Statistics & Decisions 23, 131–146 (2005)

Recursive random variables with subgaussian distributions

Ralph Neininger

Received: Mai 5, 2005; Accepted: September 23, 2005

Summary: We consider sequences of random variables with distributions that satisfy recurrences as they appear for quantities on random trees, random combinatorial structures and recursive algorithms. We study the tails of such random variables in cases where after normalization convergence to the normal distribution holds. General theorems implying subgaussian distributions are derived. Also cases are discussed with non-Gaussian tails. Applications to the probabilistic analysis of algorithms and data structures are given.

1 Introduction

A large number of quantities $(X_n)_{n\geq 0}$ of recursive combinatorial structures, random trees and recursive algorithms satisfy recurrences of the form

$$X_n \stackrel{d}{=} \sum_{r=1}^{K} X^{(r)}_{I_r^{(n)}} + b_n, \quad n \geq n_0, \tag{1.1}$$

with $K, n_0 \geq 1$, $(X_n^{(r)})_{n\geq 0}$ identically distributed as $(X_n)_{n\geq 0}$ for $r = 1, \dots, K$, a random vector $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ of integers in $\{0, \dots, n-1\}$ and a random b_n such that $(X_n^{(1)})_{n\geq 0}, \dots, (X_n^{(K)})_{n\geq 0}, (I^{(n)}, b_n)$ are independent. The symbol $\stackrel{d}{=}$ denotes equality in distribution. In applications, the $I_r^{(n)}$ are random subgroup sizes, b_n is a toll function specifying the particular quantity of a combinatorial structure and $(X_n^{(r)})_{n\geq 0}$ are copies of the quantity $(X_n)_{n\geq 0}$, that correspond to the contribution of subgroup r. Typical parameters X_n range from the depths, sizes and path lengths of trees, the number of various sub-structures or components of combinatorial structures, the number of comparisons, space requirements and other cost measures of algorithms to param-

---

AMS 1991 subject classification: Primary: 60F10, 68Q25; Secondary: 68P05
Key words and phrases: tail bound, large deviation principle, recursion, analysis of algorithms, subgaussian distribution
Research supported by an Emmy Noether fellowship of the DFG.

eters of communication models, and many more. Numerous examples that are occurring in these areas will be discussed below; see also the books of Mahmoud (1992), Sedgewick and Flajolet (1996), Szpankowski (2001), and Arratia, Barbour and Tavaré (2003).

Stochastic analysis of such quantities has been performed in many special cases, mainly with respect to the computation of averages and higher moments of X_n, limit laws and rates of convergences. Techniques in use include moment generating functions, saddle point methods, the method of moments, martingales, and various direct approaches to asymptotic normality such as representations as sums of independent or weakly dependent random variables, Stein's method and Berry–Esseen methodology.

During the last 15 years an efficient and quite universal probabilistic tool for the analysis of asymptotic distributions for recurrences as in (1.1), the contraction method, has been developed. It has been introduced for the analysis of the Quicksort algorithm in Rösler (1991) and further developed independently in Rösler (1992) and Rachev and Rüschendorf (1995), see also the survey article of Rösler and Rüschendorf (2001). It has been applied and extended since then successfully to a large number of problems.

Recently, fairly general unifying limit theorems for this type of recurrences have been obtained by the contraction method in Neininger and Rüschendorf (2004a, 2004b). Typically, the limit distribution of the normalized recurrence is uniquely characterized by a fixed point equation; we give a general outline below.

In this paper tail bounds for the quantities X_n are studied in cases, where the rescaled quantities tend to a normal limit. Revisiting all presently known applications of the contraction method that lead to a normal limit law in the area of analysis of algorithms, one finds three structurally different cases how a normal limit law has appeared in the context of the contraction method. For two of these cases we derive Gaussian tail bounds under general conditions on the expansion of the first moment of X_n. In the third case we discuss an example that leads to a large deviation principle with a rate function that increases slower than quadratic at infinity.

For particular examples of recurrence (1.1), where $I^{(n)}$ is explicitly given, sharp analytic tolls based on the analysis of generating functions usually give precise bounds. The intention of the present paper is to derive theorems that do not make use of the particular splitting vector $I^{(n)}$ and are valid for a whole class of problems that are related by a similar splitting vector. Since our theorems below need assumptions on the expansion of moments which are often derived via generating functions, analytic and probabilistic tools may be regarded complementary here.

General bounds on the upper tail for recurrences as (1.1) have been derived in Karp (1994) which also apply if the recurrence is less explicitly given than in our setting.

The paper is organized as follows. In section 2 we outline the approach of the contraction method and discuss three typical situations that lead to normal limit laws. Section 3 reviews some technical preliminaries on basic concentration inequalities. The sections 4–6 contain tail bounds for X_n for the three different cases together with applications to special examples.

We use L, C, D_1, and D_2 as generic symbols standing for constants that may change from one occurrence to another.

