle, this model also applies to heap-ordered trees for priority queue maintenance [40, 43], to tree representations of permutations [4, 40, 42], as well as to quicksort [30, 40, 42]. In addition, empirical studies by software engineers suggest that the BST model is perhaps more adequate for syntax trees and term trees than the common combinatorial model where all trees of a given size are taken with equal likelihood; a plausible reason is that the combinatorial model tends to produce trees that are often too "skinny" to model closely trees that occur naturally in this context (Gilles Kahn, private communication, 1994).

In abstract terms, the BST model produces for each size n a random plane binary tree which consists of an internal node (the root) connected to a left subtree of size K and a right subtree of size $n-1-K$. There K is a random variable uniformly distributed over its range,

$$\Pr\{K=k\} = \frac{1}{n}, \qquad k = 0, 1, \ldots, n-1, \tag{1}$$

and the subtrees recursively obey the BST model. By design, the model of BSTs applies to two closely related types of trees built on random permutations:

— the binary search tree where the first element of the permutation is placed at the root, with elements smaller and larger than the root going respectively to left and right root subtrees;

— the increasing binary tree (also called "heap-ordered" tree) where the smallest element is placed at the root, with elements left and right of the minimum going respectively to left and right root subtrees.

We refer to [30, 33, 40, 42] for basic combinatorial and probabilistic properties; for instance, the trees produced have expected height $\sim 4.31107\log n$, their path length is $\sim 2n\log n$ so that a search costs $\sim 2\log n$ on average, and the number of leaves (nodes with both descendents external) is $\sim n/3$ on average.

In this paper we investigate fine characteristics of the shape of trees produced by the BST model. Given a fixed binary tree u, called the *pattern*, we examine the *number of occurrences* $\omega_u[t]$ of u as a subtree[1] of a larger tree t called the *text*. Taken over a random BST of size n, this random variable has an expectation that is $\sim c(u)\cdot n$ for some effectively determined constant $c(u)$, its standard deviation is $\mathcal{O}(\sqrt{n})$, and its distribution is found to be asymptotically normal. Thus, an overwhelming majority of trees will behave closely like what the average case analysis predicts. Trees *not* containing pattern u, are found to have an exponentially vanishing probability, where the exponential rate is characterized in terms of zeros of Bessel function equations. (See [35] for a first connection between Bessel functions and occurrence counts.)

These phenomena are analogous to what happens in random binary strings where a pattern u has on average $\sim n/2^{|u|}$ occurrences, with a companion Gaussian law. Strings with excluded patterns (for instance no sequence of 3 identical characters in a row) have exponentially small probability, with the exponential rates being given as roots of correlation polynomials [15, 24, 40, 45]. Similar properties hold true for the combinatorial tree model as established in [18, 41].

[1]As usual, a subtree of a tree t is defined by a node of t together with *all* its descendents. We are thus counting here occurrences of "terminal" subtrees.

Such analyses, apart from being of combinatorial interest, are relevant to efficient storage representations of trees. A straightforward implementation of a binary tree of size n involves a total of $2n$ pointers amongst which $n+1$ are attached to an external node and are void of information content. By distinguishing between the 4 basic types of possible nodes —binary, left unary, right unary, and leaves— one obtains a representation of trees that only requires $n-1$ pointers. Pushing the idea further leads to pointer-free representations for small subtrees occurring at the fringe of the tree. For instance, our analysis shows that the number of pointers is reduced to $\sim 4n/5$ when fringe subtrees of the 3 types

```
    o        o     o
   / \      /       \
  o   o    o         o
```

are encoded as special pointer-free nodes. (Similar ideas are often used to obtain compacted form of digital trie dictionaries, see for instance [37].)

Pushed to its limits, the pointer-free representation gives rise to the directed acyclic graph (DAG) representation, also known in parsing and compiling as common subexpression factoring [18]. Naturally, this technique would not apply directly to BSTs, where internal nodes contain important informations, but, with minor adjustments, it is applicable to parse trees statistically governed by the BST model. We show here that the expected size K_n of the DAG associated to a BST of size n is of order at most $n/\log n$. This result contrasts with the corresponding estimate of $\mathcal{O}(n/\sqrt{\log n})$ under the combinatorial model [18] and it confirms that trees obeying the BST model are better suited for compaction.

Another consequence of the methods developed here is a distributional analysis of the storage requirements of paged BSTs, or equivalently of the number of recursive calls of quicksort under the strategy of halting on subfiles of size less than a fixed threshold b.

Some of the distributional results of Section 5 are quite similar in spirit to theorems obtained by Devroye [10]. Devroye develops a general framework for the study of local order patterns in random permutations that is based on the central limit theorem of probability theory extended to random variables with restricted dependencies. In particular, Devroye's approach yields central limit laws for leaves and for nodes with k descendents (compare with our Thm. 5). The approaches of this paper and of [10] are complementary. Devroye's method applies more naturally to problems expressed on the linear representation of permutations, especially local order patterns of close neighbours. Our method, being based on the tree decomposition, seems more suitable for recursively defined parameters of the tree structure, like occurrences of subtrees and paging. In addition, it may give local limit laws (Thm. 6) and quantify rare events (Thm. 2,3) as well as convergence rates (Thm. 4,5).

2 Mean, variance, and generating function

The probability $\lambda(u)$ for a given unlabelled tree u to be the shape of a random BST of size $|u|$ is well-known to satisfy [30, 40]

$$\lambda(u) = \prod_{v \prec u} \frac{1}{|v|},$$

where the product is over all subtrees v of u. The parameter λ has been studied in detail by Fill [14]. All analyses relative to the occurrences of pattern u involve crucially $\lambda = \lambda(u)$ as well as the pattern size $m = |u|$. We will often omit the dependence on u in parameters or generating functions.

Let $\omega[t] \equiv \omega_u[t]$ denote the number of occurrences of pattern u as a subtree of the BST t (possible labels of t are not taken into account). Then, one has the obvious recurrence:

$$\omega[t] = [\![t = u]\!] + \omega[t_0] + \omega[t_1],$$

where t_0, t_1 denote the left and right subtrees of t and where the bracket notation $[\![P]\!]$ is the indicator of P with value 1 if the predicate P is true and 0 otherwise. The *bivariate generating function* (BGF) $F(z, y)$ defined by

$$F(z, y) := \sum_t \lambda(t) y^{\omega[t]} z^{|t|}$$

is such that the coefficient $[z^n y^k]F(z, y)$ represents the probability that a random BST of size n has k occurrences of u. We have:

Lemma 1 *The bivariate generating function $F(z, y)$ satisfies the Riccati equation*

$$\frac{\partial}{\partial z} F(z, y) = F^2(z, y) + (y - 1)\lambda(u)|u|z^{|u|-1}, \qquad F(0, y) = 1. \tag{2}$$

Proof. The tree function $y^{\omega[t]}$ satisfies the recursive relation

$$y^{\omega[t]} = y^{[\![t=u]\!]} y^{\omega[t_0]} y^{\omega[t_1]},$$

which means that $y^{\omega[t]}$ is multiplicative over subtrees, except in the single case when $t = u$ for which $y^0 = 1$ should be replaced by $y^1 = y$. By the shape of the splitting probabilities —check directly by means of recurrences implied by (1), or see [40, 42] for more general approaches— this gives the integral equation

$$F(z, y) = 1 + \int_0^z F^2(w, y)\, dw + (y - 1)\lambda(u)z^{|u|}.$$

The statement then follows by differentiation with respect to z. □

Mean and variance. By a classical process, the moments of the number of occurrences of u are obtained by successive differentiation of the BGF $F(z,y)$ with respect to y, upon setting $y=1$. The easy computation is summarized by the following statement that is of folklore knowledge.