2 The contraction method

In the framework of the contraction method the quantities X_n in (1.1) are first rescaled by

$$Y_n := \frac{X_n - m(n)}{s(n)}, \quad n \geq 0, \tag{2.1}$$

where $m(n)$ and $s(n)$ are appropriately chosen, e.g., of the order of mean and standard deviation of X_n. Then, recursion (1.1) for X_n implies a modified recurrence for the scaled quantities Y_n,

$$Y_n \stackrel{d}{=} \sum_{r=1}^{K} \frac{s(I_r^{(n)})}{s(n)} Y_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq n_0, \tag{2.2}$$

with

$$b^{(n)} = \frac{1}{s(n)} \left(b_n - m(n) + \sum_{r=1}^{K} m(I_r^{(n)}) \right) \tag{2.3}$$

and conditions on independence and distributional copies as in (1.1). Then, the contraction method aims to provide theorems of the following type: Assuming that the coefficients in (2.2) are appropriately convergent,

$$\frac{s(I_r^{(n)})}{s(n)} \to A_r^*, \quad b^{(n)} \to b^*, \quad (n \to \infty) \tag{2.4}$$

with random A_r^*, b^*, then under appropriate conditions the quantities (Y_n) itself converge in distribution to a limit Y. The limit distribution $\mathcal{L}(Y)$ is obtained as a solution of the fixed point equation that is obtained from (2.2) by letting formally $n \to \infty$:

$$Y \stackrel{d}{=} \sum_{r=1}^{K} A_r^* Y^{(r)} + b^*. \tag{2.5}$$

Here, $(A_1^*, \ldots, A_K^*, b^*), Y^{(1)}, \ldots, Y^{(K)}$ are independent and $Y^{(r)} \stackrel{d}{=} Y$ for $r = 1, \ldots, K$. Usually, under constraints on the finiteness of moments of $\mathcal{L}(Y)$ the fixed point equation (2.5) has a unique solution that is the limit distribution in the corresponding limit law.

This approach has been universally developed in Neininger and Rüschendorf (2004a), where detailed conditions for convergence of Y_n are discussed.

A fixed point of (2.5) is in general not easily accessible. However, for some classes of problems the normal distribution appears as limit distribution. There are mainly three structurally different situations, in which the normal distribution appears:

The case $\sum (A_r^*)^2 = 1$, $b^* = 0$: It is well known that equation (2.5), with $\sum_{r=1}^{K} (A_r^*)^2 = 1$ and $b^* = 0$ almost surely, has exactly the centered normal distributions as solutions (excluding the degenerate case where the A_r^* only take the values 0 and 1). This is the most frequent occurrence of the normal distribution in applications in the analysis of algorithms and combinatorial structures. Various examples can be found in section 5.3 of Neininger and Rüschendorf (2004a). Subgaussian distributions are derived for some cases in Theorem 4.1.

The case $Y \stackrel{d}{=} Y$: A degenerate fixed point equation is one with $\sum_{r=1}^{K} A_r^* = 1$, where the A_r^* only take the values 0 and 1 almost surely, and $b^* = 0$. Any distribution is a solution to these fixed point equations, hence we call this case $Y \stackrel{d}{=} Y$. It appears in particular for quantities X_n with variances that are slowly varying at infinity. Limit laws for certain classes of problems where precise expansions of mean and variance are available are studied together with applications in Neininger and Rüschendorf (2004b). We derive subgaussian distributions in some cases in Theorem 5.1.

The case of deterministic A_r^* and $b^* \stackrel{d}{=} \mathcal{N}$: Equation (2.5) with deterministic $(A_1^*, \ldots, A_K^*)$ with $\sum_{r=1}^{K} |A_r^*| < 1$ and b^* being normally $\mathcal{N}(\nu, \tau^2)$ distributed has the normal distribution $\mathcal{N}(\mu, \sigma^2)$ as a solution, where mean μ and standard deviation σ are determined in terms of ν, τ, and the A_r^*, cf. equations (6.4). The solution $\mathcal{N}(\mu, \sigma^2)$ is unique under the constraint of a finite absolute first moment. The occurrence of the normal limit distribution via this fixed point equation has not yet been systematically studied. In section 6 a general normal limit law is given in Theorem 6.1, applications are mentioned, and for a particular case, non-Gaussian tails are explicitly quantified.

3 Technical preliminaries

In this section we recall basic notions, Hoeffding's Lemma and give a version of Chernoff's bounding argument; for general reference see Petrov (1975).

Definition 3.1 *A random variable X is said to have subgaussian distribution if there exists an $L > 0$ such that for all $\lambda > 0$,*

$$\mathbb{E} \exp(\lambda X) \leq \exp(L\lambda^2).$$

For centered, bounded random variables we have Hoeffding's Lemma (1963):

Lemma 3.2 (Hoeffding's Lemma) *Let X be a random variable with $a \leq X \leq b$ and $\mathbb{E}\, X = 0$. Then, for all $\lambda \in \mathbb{R}$, we have*

$$\mathbb{E} \exp(\lambda X) \leq \exp\left(\frac{(b-a)^2\lambda^2}{8}\right).$$

We will also need a bound on the moment generating function of centered random variables that are only bounded from above:

Lemma 3.3 *Let X be a random variable with $X \leq b$, $\mathbb{E}\, X = 0$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Then, there exists an $L \geq 0$ such that for all $\lambda > 0$,*

$$\mathbb{E} \exp(\lambda X) \leq \exp(L\lambda^2).$$

We may choose

$$L = \sup_{x>0} \left\{ \frac{b}{x} \wedge \frac{(e^{xb} - 1 - xb)\sigma^2}{(xb)^2} \right\} < \infty. \tag{3.1}$$

Proof: The proof resembles ideas from Bennett (1962). We have

$$\exp(\lambda X) = 1 + \lambda X + \lambda^2 X^2 \, \frac{\exp(\lambda X) - 1 - \lambda X}{(\lambda X)^2}.$$

It is easily checked that the function g defined by $g(s) := (e^s - 1 - s)/s^2$ for $s \neq 0$ and $g(0) := 1/2$ is monotonically increasing. Thus, for $\lambda > 0$, we obtain from $\lambda X \leq \lambda b$ that

$$\exp(\lambda X) \leq 1 + \lambda X + g(\lambda b)\lambda^2 X^2.$$

Taking expectations yields

$$\begin{aligned} \mathbb{E} \exp(\lambda X) &\leq 1 + g(\lambda b)\sigma^2\lambda^2 \\ &\leq \exp(g(\lambda b)\sigma^2\lambda^2). \end{aligned} \tag{3.2}$$

On the other hand, for $\lambda > 0$, we obtain from $X \leq b$ that

$$\mathbb{E} \exp(\lambda X) \leq \exp(b\lambda). \tag{3.3}$$

Combining (3.2) and (3.3) we obtain for all $\lambda > 0$,

$$\begin{aligned} \mathbb{E} \exp(\lambda X) &\leq \exp\left(\left(\frac{b}{\lambda} \wedge g(\lambda b)\sigma^2 \right) \lambda^2 \right) \\ &\leq \exp(L\lambda^2), \end{aligned}$$

with L as given in (3.1). That L is finite follows from the fact that b/x is decreasing and $g(xb)\sigma^2$ is increasing for $x \in (0, \infty)$. □

For sequences (X_n) of random variables, we will subsequently obtain subgaussian distributions for normalizations $(X_n - \mathbb{E} X_n)/s(n)$, where the constant L in Definition 3.1 can be chosen uniformly in n. In such cases the following tail bound follows via Chernoff's bounding technique:

Lemma 3.4 *Let $(X_n)_{n\geq 0}$ be a sequence of integrable random variables and $s(n) > 0$, $L > 0$ such that for all $\lambda > 0$, $n \geq 0$,*

$$\exp\left(\lambda \frac{X_n - \mathbb{E} X_n}{s(n)} \right) \leq \exp(L\lambda^2). \tag{3.4}$$

Then we have for all $t > 0$ and $n \geq 0$,

$$\mathbb{P}(X_n - \mathbb{E} X_n \geq t|\mathbb{E} X_n|) \leq \exp\left(-\frac{t^2}{4L} \left(\frac{\mathbb{E} X_n}{s(n)} \right)^2 \right). \tag{3.5}$$

If (3.4) holds for all $\lambda \in \mathbb{R}$ *and* $n \geq 0$ *then we have for all* $t > 0$ *and* $n \geq 0$,

$$\mathbb{P}(|X_n - \mathbb{E}\,X_n| \geq t|\,\mathbb{E}\,X_n|) \leq 2\exp\left(-\frac{t^2}{4L}\left(\frac{\mathbb{E}\,X_n}{s(n)}\right)^2\right). \tag{3.6}$$

Proof: Chernoff's bounding technique yields for $t > 0$

$$\begin{aligned}\mathbb{P}(X_n - \mathbb{E}\,X_n \geq t|\,\mathbb{E}\,X_n|) &= \mathbb{P}\left(\exp\left(\lambda\frac{X_n - \mathbb{E}\,X_n}{s(n)}\right) \geq \exp\left(\lambda t\frac{|\,\mathbb{E}\,X_n|}{s(n)}\right)\right)\\ &\leq \exp(L\lambda^2 - \lambda t|\,\mathbb{E}\,X_n|/s(n))\end{aligned}$$

for all $\lambda > 0$. This bound is optimized by choosing $\lambda = t|\,\mathbb{E}\,X_n|/(2Ls(n))$. For (3.6) we apply the same argument as well to $-X_n$. □

4 The case $\sum(A_r^*)^2 = 1$, $b^* = 0$

We consider a sequence $(X_n)_{n\geq 0}$ of random variables satisfying recurrence (1.1). Then subgaussian distributions appear in the following situation that is frequent in applications.

Theorem 4.1 *Assume that* $(X_n)_{n\geq 0}$ *satisfies (1.1) and that we have*

$$\|X_0\|_\infty, \ldots, \|X_{n_0-1}\|_\infty < \infty, \quad \sup_{n\geq n_0} \|b_n\|_\infty < \infty, \tag{4.1}$$

$$1 \leq n - \sum_{r=1}^{K} I_r^{(n)} \leq C \text{ almost surely,} \tag{4.2}$$

$$\mathbb{E}\,X_n = \mu n + O(1),$$

with $\mu \neq 0$ *and a constant* $C \geq 1$.

Then, there exists an $L > 0$ *such that for all* $\lambda \in \mathbb{R}$, $n \geq 1$,

$$\mathbb{E}\,\exp\left(\lambda\frac{X_n - \mathbb{E}\,X_n}{\sqrt{n}}\right) \leq \exp(L\lambda^2).$$

In particular, we have (3.6) with $s(n) = \sqrt{n \vee 1}$.

Proof: We denote

$$D_1 := \sup_{n\geq 0}|\,\mathbb{E}\,X_n - \mu n|, \quad D_2 := \sup_{n\geq n_0}\|b_n\|_\infty \tag{4.3}$$

and consider the scaled quantities

$$Y_n := \frac{X_n - \mathbb{E}\,X_n}{\sqrt{n}}, \quad n \geq 1,$$

and $Y_0 := (1/\varepsilon)(X_0 - \mathbb{E}\, X_0)$ for a $0 < \varepsilon \le 1$. Then (1.1) implies

$$Y_n \stackrel{d}{=} \sum_{r=1}^{K} \sqrt{\frac{I_r^{(n)} \vee \varepsilon}{n}}\, Y^{(r)}_{I_r^{(n)}} + b^{(n)}, \quad n \ge n_0 \tag{4.4}$$

with

$$b^{(n)} := \frac{1}{\sqrt{n}} \left(b_n - \mu \left(n - \sum_{r=1}^{K} I_r^{(n)} \right) + R_n \right) \tag{4.5}$$

with a random R_n satisfying $|R_n| \le (K+1)D_1$.