Theorem 1 (Moments of occurrences) *The number Ω_n of occurrences of a pattern u of size m in a random BST of size n has mean $\mu_n = E\{\Omega_n\}$ and variance $\sigma_n^2 = Var\{\Omega_n\}$ that satisfy*

$$\mu_n \sim \frac{2\lambda}{(m+1)(m+2)} \cdot n.$$

$$\sigma_n^2 \sim \left[\frac{2\lambda}{(m+1)(m+2)} - \frac{2\lambda^2(11m^2+22m+6)}{(m+1)^2(m+2)^2(2m+1)(2m+3)}\right] \cdot n,$$

where $\lambda = \lambda(u)$ is the probability of a BST with shape u.

Proof. We have

$$\mu_n = [z^n] \left.\frac{\partial F(z,y)}{\partial y}\right|_{y=1}, \quad \sigma_n^2 = \phi_n + \mu_n - \mu_n^2,$$

where ϕ_n is the second factorial moment,

$$\phi_n = \frac{1}{2}[z^n] \left.\frac{\partial^2 F(z,y)}{\partial y^2}\right|_{y=1}.$$

By differentiation with respect to y of the basic equation (2), the ordinary generating functions $M(z) := \sum_n \mu_n z^n$ and $\Phi(z) := \sum_n \phi_n z^n$ satisfy the first order differential equations:

$$M'(z) = \frac{2}{1-z}M(z) + \lambda m z^{m-1}, \qquad \Phi'(z) = \frac{2}{1-z}\Phi(z) + 2M^2(z).$$

These equations are a priori solvable by quadrature through the variation of constant method; both functions turn out to be rational fractions with a pole at $z=1$ of respectively 2nd and 3rd order. For dominant asymptotics of σ_n^2, one needs accordingly 2-term and 3-term expansions of $M(z)$ and $\Phi(z)$ near the singularity $z=1$. The computations are easily completed with the help of the symbolic manipulation system Maple. □

In particular, a random BST of size n has on average $\sim n/3$ nodes that are leaves, hence $\sim n/3$ binary nodes, a well-known fact. This indicates a better balancing for BSTs than for trees under the combinatorial model where these quantities are $\sim n/4$, see for instance [40].

Bessel function solutions. We now proceed with an explicit solution of the Riccati differential equation (2). Solutions to such a nonlinear equation are always reducible to quotients of solutions of second order linear differential equations. We thus perform the basic change of variables

$$F(z,y) = -\frac{w'(z,y)}{w(z,y)}, \tag{3}$$

where $w'(z,y) = w'_z(z,y)$ is the partial derivative with respect to z. The Riccati equation of Lemma 1 induces the second order equation

$$\frac{\partial^2}{\partial z^2} w(z,y) - \Lambda z^{m-1} w(z,y) = 0, \qquad \Lambda = \lambda \cdot m \cdot (1-y)$$

with, again, $\lambda = \lambda(u)$, $m = |u|$. This equation is readily solved by indeterminate coefficients: with $w_n = [z^n]w(z)$, we have the recurrence:

$$(n+m)(n+m+1)w_{n+m+1} = \Lambda w_n.$$

Introduce now the two functions

$$\begin{aligned} A_m(z) &= 1 + \frac{z}{m(m+1)} + \frac{z^2}{m(m+1)(2m+1)(2m+2)} + \cdots \\ B_m(z) &= 1 + \frac{z}{(m+1)(m+2)} + \frac{z^2}{(m+1)(m+2)(2m+2)(2m+3)} + \cdots, \end{aligned} \tag{4}$$

so that any solution to the linear equation is a linear combination of $A_m(z)$ and $zB_m(z)$. These are normalized Bessel functions [46] of orders $-1/(m+1)$ and $1/(m+1)$ respectively. In effect, defining

$$\mathcal{J}_a(z) = \sum_{r=0}^{\infty} \frac{z^r}{r!\,(a+1)(a+2)\cdots(a+r)},$$

one has

$$A_m(z) = \mathcal{J}_{-1/(m+1)}\left(\frac{z}{(m+1)^2}\right), \qquad B_m(z) = \mathcal{J}_{1/(m+1)}\left(\frac{z}{(m+1)^2}\right).$$

The initial conditions arising from the combinatorics of the problem entail that $-w'(z,y)/w(z,y) = 1 + z + \mathcal{O}(z^2)$ at the origin (for $|u| \geq 2$), and one may impose additionally $w(z) = 1 + \mathcal{O}(z)$, which fully determines $w(z)$. Hence:

Lemma 2 *The bivariate generating function of the number of occurrences of pattern u is given by*

$$F(z,y) = -\frac{w'(z,y)}{w(z,y)}, \qquad w(z,y) = A_m(\Lambda z^{m+1}) - zB_m(\Lambda z^{m+1}), \tag{5}$$

with $w'(z,y) = w'_z(z,y)$, $\Lambda = |u|\lambda(u)(1-y)$, $m = |u| \geq 2$, and A_m, B_m are normalized Bessel functions given by (4).

3 Analysis of generating functions

For the reader's convenience, we summarize here the basic principles of complex analysis that are required in our subsequent analyses of the BST model.

The generating functions that occur throughout are *meromorphic*, in the main variable z, which means that they are quotients of analytic (holomorphic) functions. This fact is a direct reflection of the recursive binary form of the BST model, which leads to Riccati equations, hence to quotients of analytic functions by the basic linearizing transformation of (3). It is then well-known that the location of polar singularities of a function dictates the asymptotic form of its coefficients: this is a simple consequence of Cauchy's coefficient formula detailed in Henrici's book [25] and briefly recalled when we first encounter it in Eq. (8). This technique falls into the broad class of singularity analysis methods, and is used in the univariate case for proving Theorems 2,3.

Our interest here is largely with multivariate asymptotics, where it is required to extract information on coefficients

$$f_{n,k} = [z^n y^k] F(z, y),$$

of a bivariate generating function $F(z, y)$, like in Lemma 2. As is common practice, we first perform one level of inversion, resulting in estimates of

$$f_n(y) = [z^n] F(z, y),$$

for some values of y. At this stage, the problem of estimating $f_n(y)$ belongs to univariate asymptotics but is parameterized by y. Singularity analysis techniques allow for uniform error bounds, a crucial feature for probability estimates.