Since $Y_0, \dots, Y_{n_0-1}$ are centered and bounded random variables, Hoeffding's Lemma implies that there exists a $Q > 0$ such that, for all $\lambda \in \mathbb{R}$ and all $j = 0, \dots, n_0 - 1$ the bound $\mathbb{E}\, \exp(\lambda Y_j) \le \exp(Q\lambda^2)$ holds. We show by induction that there exists $L \ge Q$ such that for all $\lambda \in \mathbb{R}$ and all $j \ge 0$,

$$\mathbb{E}\, \exp(\lambda Y_j) \le \exp(L\lambda^2). \tag{4.6}$$

The assertion is true for $j = 0, \dots, n_0 - 1$ since $L \ge Q$. For the induction step we assume that (4.6) holds for all $j = 0, \dots, n-1$. Denoting by Υ_n the distribution of the vector $(I^{(n)}, b^{(n)})$ we obtain with (4.4), conditioning on $(I^{(n)}, b^{(n)})$, the induction hypothesis, and the notation $\mathbf{j} = (j_1, \dots, j_K)$ that

$$\begin{aligned}
\mathbb{E}\, \exp(\lambda Y_n) &= \mathbb{E}\, \exp\left(\lambda \sum_{r=1}^{K} \sqrt{\frac{I_r^{(n)} \vee \varepsilon}{n}}\, Y^{(r)}_{I_r^{(n)}} + \lambda b^{(n)} \right) \\
&= \int \mathbb{E}\, \exp\left(\lambda \sum_{r=1}^{K} \sqrt{\frac{j_r \vee \varepsilon}{n}}\, Y^{(r)}_{j_r} + \lambda\beta \right) d\Upsilon_n(\mathbf{j}, \beta) \\
&\le \int \exp\left(L\lambda^2 \sum_{r=1}^{K} \frac{j_r \vee \varepsilon}{n} + \lambda\beta \right) d\Upsilon_n(\mathbf{j}, \beta) \\
&= \exp(L\lambda^2)\, \mathbb{E}\, \exp\left(L\lambda^2 \left(\sum_{r=1}^{K} \frac{I_r^{(n)} \vee \varepsilon}{n} - 1 \right) + \lambda b^{(n)} \right).
\end{aligned}$$

Hence, for the induction step it is sufficient to show that

$$\sup_{n \ge n_0} \mathbb{E}\, \exp\left(L\lambda^2 \left(\sum_{r=1}^{K} \frac{I_r^{(n)} \vee \varepsilon}{n} - 1 \right) + \lambda b^{(n)} \right) \le 1. \tag{4.7}$$

By (4.2) we obtain

$$\sum_{r=1}^{K} \frac{I_r^{(n)} \vee \varepsilon}{n} - 1 \le \frac{-1 + K\varepsilon}{n} \le -\frac{1}{2n} \tag{4.8}$$

for $0 < \varepsilon \leq 1 \wedge (2/K)$. By (4.1), (4.2), (4.3) and (4.5) we obtain

$$\|b^{(n)}\|_\infty \leq \frac{1}{\sqrt{n}} (\|b_n\|_\infty + \mu C + (K+1)D_1) \leq \frac{M}{\sqrt{n}}$$

with

$$M := D_2 + \mu C + (K+1)D_1.$$

Moreover, $\mathbb{E}\, Y_n = 0$ implies $\mathbb{E}\, b^{(n)} = 0$. Hence, Hoeffding's Lemma implies

$$\mathbb{E}\, \exp(\lambda b^{(n)}) \leq \exp\left(\frac{\lambda^2 (2\|b^{(n)}\|_\infty)^2}{8}\right) \leq \exp\left(\frac{(M\lambda)^2}{2n}\right). \tag{4.9}$$

Combining (4.8) and (4.9) we obtain with ε as above

$$\mathbb{E}\, \exp\left(L\lambda^2 \left(\sum_{r=1}^{K} \frac{I_r^{(n)} \vee \varepsilon}{n} - 1\right) + \lambda b^{(n)}\right) \leq \exp\left(-\frac{L\lambda^2}{2n} + \frac{(M\lambda)^2}{2n}\right) \leq 1$$

if $L \geq M^2$. Hence the induction step is completed by choosing $L := M^2 \vee Q$. □

We give a couple of applications of Theorem 4.1 on the probabilistic analysis of algorithms and data structures:

Number of leaves in random binary search trees: The number of leaves X_n in a random binary search tree with n elements satisfies recurrence (1.1) with $K = 2$, $n_0 = 2$, $X_0 = 0$, $X_1 = 1$, $b_n = 0$ and $I_1^{(n)} \stackrel{d}{=} \mathrm{unif}\{0, \ldots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$. It is well known that for this quantity $\mathbb{E}\, X_n = (n+1)/3 = n/3 + O(1)$ holds, see Mahmoud (1986), Devroye (1991) and Flajolet, Gourdon and Martínez (1997). Hence, all conditions of Theorem 4.1 are satisfied and subgaussian distributions are implied. In particular, (3.6) is implied with $s(n) = \sqrt{n \vee 1}$.