The quantity $f_n(y)$ is by construction a probability generating function and one more inversion is required in order to recover the individual probabilities $f_{n,k}$ as

$$f_{n,k} = [y^k] f_n(y).$$

The most direct approach consists in appealing to *Levy's continuity theorem* for characteristic functions; this implies estimating $f_n(y)$ for $y = e^{i\theta}$, but only θ near 0 is required because of scaling. Thus, we have a *perturbation* of the univariate problem at $y = 1$. It turns out that $f_n(y)$ is a "quasi-power", meaning that it behaves very nearly like the powers of a fixed function, so that the scaled version of $f_n(y)$ behaves like the characteristic function of a Gaussian variable, see Eq. (13,14). In this way, a central limit law is derived in Theorems 4–5. Analytically, the process is then essentially equivalent to Fourier inversion. Additional results derive if one can estimate *globally* $f_n(y)$ by quasi-powers in larger regions, like $|y| = 1$, and not merely *locally* near $y = 1$. In that case, the recovery of $f_{n,k}$ from $f_n(y)$ is achieved by subjecting a Cauchy coefficient integral,

$$f_{n,k} = \frac{1}{2i\pi} \oint f_n(y) \frac{dy}{y^{k+1}},$$

to the saddle point method. Large powers and quasi-powers are known to lead to local limit laws of the Gaussian type, and Theorem 6 is an instance of this method.

In order to carry out this program, one must analyse the way the poles of F depend on y. By the basic linearizing transformation, this requires analysing the behaviour of roots of various sorts of equations

$$S(z, y) = 0,$$

where S is analytic in both variables z, y; see Lemma 2 with $S \equiv w$.

For this category of problems, we refer to the very clear treatment by Hille in [26]. The Weierstrass preparation theorem and the implicit function theorem (see Section 9.4 of [26]) assert the following: if, for at y_0, the equation $S(z, y_0) = 0$ has an mth order root in z at $z = z_0$, then there exist also m roots $\{\zeta_j\}_{j=1}^{m}$ of $S(\zeta, y)$ that are near z_0 when y is sufficiently near to y_0. Furthermore, these local roots are algebraic functions, in the sense that they satisfy

$$\zeta^m + g_1(y)\zeta_{m-1} + \cdots + g_m(y) = 0,$$

for some functions $g_j(y)$ analytic at y_0 with $g_j(y_0) = 0$. Thus, functions defined implicitly by bivariate analytic equations have a locally predictable behaviour. We refer again to [26, p. 265–275] for details. In particular, there is no "spontaneous" appearance of roots of analytic equations $S(z, y) = 0$, as these roots vary continuously on the Riemann sphere. Consequently, poles (in the z-plane) of functions like $F(z, y)$ have a dependence on the parameter y that is of an algebraic form (and governed by Puiseux expansions [27, Sec. 12.3]). We make use of these properties throughout Sections 5–7.

4 Trees with excluded patterns

We now estimate the probability that a tree does not contain a given pattern u. This problem is of combinatorial interest as it corresponds to enumerating permutations with certain types of forbidden patterns, given the equivalence between the BST model, heap-ordered trees, and permutations. More importantly, the analysis paves the way for the distributional results of the next section.

Asymptotic analysis of univariate and bivariate generating functions derived from $F(z, y)$ depends on locating the zeros of $w(z, y)$ where y is a parameter. We thus define the function $\alpha_u(y)$ to be the root of smallest modulus of the Bessel type equation

$$A_m(\Lambda\alpha^{m+1}) - zB_m(\Lambda\alpha^{m+1}), \qquad \Lambda = (1 - y)|u|\lambda(u). \tag{6}$$

This definition specifies $\alpha_u(y)$ unambiguously, for $|1 - y|$ not too large, as is shown by the following lemma.

Lemma 3 *For any constant c, $|c| < \frac{5}{2}$, and any pattern u of size $m = |u| > 4$, the equation*

$$A_m(c|u|\lambda(u)z^{|u|+1}) - zB_m(c|u|\lambda(u)z^{|u|+1}) = 0 \tag{7}$$

admits exactly one root in the domain $|z| \leq \frac{11}{10}$.

l

Proof. The idea of the proof is that, since $\lambda(u)$ is small, the equation is a small perturbation of $1 - z = 0$ corresponding to the first two Taylor terms of (5), so that Rouché's theorem [26] applies.

First, a uniform exponential upper bound on $\lambda(u)$ as a function of $m = |u|$ is needed. Asymptotically, it is known that $\lambda(u)$ decays at least exponentially with $|u|$, see [14], and it is not hard to see that for all $m \geq 4$, one has:

$$|u|\lambda(u) \leq 2^{-|u|/4}.$$

(The actual asymptotic growth of $|u|\lambda(u)$ is exponentially smaller than the upper bound above, so that any rough asymptotic analysis complemented by exhaustive verifications for small m suffices to establish such a bound.)

Then the proof is completed by decomposing the left handside of (7) as $1 - z + R(z)$ where $R(z)$ is the sum of terms of degree > 2. On $|z| \leq \frac{11}{10}$, the quantity $R(z)$ is majorized by the sum of a geometric progression, and a simple computation shows that $|R(z)| < |1 - z|$ on $|z| = \frac{11}{10}$. By Rouché's theorem, the number of solutions of the full equation (7) is the same as that of $1 - z = 0$ inside $|z| \leq \frac{11}{10}$. $\square$

The proof above leaves aside six cases corresponding to the six types of trees of sizes $2, 3, 4$ based on the values of $\langle |u|, \lambda(u)\rangle$. These cases are exhausted by applying Rouché's theorem to Taylor truncations of low degrees. For instance, for $c = 1$, the roots of smallest modulus (also called "dominant" roots) of the characteristic equation (7) are well separated from subdominant roots and given by the following table,

$m = 2$	$(o, (o, \square, \square), \square)$	1.10732
$m = 3$	$(o, (o, (o, \square, \square), \square), \square)$	1.01762
	$(o, (o, \square, \square), (o, \square, \square))$	1.03748

with values $1.00861, 1.00567, 1.00280$ for size $m = 4$.

Theorem 2 (Excluded patterns) *The probability $e_{u,n}$ that a random BST of size n does not contain the pattern u satisfies*

$$e_{u,n} = \alpha_u(0)^{-n-1}(1 + \mathcal{O}(K^{-n}))$$

where K is a constant strictly larger than 1, and $\alpha_u(0)$ is the smallest positive root of Equation (6) with $y = 0$, namely:

$$A_m(|u|\lambda(u)\alpha^{|u|+1}) - zB_m(|u|\lambda(u)\alpha^{|u|+1}) = 0.$$

Proof. The probability is by definition $[z^n]F(z,0)$. By Lemma 3, the function $F(z,0)$ has a unique dominant singularity that is a simple pole, at $\alpha_u(0)$ that is positive and satisfies $\alpha_u(0) \le \frac{11}{10}$ for $|u| > 4$. (The cases $|u| \le 4$ are covered by the remarks following Lemma 3.) The function $-F$ is a logarithmic derivative, so that its residue at $\alpha_u(0)$ equals 1. Therefore,

$$F(z,0) = \frac{1}{\alpha_u(0) - z} + R(z),$$

where $R(z)$ is analytic in a circle of radius strictly larger than $\alpha_u(0)$.