Binary search trees with bounded toll functions: Binary search tree recurrences have been studied for general toll functions b_n in Devroye (2002/03) and Hwang and Neininger (2002). These are quantities X_n that satisfy recurrence (1.1) with $K = 2$, $n_0 = 1$, $X_0 = 0$, and $I_1^{(n)} \stackrel{d}{=} \mathrm{unif}\{0, \ldots, n-1\}$, $I_2^{(n)} = n - 1 - I_1^{(n)}$. We consider the case of uniformly bounded toll functions b_n, i.e., $\sup_{n \geq 1} \|b_n\|_\infty < \infty$ and assume that

$$\mu := \sum_{k=1}^{\infty} \frac{\mathbb{E}\, b_k}{(k+1)(k+2)} \neq 0. \tag{4.10}$$

It is well known that for the binary search tree recurrences we have

$$\mathbb{E}\, X_n = \mathbb{E}\, b_n + 2(n+1) \sum_{k=1}^{n-1} \frac{\mathbb{E}\, b_k}{(k+1)(k+2)},$$

see, e.g., Lemma 1 in Hwang and Neininger (2002). Hence, $\sup_{n\geq 1} \|b_n\|_\infty < \infty$, (4.10), and $\sum_{k=n}^{\infty} 1/k^2 = O(1/n)$ imply $\mathbb{E}\, X_n = \mu n + O(1)$ with μ given in (4.10). Thus, Theorem 4.1 yields subgaussian distributions for all binary search tree recurrences with uniformly bounded toll functions satisfying (4.10). Various examples of such quantities relevant in the analysis of tree traversing algorithms and secondary cost measures of Quicksort are given in section 6 of Hwang and Neininger (2002).

Size of m-ary search trees: The size X_n of random m-ary search trees, $m \geq 3$, satisfies recurrence (1.1) with $K = m$, $n_0 = m$, $X_0 = 0$, $X_1 = \cdots = X_{m-1} = 1$, $b_n = 1$, and $I^{(n)}$ being a certain mixture of multinomial distributions with $\sum_{1\leq r\leq m} I_r^{(n)} = n - m + 1$. It is known that $\mathbb{E}\, X_n = (2(H_m - 1))^{-1} n + O(1)$ for all $3 \leq m \leq 13$, see Mahmoud and Pittel (1989), Lew and Mahmoud (1994), and Chern and Hwang (2001). Here, H_m denotes the mth harmonic number $H_m = \sum_{1\leq k\leq m} 1/k$. Hence, all conditions of Theorem 4.1 are satisfied and we obtain subgaussian distributions for the size of random m-ary search trees for $3 \leq m \leq 13$. For a discussion of phase changes in m-ary search trees see Hwang (2003).

Number of leaves in random quadtrees: The number of leaves X_n in a d-dimensional random (point) quadtree with n elements satisfies recurrence (1.1) with $K = 2^d$, $n_0 = 2$, $X_0 = 0$, $X_1 = 1$, $b_0 = 0$ and $I^{(n)}$ a mixture of multinomial distributions with $\sum_{1\leq r\leq 2^d} I_r^{(n)} = n - 1$. Various parameters of random quadtrees have systematically been studied in Flajolet et al. (1995) and in Chern, Fuchs and Hwang (2004). In particular, we have $\mathbb{E}\, X_n = \mu_d n + O(1)$ for $1 \leq d \leq 6$ with constants $\mu_d > 0$. Hence, Theorem 4.1 can be applied and we obtain subgaussian distributions for the number of leaves in d-dimensional random quadtrees for $d = 1, \ldots, 6$. The case $d = 1$ is the binary search tree case discussed above.

5 The case $Y \stackrel{d}{=} Y$

In this section we consider recursions (1.1) with $K = 1$,

$$X_n \stackrel{d}{=} X_{I_n} + b_n, \quad n \geq n_0, \tag{5.1}$$

with conditions as in (1.1) and the abbreviation $I_n = I_1^{(n)}$. We have subgaussian distributions for the following logarithmic growth.

Theorem 5.1 *Assume that $(X_n)_{n\geq 0}$ satisfies (5.1) and that for some $\eta < 1$, $\mu > 0$ and $n_1 \geq n_0$ we have*

$$\|X_0\|_\infty, \ldots, \|X_{n_0-1}\|_\infty < \infty, \quad \sup_{n\geq n_0} \|b_n\|_\infty < \infty,$$

$$\sup_{n\geq n_1} \mathbb{E} \left(\log \left(\frac{I_n \vee 1}{n} \right) \right)^2 < \infty, \tag{5.2}$$

$$\mathbb{E} \left(\frac{I_n \vee 2}{n} \right)^k \leq \eta^k, \; k \geq 1, \; n \geq n_1, \tag{5.3}$$

$$\mathbb{E}\, X_n = \mu \log n + O(1).$$

Then there exists an $L > 0$ *such that for all* $\lambda > 0$ *and* $n \geq 2$,

$$\mathbb{E}\exp\left(\lambda \frac{X_n - \mathbb{E}\, X_n}{\sqrt{\log n}}\right) \leq \exp(L\lambda^2).$$

In particular, we have (3.5) with $s(n) = \sqrt{\log n}$.