The result follows by singularity analysis of meromorphic functions, according to a classical process. For small enough ϵ, ϵ', one has

$$\begin{aligned} [z^n]F(z,0) &= \frac{1}{2i\pi}\int_{|z|=\epsilon} F(z,0)\,\frac{dz}{z^{n+1}} \\ &= \alpha_u(0)^{-n-1} + \frac{1}{2i\pi}\int_{|z|=(1+\epsilon')\alpha_u(0)} F(z,0)\,\frac{dz}{z^{n+1}} \\ &= \alpha_u(0)^{-n-1} + \mathcal{O}(K^{-n}\alpha_u(0)^{-n-1}), \end{aligned} \tag{8}$$

by Cauchy's coefficient formula, the residue theorem, and the triangular inequality. □

The same argument proves that, for small enough $|y|$,

$$[z^n]F(z,y) \sim \alpha_u(y)^{-n-1}, \tag{9}$$

with a uniform exponentially small error term. Now, the probability that a random BST of size n has k occurrences of a pattern u is obtained by differentiating k times:

$$\Pr\left\{\omega_u[t] = k \;\middle|\; |t| = n\right\} = \frac{1}{k!}\left(\frac{d^k}{dy^k}[z^n]\,F(z,y)\right)_{y=0}.$$

Differentiation of asymptotic expansions like (9) is valid for analytic functions, so that the probability of k occurrences, for any *fixed* k satisfies

$$\Pr\left\{\omega_u[t] = k \;\middle|\; |t| = n\right\} \sim P_u^{\langle k\rangle}(n)\alpha_u^{-n-1}.$$

There, $P_u^{\langle k\rangle}$ is a polynomial of degree k whose coefficients depend on u by way of values of α_u and its derivatives at 0. Retaining only dominant asymptotics leads to the following theorem.

Theorem 3 (Poisson law for rare occurrences) *Given a fixed pattern u, for each* fixed *k, one has as $n \to +\infty$:*

$$\Pr\left\{\omega_u[t] = k \;\middle|\; |t| = n\right\} = \alpha_u(0)^{-n-1}\cdot\frac{(\mu n)^k}{k!}\left(1 + \mathcal{O}(\frac{1}{n})\right), \qquad \mu = -\frac{\alpha_u'(0)}{\alpha_u(0)}. \tag{10}$$

For small number of occurrences, the asymptotic probability of this rare event is thus the product of a Poisson law of parameter μn and of an exponentially small scaling factor.

Permutations. Theorem 2 is in line with known enumerations of permutations with excluded patterns, a fact to be expected since the BST model is also isomorphic to the model of heap-ordered (*i.e.*, increasing) trees that itself bijectively correspond to permutations [4, 40, 43]. For instance the exponential generating function of alternating (or "up-down") permutations [7] is

$$A(z) = \tan z = \frac{\sin z}{\cos z},$$

also the negative of a logarithmic derivative, satisfying the Riccati equation $A' = 1 + A^2$. Accordingly, the function $A(z)$ is meromorphic and the probability for a permutation to be alternating decays like $(\pi/2)^{-n}$, an estimate of the same form as Theorem 2; as alternating permutations correspond to trees that exclude unary branching nodes, this is a type of excluded pattern that is only marginally different from the ones considered here.

5 Gaussian limit laws

Finding the asymptotic distribution of the number of occurrences of a fixed pattern belongs to bivariate asymptotics. The starting point of our approach is the bivariate generating function $F(z, y)$. The method consists in analysing the meromorphic function $F(z, y)$ and its z-coefficients in the vicinity of $y = 1$ by a technique of singularity perturbation. In this way, one proves that the *probability generating function* (PGF) of the number of occurrences is, near $y = 1$, well-approximated by a large power of the fixed function $\alpha_u(y)$, like in Eq. (9). In the "pure" case of exact powers, this situation yields a Gaussian limit distribution, in accordance with the central limit theorem of probability theory. Here, PGFs obey a general scheme of "quasi-powers" already studied by Bender, Richmond, Hwang, and others [3, 19, 21, 29].

Suitable adaptations of the technique also lead to a distributional analysis of paging, where the bivariate GF is only known *implicitly* through differential equations. A local limit law for leaves is also proved by means of typical saddle point arguments.

5.1 Pattern occurrences

The first result, also called a *central limit law* (of the Gaussian type), describes the probability of deviating more than a certain number of standard deviations from the mean in terms of the Gaussian error function.

> **Central Limit Law** (CLL): a sequence of random variables Ω_n with mean μ_n and standard deviation σ_n satisfies a central limit law if
>
> $$\sup_{x \in R} \left| \Pr\left\{ \frac{\Omega_n - \mu_n}{\sigma_n} \leq x \right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw \right| < \epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to +\infty$.

An upper bound ϵ_n is called a *speed of convergence* to the central limit. Clearly, a CLL is equivalent to a Gaussian approximation for the partial sums $\sum_{j\leq k} f_{n,j}$, where $f_{n,j} = \Pr\{\Omega_n = j\} = [z^n y^j]F(z,y)$.

Theorem 4 (Central law for occurrences) *Given a pattern u, the number of occurrences Ω_n of u in a random BST of size n obeys a central limit law (LLL) with speed of convergence $\mathcal{O}(1/\sqrt{n})$. The mean and variance μ_n, σ_n^2 given by Theorem 1.*

Proof. The quantity $f_n(y) = [z^n]F(z,y)$ is the probability generating function (PGF) of Ω_n, that is to say $f_n(y) = E\left\{y^{\Omega_n}\right\}$. The semi-normalized variable $\Omega_n^* = (\Omega_n - \mu_n)/\sqrt{n}$ then has characteristic function equal to

$$\xi_n(t) = E\left\{e^{it\Omega_n^*}\right\} = e^{-i\mu_n t/\sqrt{n}} f_n(e^{it/\sqrt{n}}). \tag{11}$$

By Lévy's continuity theorem for characteristic functions [5, 32], it is enough to prove pointwise convergence of $\xi_n(t)$ (for any fixed t) to the characteristic function of a Gaussian variable as $n \to \infty$. In that case, the argument of $f_n(y)$ lies in a complex neighbourhood of 1 that is of vanishingly small radius. Thus, only a local analysis of $f_n(y)$ near 1 is needed.

The analysis of trees with excluded patterns applies almost *verbatim* in this context. By the implicit function theorem and the preparation theorem of Weierstrass (see the previous section), there is a small complex neighbourhood of 1 such that the function $\alpha_u(y)$ is analytic. In such a small neighbourhood, we have, by the analysis of meromorphic functions and by Lemma 3,

$$f_n(y) = \alpha_u(y)^{-n-1}(1 + \mathcal{O}(K^{-n})). \tag{12}$$

In other words, $f_n(y)$ is closely approximated by a large power of a fixed function, a situation conducing to normal laws.

Combining (11,12), we get

$$\log \xi_n(t) = -it\mu_n n^{-1/2} - (n+1)\log \alpha_u(e^{itn^{-1/2}}) + \mathcal{O}(K^{-n}). \tag{13}$$

We have $\alpha_u(1) = 1$, so that, from (13), as $n \to \infty$,

$$\log \xi_n(t) = \frac{-it}{\sqrt{n}}(\mu_n + n\alpha_u'(1)) + \frac{t^2}{2}\left(\alpha_u''(1) + \alpha_u'(1) - (\alpha_u'(1))^2\right) + \mathcal{O}(n^{-1/2}). \tag{14}$$

The derivatives $\alpha_u'(1), \alpha_u''(1)$ are readily computed from the bivariate Taylor expansion of the entire function $w(z,y)$ at $(z,y) = (1,1)$. It is found that $\alpha_u'(1) = -\lim_{n\to\infty} \mu_n/n$, so that

$$\log \xi_n(t) = -s^2\frac{t^2}{2} + \mathcal{O}(n^{-1/2}),$$

for a constant s that also equals $\lim_{n\to\infty} \sigma_n/\sqrt{n}$ and is expressible in terms of $\alpha'_u(1), \alpha''_u(1)$. This implies that the variable $\frac{1}{s}\Omega_n^*$ converges in distribution to a standard normal variable. In passing, the computation provides an independent check of the variance computation done earlier in Theorem 1.

Following Feller and Hwang [13, 29], it is also possible to bound the speed of convergence to the Gaussian limit by means of the Berry-Esseen inequalities that relate the distance between distribution functions and characteristic functions [13, 32]. Let F, G be two distribution functions with characteristic functions $\phi(t), \gamma(t)$, G being assumed to have a density G', and let $||h||_\infty$ be the sup norm, $||h||_\infty = \sup_{x\in R} |h(x)|$. Then, the Berry-Esseen inequalities state that

$$||F - G||_\infty \leq \frac{24}{\pi}\frac{||G'||_\infty}{T} + \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\phi(t)-\gamma(t)}{t}\right| dt, \tag{15}$$

for any $T > 0$. Now, the main estimate (13) applies with a uniform error term of the form $\mathcal{O}(t^3 n^{-1/2})$ provided $|t| \leq c_1 n^{1/2}$, where c_1 is a positive constant whose value is dictated by the radius of the analyticity region of $\alpha_u(y)$ at 1. Taking then $T = c_1 n^{1/2}$ in (15) entails that the speed of convergence in the central limit law is $\mathcal{O}(1/\sqrt{n})$. (See also Hwang's work [29] for an interesting analytic framework of considerable generality.) □

5.2 Paged trees

Fix a "bucket size" parameter $b \geq 2$. Given a tree t, its *b-index* is a tree that is constructed by retaining only those internal nodes of t which correspond to subtrees of size $> b$. Such an index is well-suited to "paging", where one has a two-level hierarchical memory structure: the index resides in main memory and the rest of the tree is kept in pages of capacity b on peripheral storage, see for instance [28, 33]. We let $\iota[t] = \iota_b[t]$ denote the size —number of nodes— of the b-index of t. The analysis is then clearly equivalent to determining the total number of occurrences of all patterns of size $> b$, or dually those of size $\leq b$.

Theorem 5 (Paging distribution) *For fixed $b \geq 2$, the size I_n of the b-index constructed on a random BST of size n has average μ_n and variance σ_n^2 that satisfy*

$$\mu_n = \frac{2(n+1)}{b+2} - 1, \qquad \sigma_n^2 = \frac{2}{3}\frac{(b-1)b(b+1)}{(b+2)^2}(n+1). \tag{16}$$

The random variable I_n obeys a central limit law with speed of convergence $\mathcal{O}(1/\sqrt{n})$.

Proof. Like in Lemma 1, the bivariate generating function

$$G(z,y) := \sum_t \lambda(t) y^{\iota[t]} z^{|t|}$$

satisfies a Riccati equation that reflects the root decomposition of trees,

$$\frac{\partial}{\partial z}G(z,y) = yG^2(z,y) + (1-y)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right). \tag{17}$$

Mean and variance follow by differentiation at $z = 1$, like in the case of Theorem 1. (The result for the mean is well-known, refer to quantity A_n in the analysis of quicksort on p. 122 of [30].)

Multiplying both sides of (17) by y now gives an equation satisfied by $H(z,y) := yG(z,y)$,

$$\frac{\partial}{\partial z}H(z,y) = yH^2(z,y) + y(1-y)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right),$$

that may as well be taken as a starting point since $H(z,y)$ is the bivariate GF of parameter $1+\iota_b$ (a quantity also equal to the number of external pages). In order to apply the linearization transformation, one sets

$$H(z,y) = -\frac{X'_z(z,y)}{X(z,y)},$$

so that

$$\frac{\partial^2}{\partial z^2}X(z,y) + y(y-1)A(z)X(z,y) = 0, \qquad A(z) = \frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right), \tag{18}$$

with $X(0,y) = 1, X'_z(0,y) = -y$. By the classical existence theorem of Cauchy, the solution of (18) is an entire function of z for each fixed y, as the linear differential equation has no singularity at a finite distance. Furthermore, the dependency of X on y is also locally everywhere analytic; see the remarks of [44, Sec. 24], for which a proof derives by inspection of the classical existence proof based on indeterminate coefficients and majorant series. Thus, $X(z,y)$ is actually an entire function of *both* complex variables z and y. As a consequence, for any fixed $y = y_0$, the function $H(z,y_0)$ is a meromorphic function of z whose coefficients are amenable to singularity analysis.

In order to proceed further, we need to prove that, in a sufficiently small neighbourhood of $y = 1$, $X(z,y)$ has only one simple root, corresponding for $H(z,y)$ to a unique dominant and simple pole. This fact itself derives from the Preparation Theorem of Weierstrass (see the discussion in the previous section): *in the vicinity of any point (z_0,y_0) with $X(z_0,y_0) = 0$, the roots of the bivariate analytic equation $X(z,y) = 0$ are locally branches of an algebraic function.* Here, we have $X(z,1) \equiv 1-z$. Thus, as y tends to 1, all solutions of $X(z,y)$ must escape to infinity except for one branch $\beta(y)$ that satisfies $\beta(1) = 1$. By the nonvanishing of $X'_y(z,1)$ and the implicit function theorem, the function $\beta(y)$ is additionally an analytic function about $y = 1$.

The argument is now completed like in the proof of Theorem 4. We have, for y in a sufficiently complex neighbourhood of 1,

$$[z^n]H(z,y) = \beta(y)^{-n-1}\left(1 + \mathcal{O}(K^{-n})\right),$$

for some fixed constant $K > 0$. Thus, the probability GF of $\iota_b[t]$ over trees of size n is asymptotic to a large power, and by the computation of (11–14), the Gaussian limit results. □

Sets of patterns. Note that the proof architecture is robust enough to survive the disappearance of explicit Bessel function solutions. For this reasons, it applies in full generality to *any finite collection of patterns*, S. In this way, one can prove the following result: *the occurrence count*

$$\omega_S[t] := \sum_{s \in S} \omega_s[t],$$

under the BST model of size n, has an expectation μ_n and a variance σ_n that are linear in n; in addition, it satisfies a central limit law. In the general context of this paper, the Laplace transform of the distribution of ω also exists for real y in a real interval strictly containing 1, and this fact implies *exponential tails* for large deviations from the mean, see [5, 19, 29]. Thus, *there exist positive constants A, B (depending only on S) such that*

$$\Pr\left\{ |\omega_S[t] - \mu_n| \geq x\sqrt{n} \ \middle|\ |t| = n \right\} \leq Ae^{-Bx}.$$

5.3 Local laws

As seen in Section 5.1, a *central limit law* approximates the distribution function of the number of occurrences by a Gaussian error function. A *local limit law* (of the Gaussian type) means a direct approximation of the the probabilities by the Gaussian *density* (*i.e.*, the normalized form of the function $e^{-x^2/2}$).