Proof: We denote

$$D_1 := |\,\mathbb{E}\, X_0| \vee \sup_{n\geq 1} |\,\mathbb{E}\, X_n - \mu \log n|, \quad D_2 := \sup_{n\geq n_0} \|b_n\|_\infty$$

and consider the scaled quantities

$$Y_n := \frac{X_n - \mathbb{E}\, X_n}{\sqrt{\log n}}, \quad n \geq 2,$$

and $Y_n := (\log 2)^{-1/2}(X_n - \mathbb{E}\, X_n)$ for $n = 0, 1$. Then (5.1) implies

$$Y_n \stackrel{d}{=} \sqrt{\frac{\log(I_n \vee 2)}{\log n}} Y_{I_n} + b^{(n)}, \quad n \geq n_0$$

with

$$b^{(n)} := \frac{1}{\sqrt{\log n}} (b_n + \mu \log((I_n \vee 1)/n) + R_n)$$

with a random R_n satisfying $|R_n| \leq 2D_1$.

We show that there exists an $L \geq 0$ such that $\mathbb{E}\exp(\lambda Y_n) \leq \exp(L\lambda^2)$ for all $\lambda > 0$ and $n \geq 0$. We proceed as in the proof of Theorem 4.1 by induction. Note, that all $X_0, \dots, X_{n_1-1}$ are uniformly bounded, so that the subgaussian distribution for $Y_0, \dots, Y_{n_1-1}$ follows form Hoeffding's Lemma as in the proof of Theorem 4.1. For the induction step we argue analogously to the proof of Theorem 4.1 to obtain

$$\mathbb{E}\exp(\lambda Y_n) \leq \exp(L\lambda^2)\,\mathbb{E}\exp\left(L\lambda^2\left(\frac{\log(I_n \vee 2)}{\log n} - 1\right) + \lambda b^{(n)}\right).$$

Hence, it is sufficient to show that

$$\sup_{n\geq n_1} \mathbb{E}\exp\left(\frac{L\lambda^2}{\log n}\log\left(\frac{I_n \vee 2}{n}\right) + \lambda b^{(n)}\right) \leq 1.$$

By the Cauchy–Schwarz inequality it is sufficient to show

$$\sup_{n\geq n_1} \mathbb{E}\exp\left(\frac{2Q\lambda^2}{\log n}\log\left(\frac{I_n \vee 2}{n}\right)\right) \mathbb{E}\exp(2\lambda b^{(n)}) \leq 1.$$

By Lemma 3.3 there exists a $Q \geq 0$ such that for all $n \geq n_1$ and $\lambda > 0$,

$$\begin{aligned}\mathbb{E}\exp(2\lambda b^{(n)}) &= \mathbb{E}\exp\left(\frac{2\lambda}{\sqrt{\log n}}(b_n + \mu\log((I_n \vee 1)/n) + R_n)\right)\\ &\leq \exp(Q\lambda^2/\log n),\end{aligned}$$

since $b_n + \mu \log((I_n \vee 1)/n) + R_n$ is centered, uniformly upper bounded and has uniformly bounded variance according to (5.2).

By condition (5.3) we obtain

$$\begin{aligned}\mathbb{E} \exp\left(\frac{2L\lambda^2}{\log n} \log\left(\frac{I_n \vee 2}{n}\right)\right) &= \mathbb{E} \left(\frac{I_n \vee 2}{n}\right)^{2L\lambda^2/\log n} \\ &\leq \eta^{2L\lambda^2/\log n} \\ &= \exp(2L\log(\eta)\lambda^2/\log n).\end{aligned}$$

Now the bound on the moment generating function follows choosing $L \geq Q/(2\log(1/\eta))$ and sufficiently large, so that the initial quantities $Y_0, \ldots, Y_{n_1-1}$ satisfy the same bound. □

Conditions (5.2), (5.3) require that I_n does not have too much mass on small or large values. This is somehow similar to the conditions (9) in Theorem 2.1 in Neininger and Rüschendorf (2004b), where a normal limit law for the same type of recurrences is studied. However, the conditions here in (5.2), (5.3) are more restrictive which makes the theorem less useful for practical applications. In particular, $I_n \stackrel{d}{=} \text{unif}\{0, \ldots, n-1\}$ does not satisfy (5.2), (5.3). Theorem 5.1 is more tailored for I_n that have, e.g., Binomial distributions $B(n-1, p)$ or distributions with similar tail properties as the Binomials. A typical application of Theorem 5.1 are depths of random nodes in asymmetric digital search trees, see, e.g., Louchard, Szpankowski and Tang (1999), where more refined estimates are given.

6 The case of deterministic A_r^* and $b^* \stackrel{d}{=} \mathcal{N}$:

In this section we consider $(X_n)_{n\geq 0}$ satisfying (1.1) so that after normalization as in (2.1) we obtain (2.2) with (2.3) and assume that we have the convergences in (2.4),

$$A_r^{(n)} \to A_r^*, \quad b^{(n)} \to b^*, \tag{6.1}$$

with deterministic $(A_1^*, \ldots, A_K^*)$ and b^* being normally $\mathcal{N}(\nu, \tau^2)$ distributed. It is easily checked that the arising fixed point equation (2.5) is then solved by a normal distribution if $\sum A_r^* < 1$. This allows to derive the following central limit theorem.

Theorem 6.1 *Assume that $(X_n)_{n\geq 0}$ satisfies (1.1) with $X_0, \ldots, X_{n_0-1}$ being L_1 integrable and that there are functions $m : \mathbb{N}_0 \to \mathbb{R}$ and $s : \mathbb{N}_0 \to \mathbb{R}_{>0}$ such that we have the convergences (6.1) weakly and with first absolute moment with deterministic $(A_1^*, \ldots, A_K^*)$, $0 < \sum_{r=1}^K A_r^* < 1$, and $b^* \sim \mathcal{N}(\nu, \tau^2)$, $\nu \in \mathbb{R}$, $\tau > 0$. Then we have*

$$\mathbb{E} X_n = m(n) + \mu s(n) + o(s(n)), \tag{6.2}$$

$$\frac{X_n - m(n)}{s(n)} \xrightarrow{d} \mathcal{N}(\mu, \sigma^2), \tag{6.3}$$

where

$$\mu = \frac{\nu}{1 - \sum_{r=1}^{K} A_r^*},$$

$$\sigma^2 = \frac{1}{1 - \sum_{r=1}^{K} (A_r^*)^2} \left(\tau^2 + \nu^2 + 2\nu\mu \sum_{r=1}^{K} A_r^* \right) - \mu^2 > 0. \tag{6.4}$$

If $X_0, \dots, X_{n_0 - 1}$ are moreover square integrable and the convergences (6.1) hold additionally with second moment, then

$$\mathrm{Var}(X_n) = \sigma^2 s^2(n) + o(s^2(n)).$$

Proof: The theorem is covered by general theorems of the contraction method. Parts (6.2) and (6.3) follow applying Theorem 5.1 in Neininger and Rüschendorf (2004a) with the parameter s there chosen to be $s = 1$ and noting that the fixed point equation (42) there is solved by $\mathcal{N}(\nu, \sigma^2)$ with μ, σ^2 as given in (6.4). Part (6.5) follows by applying that same Theorem 5.1 with $s = 2$. □

As an exemplary application we discuss the size of a random skip list, see, e.g., Pugh (1989), Papadakis, Munro, Poblete (1990), and Devroye (1992). Roughly, to build a skip list with parameter $p \in (0, 1)$, n elements are stored in a level 1 linked list. Each item of the level i list, $i \geq 1$, is included in the level $i + 1$ list independently with probability p. Certain pointers are used between the elements to support dictionary operations making skip lists a practical alternative to search trees. Here, we are only interested in the total number X_n of elements stored in the lists of all levels $i = 1, 2, \dots$ We call X_n the size of the random skip list for n elements. It satisfies (1.1) with $K = 1$, $I_1^{(n)} \sim B(n, p)$, $b_n = n$, $n_0 = 1$, and $X_0 = 0$. To apply Theorem 6.1 we choose $m(n) = (1/(1 - p))n$ and $s(n) = \sqrt{n} \vee 1$. By the strong law of large numbers we have

$$\frac{s(I_1^{(n)})}{s(n)} \to \sqrt{p}$$

almost surely, by the central limit theorem we have

$$\frac{1}{\sqrt{n}} \left(n - m(n) + m(I_1^{(n)}) \right) = \sqrt{\frac{p}{1-p}} \frac{I_1^{(n)} - pn}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}\left(0, \frac{p}{1-p}\right).$$

Note that both convergences also hold with first and second moment. Thus, Theorem 6.1 can be applied with $A_1^* = \sqrt{p}$ and $b^* = \mathcal{N}(0, p/(1 - p))$, and yields:

Corollary 6.2 *The size X_n of a random skip list with n elements and parameter $p \in (0, 1)$ satisfies*

$$\mathbb{E}\, X_n = \frac{1}{1-p} n + o(\sqrt{n}), \quad \mathrm{Var}(X_n) = \frac{p}{(1-p)^2} n + o(n) \tag{6.5}$$

and

$$\frac{X_n - (1-p)^{-1}n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, \frac{p}{(1-p)^2}\right).$$

For this particular recurrence a large deviation principle can directly be derived.

Theorem 6.3 *The size X_n of a random skip list with n elements and parameter $p \in (0, 1)$ satisfies for all $t > (1-p)^{-1}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_n > tn) = -I(t),$$

and for all $t < (1-p)^{-1}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_n < tn) = -I(t).$$

The rate function is given by

$$I(t) = \begin{cases} \left(\log \frac{1}{p}\right) t + (t-1)\log(t-1) - t\log(t) + \log \frac{p}{1-p}, & t \geq 1, \\ +\infty, & t < 1. \end{cases}$$

Proof: By construction, each of the n elements in the skip list is stored in a number of levels that is geometrically G_{1-p} distributed, i.e., $\mathbb{P}(G_{1-p} = k) = (1-p)p^{k-1}$, $k = 1, 2, \ldots$, and independent of the space requirements of the other elements, see Devroye (1992). Hence, X_n is distributed as a sum of n independent, identically G_{1-p} distributed random variables, thus X_n has the negative binomial distribution with parameters n and $1-p$. Cramér's theorem on large deviations applies. I as given in the theorem is the rate function of a G_{1-p} distributed random variable. □

From the perspective of the previous proof, Corollary 6.2 is directly implied by the central limit theorem for sums of independent random variables and it follows that both error terms in (6.5) are zero. However, since $I(t)/t \to \log(1/p)$ as $t \to \infty$, this application exemplifies different tails than the ones obtained in sections 4 and 5. Moreover, it gives an indication for the tails for slight perturbations of this recurrence, where a representation as a sum of independent random variables may not exist.

Theorem 6.1 can be applied to a series of problems that have been studied individually in the literature. In particular, it covers the number of coin flips in the "leader election problem", see Prodinger (1993) and Fill, Mahmoud and Szpankowski (1996), the number of coin flips for a maximum finding algorithm in a broadcast communication model, see Theorem 22 in Chen and Hwang (2003), the complexity of bucket selection, see Theorem 2 in Mahmoud, Flajolet, Jacquet, and Regniér (2000), and, with a slight modification, the distance of two randomly chosen nodes in a random binary search tree, see Mahmoud and Neininger (2003).