> **Local Limit Law** (*LLL*): a sequence of random variables Ω_n with mean μ_n and standard deviation σ_n satisfies a local limit law if, for x in any fixed compact subset of $]0, +\infty[$,
>
> $$\sup_x \left| \sigma_n \Pr\{\Omega_n = \lfloor \mu_n + x\sigma_n \rfloor\} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| \to 0.$$

Clearly, an LLL provides a direct estimation for the coefficients $f_{n,k} = \Pr\{\Omega_n = k\} = [z^n y^k]F(z,y)$. Local limit law usually accompany central limit laws, but their proofs require strong regularity conditions on the distribution; see [2] for a clear discussion.

Theorem 6 (Local law for leaves) *A local limit law in the sense of* (LLL) *holds for the number of leaves in random BSTs, that is to say, the number of occurrences of the particular pattern '*o*'. The mean* μ_n *and variance* σ_n *are*

$$\mu_n = \frac{n+1}{3}, \qquad \sigma_n = \frac{2(n+1)}{45}.$$

We present here a proof schema based on a lemma of greater generality that also clearly delineates the power and limitations of the analytic method. For an arbitrary pattern u, $\alpha_u(y)$ may be defined as the root of smallest modulus of the transcendental equation (7). By Lemma 3 and subsequent remarks, this function is uniquely defined in a disc that properly contains the unit disc; for instance, we may take $|y| < \frac{3}{2}$, for $|u| > 4$, by Lemma 3. As already noted, $\alpha_u(y)$ that is single-valued and locally analytic in such a disc is also globally an analytic function of y.

Lemma 4 *Given a pattern* u*, assume that the uniqueness condition holds:*

Condition (C)*:* $|\alpha_u(y)| \neq 1$ *for all* y *with* $|y| = 1$, $y \neq 1$.

Then the random variable Ω_n *satisfies a local limit law in the sense of* (LLL).

Proof (sketch). The condition (C) means that for $|y| = 1$, the function $F(z, y)$ has, when y varies, a pole of modulus 1 in the sole case when $y = 1$. Also, for $|y| = 1$ and $y \neq 1$, we must have $|\alpha_u(y)|$ always on the same side of 1, and, in fact, we have $|\alpha_u(y)| > 1$ for all $y \neq 1$ by virtue of the property $[z^n]F(z, y) = \mathcal{O}(1)$ when $|y| = 1$.

By singularity analysis of the meromorphic function $F(z, y)$, we have, by an argument used repeatedly before,

$$f_n(y) \equiv [z^n]F(z, y) = \alpha_u(y)^{-n-1}\left(1 + \mathcal{O}(K^{-n})\right),$$

for some $K > 1$. The individual probabilities $f_{n,k}$ can then be recovered by Cauchy's coefficient formula,

$$f_{n,k} \equiv [y^k]f_n(y) = \frac{1}{2i\pi}\int_{|y|=1} \alpha_u(y)^{-n-1}\,\frac{dy}{y^{k+1}} + \varepsilon_{n,k},$$

where $\varepsilon_{n,k}$ is exponentially small.

As is classically known, coefficients of large indices in large powers can be extracted by the saddle point method and, granted unicity of the saddle point on the contour (here $|y| = 1$), the consequence is a local limit law (LLL). We refer to [17, 20, 22, 23, 29] for this fact that is also an offspring of analytic approaches to local and central limit theorems originally stemming from the work of Daniels [8]. Here, there is a saddle point at $y = 1$ and the argument establishes the local limit law, assuming the uniqueness condition (C). □

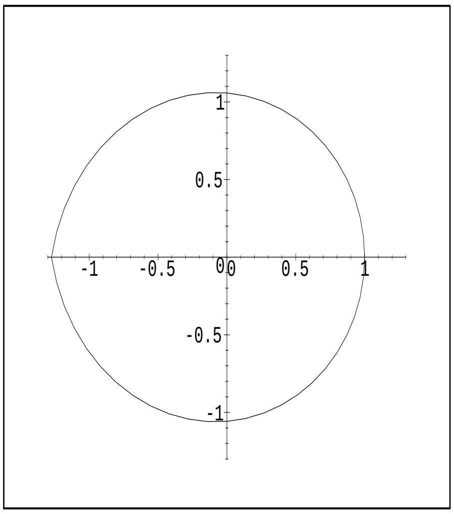

Figure 1: A polar plot of $|\alpha_u(e^{i\theta})|$ as a function of θ in the case of leaves (u='o').

Note also that *a speed of convergence of* $\mathcal{O}(1/\sqrt{n})$ *in the local law derives from the saddle point method.*

Proof (Theorem 6). The preceding discussion specializes easily to the case of leaves, where explicit expressions for $F(z,y)$ are available. The Riccati ODE,

$$\frac{\partial}{\partial z}F(z,y) = F^2(z,y) + (y-1), \qquad F(0,y) = 1,$$

reduces to a second order linear ODE that has constant coefficients. From there, one finds

$$F(z,y) = \frac{1-\delta\tanh(\delta z)}{1-\delta^{-1}\tanh(\delta z)}, \qquad \delta = (1-y)^{1/2}. \tag{19}$$

The function $\alpha_u(y)$ is then explicit,

$$\alpha_u(y) = \frac{1}{\delta}\operatorname{atanh}(\delta) = \frac{1}{2\delta}\log\frac{1+\delta}{1-\delta},$$

and the uniqueness condition of Lemma 4 is easily checked to hold. □

As an illustration, we have

$$\alpha_u(1) = 1,\ |\alpha_u(e^{\pm i})| = 1.02267,\ |\alpha_u(e^{\pm 2i})| = 1.09659,\ |\alpha_u(e^{\pm 3i})| = 1.24410,$$

and Figure 1 offers a polar plot of $\xi(\theta) = |\alpha_u(e^{i\theta})|$, with $\xi(\theta)$ plotted on the ray of angle θ. More generally, we believe the uniqueness condition (C) to hold true for *any* fixed pattern u, not just leaves. In any particular instance, the condition of Lemma 4 could at least be tested by numerical analysis and carefully controlled error bounds.