Acknowledgments
I thank Luc Devroye and Hsien-Kuei Hwang for helpful discussions and encouragement during the preparation of the manuscript and two referees for careful reading of the paper.

References

[1] Arratia, R., Barbour, A. D., and Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach.* EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich.

[2] Bennett, G. (1962) Probability inequalities for the sum of independent random variables. *J. Am. Stat. Assoc.* **57**, 33–45.

[3] Chen, W.-M. and Hwang, H.-K. (2003) Analysis in distribution of two randomized algorithms for finding the maximum in a broadcast communication model. *J. Algorithms* **46**, 140–177.

[4] Chern, H.-H., Fuchs, M., and Hwang, H.-K. (2004) Phase changes in random point quadtrees. Preprint.

[5] Chern, H.-H. and Hwang, H.-K. (2001) Phase changes in random m-ary search trees and generalized quicksort. *Random Struct. Algorithms* **19**, 316–358.

[6] Devroye, L. (1991) Limit laws for local counters in random binary search trees. *Random Struct. Algorithms* **2**, 303–315.

[7] Devroye, L. (1992) A limit theory for random skip lists. *Ann. Appl. Probab.* **2**, 597–609.

[8] Devroye, L. (2002/03) Limit laws for sums of functions of subtrees of random binary search trees. *SIAM J. Comput.* **32**, 152–171.

[9] Fill, J. A., Mahmoud, H. M., and Szpankowski, W. (1996) On the distribution for the duration of a randomized leader election algorithm. *Ann. Appl. Probab.* **6**, 1260–1283.

[10] Flajolet, P., Gourdon, X., and Martínez, C. (1997) Patterns in random binary search trees. *Random Struct. Algorithms* **11**, 223–244.

[11] Flajolet, P., Labelle, G., Laforest, L., and Salvy, B. (1995) Hypergeometrics and the cost structure of quadtrees. *Random Struct. Algorithms* **7**, 117–144.

[12] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30.

[13] Hwang, H.-K. (2003) Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.* **31**, 609–629.

[14] Hwang, H.-K. and Neininger, R. (2002) Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput.* **31**, 1687–1722.

[15] Karp, R. M. (1994) Probabilistic recurrence relations. *J. Assoc. Comput. Mach.* **41**, 1136–1150.

[16] Lew, W. and Mahmoud, H. M. (1994) The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.* **23**, 1050–1074.

[17] Louchard, G., Szpankowski, W., Tang, J. (1999) Average profile of the generalized digital search tree and the generalized Lempel–Ziv algorithm. *SIAM J. Comput.* **28**, 904–934.

[18] Mahmoud, H. M. (1986) The expected distribution of degrees in random binary search trees. *Comput. J.* **29**, 36–37.

[19] Mahmoud, H. M. (1992) *Evolution of Random Search Trees.* John Wiley & Sons Inc.

[20] Mahmoud, H. M., Flajolet, P., Jacquet, P., and Régnier, M. (2000) Analytic variations on bucket selection and sorting. *Acta Inf.* **36**, 735–760.

[21] Mahmoud, H. M. and Neininger, R. (2003) Distribution of distances in random binary search trees. *Ann. Appl. Probab.* **13**, 253–276.

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

[23] Neininger, R. and Rüschendorf, L. (2004a) A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.* **14**, 378–418.

[24] Neininger, R. and Rüschendorf, L. (2004b) On the contraction method with degenerate limit equation. *Ann. Probab.* **32**, 2838–2856.

[25] Papadakis, T., Munro, J. I., and Poblete, P. (1990) *Analysis of the expected search cost in skip lists.* SWAT 90 (Bergen, 1990), 160–172, Lecture Notes in Comput. Sci., 447. Springer, Berlin.

[26] Petrov, V. V. (1975) *Sums of Independent Random Variables.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer, Berlin–Heidelberg–New York.

[27] Prodinger, H. (1993) How to select a loser. *Discrete Math.* **120**, 149–159.

[28] Rachev, S. T. and Rüschendorf, L. (1995) Probability metrics and recursive algorithms. *Adv. Appl. Probab.* **27**, 770–799.

[29] Rösler, U. (1991) A limit theorem for "Quicksort". *RAIRO Inf. Théor. Appl.* **25**, 85–100.

[30] Rösler, U. (1992) A fixed point theorem for distributions. *Stochastic Processes Appl.* **42**, 195–214.

[31] Rösler, U. and Rüschendorf, L. (2001) The contraction method for recursive algorithms. *Algorithmica* **29**, 3–33.

[32] Pugh, W. (1989) Skip lists: a probabilistic alternative to balanced trees. *Algorithms and Data Structures (Ottawa, ON, 1989)*, 437–449, Lecture Notes in Comput. Sci., 382. Springer, Berlin.

[33] Sedgewick, R. and Flajolet, P. (1996) *An Introduction to the Analysis of Algorithms.* Addison-Wesley, Amsterdam.

[34] Szpankowski, W. (2001) *Average Case Analysis of Algorithms on Sequences.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York.

Ralph Neininger
Fachbereich Mathematik
J. W. Goethe Universität
Robert-Mayer-Str. 10
60325 Frankfurt a. M.
Germany
neiningr@math.uni-frankfurt.de