Permutations. By the bijective correspondence between increasing trees and permutations, $f_{n,k}$ also equals the probability for a random permutation σ of size n to have k *peaks*, that is to say configurations such that $\sigma_{i-1} < \sigma_i > \sigma_{i+1}$. Thus, Theorems 4, 6 provide asymptotic laws for permutations counted according to size and number of peaks; see also [9, Ch. 10]. The situation of peaks then appears to be analogous to that of Eulerian numbers [7, 30, 40] that count permutations according to size and number of ascents (configurations such that $\sigma_i < \sigma_{i+1}$), where both a central law [9, 39] and a local law [2, 31] are known.

6 Factored representations of trees

We consider finally a global parameter $\kappa[t]$ of trees that represents the number of structurally different subtrees (*i.e.*, number of different subtree shapes) that occur in t. This parameter is of intrinsic interest as an indicator of the structural "richness" of t. It also measures the optimal storage complexity of tree t when all common subtrees are factored and represented only once. Then, $\kappa[t]$ measures the number of nodes of the maximally factored DAG (directed acyclic graph) corresponding to t, a quantity that intervenes in parsing and data compression applications [18]. We call $\kappa[t]$ the *DAG size* of tree t.

By its definition,

$$\kappa[t] = \sum_{u \in \mathcal{T}} [\![u \in t]\!] = \sum_{u \in \mathcal{T}} [\![\omega_u[t] \geq 1]\!],$$

where the sum is over the set $\mathcal{T}$ of all tree shapes, $[\![\cdot]\!]$ is the indicator function, and $u \in t$ is true if u occurs (at least once) as a subtree of t. We denote by K_n the average value of $\kappa[t]$ under the BST model of index n, so that

$$K_n = \sum_{|t|=n} \lambda(t)\, \kappa[t],$$

where $\lambda(t)$ is the probability of the tree shape t under the BST model.

Under the combinatorial model where all trees are taken equally likely, the corresponding expectation $\widetilde{K}_n$ grows like $n/\sqrt{\log n}$, see [18]. In the BST model where trees tend to be more balanced, we expect *a priori* fewer different subtrees, that is to say $K_n \ll \widetilde{K}_n$. The following simple bound justifies this observation.

Theorem 7 *The average value of the DAG size of a random BST of size n satisfies the upper bound,*

$$K_n \leq \frac{4(\log 2)\, n}{\log n} + \mathcal{O}\left(\frac{n\, \log\log n}{(\log n)^2}\right).$$

Proof. Fix a cutpoint paramemeter b (to be adjusted later). An upper bound $\upsilon_b[t]$ on $\kappa[t]$ is

$$\upsilon_b[t] = B_0 + B_1 + \cdots + B_b + \iota_b[t],$$

where $\iota_b[t]$ is the number of nodes in t whose subtrees have size $> b$, and

$$B_k = \frac{1}{k+1}\binom{2k}{k}$$

is the Catalan number that counts the number of binary trees with k internal nodes. Combinatorially, $\upsilon_b[t]$ is the size of an approximate DAG representation where all trees of size less than b are represented once, irrespective of their possible nonoccurrence in t, and nodes commanding subtrees of size $\geq b$ are each represented irrespective of the fact that they may be associated to repeated subtrees. In other words, $\upsilon_b[t]$ is the size of a partly redundant and partially factored DAG representation.

Let $U_{b,n}$ and $I_{b,n}$ be the expectations of υ_b and ι_b under the BST model of size n. The analysis of ι_b is exactly that of paging in Section 5.2, and we have by Theorem 5, for $n > b$,

$$I_{b,n} = \frac{2n}{b+2} - \frac{b}{b+2}. \tag{20}$$

As the Catalan numbers grow like $4^k k^{-3/2}$, we also have, for *any* b:

$$B_0 + B_1 + \cdots B_b = \mathcal{O}(\frac{4^b}{b^{3/2}}). \tag{21}$$

Thus, we have

$$K_n < \frac{2n}{b} + \mathcal{O}(\frac{4^b}{b^{3/2}}).$$

Equipped with this family of upper bounds, we can now optimize the choice of the cutpoint b in (20,21). Adopting

$$b = \lfloor \frac{\log n - \log\log n}{\log 4} \rfloor,$$

so that

$$\frac{4^b}{b^{3/2}} = \mathcal{O}\left(\frac{n}{(\log n)^{5/2}}\right), \qquad I_{b,n} = 4(\log 2)\,\frac{n}{\log n} + \mathcal{O}\left(\frac{n\log\log n}{(\log n)^2}\right)$$

Values of y	Combinatorial property	
$y = 0$	Excluded pattern	Thm. 2
$y \approx 0$	Rare occurrences	Thm. 3
$y = 1$	Moments	Thm. 1
$y \approx 1$	Central limit law	Thm. 4, 5
$\lvert y \rvert = 1$	Local limit law	Thm. 6

Figure 2: The correspondence between regions of the auxiliary variable y and combinatorial properties of pattern occurrences.

yields the stated inequality. □

Based on simulations and heuristic analysis, we have reasons to believe that the upper bound on K_n is of the right order. We know by Thm. 1 that any fixed pattern occurs almost surely in a large random tree. The argument of the upper bound suggests that, *almost surely* in a large BST of size n, *all* the patterns of size a bit less than $\log_4 n$ occur *at least once* while *all* the patterns of size a bit more than $\log_4 n$ occur *at most once.* (Analogous laws are known for random strings [15].) Second moment methods based on Theorem 1 seem to crude to establish such properties. Perhaps the thresholds could be precisely quantified along the lines of Theorems 2, 3, 4, by allowing for uniform error terms when $|u|$ grows with n; see [15] for strings. This would lead to a precise asymptotic estimation of K_n.

7 Conclusions

There are two aspects of possible interest in the present work, one relative to the "physics" of random trees and permutations, the other concerning methodologies for multivariate asymptotics.

Regarding methodology, the analysis of pattern occurrences as presented here is attached to the general domain of bivariate asymptotics. Here, the combinatorial problems translate into nonlinear differential equations that, being of the Riccati type, lead to linear second-order ODE's. Singularities in the main variable z drive the asymptotic behaviour of the bivariate generating function, with the auxiliary variable y entering as a parameter. It is interesting to note, and typical, that different regions of values of the auxiliary variable provide information on excluded patterns, rare ocurrences, as well as central or local limit laws, as summarized in Fig. 2. Globally, the process belongs to the class of *singularity perturbation* methods; see [4, 12, 16] for related situations.

Regarding the physics of the problem, random structures of a large size tend to have any small subconfiguration whose occurrences obey a Gaussian law, whether local or global. This is a well-established fact for random strings (see, *e.g.*, [15]) with implications in computational biology [45, Chap. 12], for random combinatorial trees as implied by the results of [18, 41], as well as for random graphs of various sorts [6]. Our work adds random binary search trees to the collection, and recent analyses by Devroye [11] suggest that such universal behaviour should persist for many other types of trees. Consideration of multiway search trees that exhibit some sort of Gaussian behaviour, as shown by Mahmoud and Pittel [34], also supports this expectation.

Acknowledgements. This work was supported in part by the Long Term Research Project ALCOM-IT (# 20244) of the European Union.

References

[1] ARAGON, C. R., AND SEIDEL, R. G. Randomized search trees. In *30th Annual Symposium on Foundations of Computer Science* (1989), IEEE Computer Society, pp. 540–545.

[2] BENDER, E. A. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory 15* (1973), 91–111.

[3] BENDER, E. A., AND RICHMOND, L. B. Central and local limit theorems applied to asymptotic enumeration II: Multivariate generating functions. *Journal of Combinatorial Theory,* Series A *34* (1983), 255–265.

[4] BERGERON, F., FLAJOLET, P., AND SALVY, B. Varieties of increasing trees. In *CAAP'92* (1992), J.-C. Raoult, Ed., vol. 581 of *Lecture Notes in Computer Science*, pp. 24–48. Proceedings of the 17th Colloquium on Trees in Algebra and Programming, Rennes, France, February 1992.

[5] BILLINGSLEY, P. *Probability and Measure*, 2nd ed. John Wiley & Sons, 1986.

[6] BOLLOBÁS, B. *Random Graphs.* Academic Press, 1985.

[7] COMTET, L. *Advanced Combinatorics.* Reidel, Dordrecht, 1974.

[8] DANIELS, H. E. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics 25* (1954), 631–650.

[9] DAVID, F. N., AND BARTON, D. E. *Combinatorial Chance.* Charles Griffin, London, 1962.

[10] DEVROYE, L. Limit laws for local counters in random binary trees. *Random Structures and Algorithms 2*, 3 (1991), 302–315.

[11] DEVROYE, L. Universal limit laws for depths in random trees. Manuscript, 1995. 35 pages.

[12] DRMOTA, M. Systems of functional equations. Manuscript, Aug. 1996. To appear in *Random Structures and Algorithms*, 1997.

[13] FELLER, W. *An Introduction to Probability Theory and Its Applications*, vol. 2. John Wiley, 1971.

[14] FILL, J. A. On the distribution of binary search trees under the random permutation model. *Random Structures and Algorithms 8*, 1 (1996), 1–25.

[15] Flajolet, P., Kirschenhofer, P., and Tichy, R. F. Deviations from uniformity in random strings. *Probability Theory and Related Fields 80* (1988), 139–150.

[16] Flajolet, P., and Lafforgue, T. Search costs in quadtrees and singularity perturbation asymptotics. *Discrete and Computational Geometry 12*, 4 (1994), 151–175.

[17] Flajolet, P., and Sedgewick, R. The average case analysis of algorithms: Saddle point asymptotics. Research Report 2376, Institut National de Recherche en Informatique et en Automatique, 1994. 55 pages.

[18] Flajolet, P., Sipala, P., and Steyaert, J.-M. Analytic variations on the common subexpression problem. In *Automata, Languages, and Programming* (1990), M. S. Paterson, Ed., vol. 443 of *Lecture Notes in Computer Science*, pp. 220–234. Proceedings of the 17th ICALP Conference, Warwick, July 1990.

[19] Flajolet, P., and Soria, M. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics 114* (1993), 159–180.

[20] Gamkrelidze, R. V., Ed. *Analysis I, Integral Representations and Asymptotic Methods*, vol. 13 of *Encyclopedia of Mathematical Sciences.* Springer Verlag, 1989.

[21] Gao, Z., and Richmond, L. B. Central and local limit theorems applied to asymptotic enumerations IV: Multivariate generating functions. *Journal of Computational and Applied Mathematics 41* (1992), 177–186.

[22] Gardy, D. Méthode de col et lois limites en analyse combinatoire. *Theoretical Computer Science 92*, 2 (1992), 261–280.

[23] Greene, D. H., and Knuth, D. E. *Mathematics for the analysis of algorithms.* Birkhauser, Boston, 1981.

[24] Guibas, L. J., and Odlyzko, A. M. Periods in strings. *Journal of Combinatorial Theory,* Series A *30* (1981), 19–42.

[25] Henrici, P. *Applied and Computational Complex Analysis.* John Wiley, New York, 1977. 3 volumes.

[26] Hille, E. *Analytic Function Theory*, vol. 1. Blaisdell Publishing Company, Waltham, 1959.

[27] Hille, E. *Analytic Function Theory*, vol. 2. Blaisdell Publishing Company, Waltham, 1962.

[28] Hoshi, M., and Flajolet, P. Page usage in a quadtree index. *BIT 32* (1992), 384–402.

[29] Hwang, H.-K. *Théorèmes limites pour les structures combinatoires et les fonctions arithmetiques.* PhD thesis, École Polytechnique, Dec. 1994.

[30] Knuth, D. E. *The Art of Computer Programming*, vol. 3: Sorting and Searching. Addison-Wesley, 1973.

[31] Lesieur, L., and Nicolas, J.-L. On the Eulerian numbers $M_n = \max A_{n,k}$. *European Journal of Combinatorics 13* (1992), 379–399.

[32] Lukacs, E. *Characteristic Functions.* Griffin, London, 1970.

[33] Mahmoud, H. *Evolution of Random Search Trees.* John Wiley, New York, 1992.

[34] Mahmoud, H. M., and Pittel, B. Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms 10* (1989), 52–75.

[35] Martínez, C. *Statistics under the BST Model.* PhD thesis, Universitat Politècnica de Catalunya, Apr. 1992.

[36] Motwani, R., and Raghavan, P. *Randomized Algorithms.* Cambridge University Press, 1995.

[37] Nicodème, P. Compact balanced tries. In *Algorithms, Software, Architecture, Information Processing 92* (1992), J. van Leeuwen, Ed., Elsevier Science Publishers, pp. 19–27. Proceedings of IFIP Congress, Ljubljana, 1971.

[38] Roura, S., and Martínez, C. Randomization of search trees by subtree size. In *Algorithms—ESA'96* (1996), J. Diaz and M. Serna, Eds., no. 1136 in Lecture Notes in Computer Science, pp. 91–106. Proceedings of the Fourth European Symposium on Algorithms, Barcelona, September 1996.

[39] Sachkov, V. *Verojatnostnye Metody v Kombinatornom Analize.* Nauka, Moscow, 1978.

[40] Sedgewick, R., and Flajolet, P. *An Introduction to the Analysis of Algorithms.* Addison-Wesley Publishing Company, 1996. 512 pages. (ISBN 0-201-4009-X).

[41] Steyaert, J.-M., and Flajolet, P. Patterns and pattern-matching in trees: an analysis. *Information and Control 58*, 1–3 (July 1983), 19–58.

[42] Vitter, J. S., and Flajolet, P. Analysis of algorithms and data structures. In *Handbook of Theoretical Computer Science*, J. van Leeuwen, Ed., vol. A: Algorithms and Complexity. North Holland, 1990, ch. 9, pp. 431–524.

[43] Vuillemin, J. A unifying look at data structures. *Communications of the ACM 23*, 4 (Apr. 1980), 229–239.

[44] Wasow, W. *Asymptotic Expansions for Ordinary Differential Equations.* Dover, 1987. A reprint of the John Wiley edition, 1965.

[45] Waterman, M. S. *Introduction to Computational Biology.* Chapman & Hall, 1995.

[46] Whittaker, E. T., and Watson, G. N. *A Course of Modern Analysis*, fourth ed. Cambridge University Press, 1927. Reprinted 1973.