LIMIT LAWS FOR SUMS OF FUNCTIONS OF SUBTREES OF RANDOM BINARY SEARCH TREES

Luc Devroye
luc@cs.mcgill.ca

ABSTRACT. We consider sums of functions of subtrees of a random binary search tree, and obtain general laws of large numbers and central limit theorems. These sums correspond to random recurrences of the quicksort type, $X_n \stackrel{\mathcal{L}}{=} X_{I_n} + X'_{n-1-I_n} + Y_n$, $n \geq 1$, where I_n is uniformly distributed on $\{0, 1, \ldots, n-1\}$, Y_n is a given random variable, $X_k \stackrel{\mathcal{L}}{=} X'_k$ for all k, and given I_n, X_{I_n} and X'_{n-1-I_n} are independent. Conditions are derived such that $(X_n - \mu n)/\sigma\sqrt{n} \stackrel{\mathcal{L}}{\to} \mathcal{N}(0,1)$, the normal distribution, for some finite constants μ and σ.

KEYWORDS AND PHRASES. Binary search tree, data structures, probabilistic analysis, limit law, convergence, toll functions, Stein's method, random trees.

CR CATEGORIES: 3.74, 5.25, 5.5.

Research of the first author was sponsored by NSERC Grant A3456 and by FCAR Grant 90-ER-0291. Address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6.

Introduction.

In this note, we consider a random binary search tree with n nodes obtained by inserting, in the standard manner, the values $\sigma_1, \ldots, \sigma_n$ of a random permutation of $\{1, \ldots, n\}$ into an initially empty tree. Equivalently, the search tree is obtained by inserting n i.i.d. uniform $[0, 1]$ random variables $X_1, \ldots, X_n$. Most shape-related quantities of the tree have been well–studied, including the expected depth and the exact distribution of the depth of X_n (Knuth, 1973; Lynch, 1965), the limit theory for the depth (Mahmoud and Pittel, 1984, Devroye, 1988), the first two moments of the internal path length (Sedgewick, 1983), the limit theory for the height of the tree (Pittel, 1984; Devroye, 1986, 1987), and various connections with the theory of random permutations (Sedgewick, 1983) and the theory of records (Devroye, 1988). Surveys of known results can be found in Vitter and Flajolet (1990), Mahmoud (1992) and Gonnet (1984). Search trees are also useful in the analysis of quicksort. For recurrences of the quicksort type, we have $X_n \stackrel{\mathcal{L}}{=} X_{I_n} + X'_{n-1-I_n} + f(n)$, $n \geq 1$, $f(n) > 0$ for some $n > 0$, $f(0) = 0$, where I_n is uniformly distributed on $\{0, 1, \ldots, n-1\}$. X_n represents the number of comparisons in quicksort, and $f(n) = n - 1$. Other choices for $f(.)$ are of importance elsewhere. It is not hard to see that X_n is identical to the sum over all nodes u in a random binary search tree of $f(N(u))$ where $N(u)$ is the size of the subtree at u. The purpose of this note is to obtain central limit theorems for this class of random variables, regardless of the choice of f within a large class of functions.

In general, one might study the following class of tree parameters for random binary search trees: Let f be a mapping from the space of all permutations to the real line, and set

$$X_n = \sum_u f(S(u))$$

where $S(u)$ is the random permutation associated with the subtree rooted at node u in the random binary search tree. More precisely, model a random binary search tree as follows. We let $U_1, \ldots, U_n$ be i.i.d. uniform $[0, 1]$-valued random variables, and construct the unique binary search tree for $(1, U_1), \ldots, (n, U_n)$ with the property that

(i) It is a random binary search tree with respect to the first coordinates in the pairs.

(ii) It is a heap with respect to the second coordinates, which can be regarded as time stamps, with increasing values as one travels from the root down any path.

A permutation is clearly described by any subset of $(1, U_1), \ldots, (n, U_n)$. It is this unique description we follow. For example, the root of the binary search tree contains that pair (i, U_i) with smallest U_i value, the left subtree contains all pairs (j, U_j) with $j < i$, and the right subtree contains those pairs with $j > i$. Each node u can thus (recursively) be associated with a subset $S(u)$ of $(1, U_1), \ldots, (n, U_n)$. The pair that sticks with u is that with the smallest second component in $S(u)$.

With this embedding and representation, X_n is a sum over all nodes of a certain function of the permutation associated with each node. This definition is very broad. As each permutation uniquely determines subtree shape, a special case includes the functions of subtree shapes.

Example 1: The toll functions. In the first class of applications, we let $N(u)$ be the size of the subtree rooted at u (thus, if u is the overall root, $N(u) = n$), and set $f(S(u)) = g(|S(u)|)$. Define

$$X_n = \sum_u g(N(u)) \ .$$

Examples of such tree parameters abound:

A. If $g(n) \equiv 1$ for $n > 0$, then $X_n = n$.

B. If $g(n) = \mathbb{1}_{[n=k]}$ for fixed $k > 0$, then X_n counts the number of subtrees of size k.

C. If $g(n) = \mathbb{1}_{[n=1]}$, then X_n counts the number of leaves.

D. If $g(n) = n - 1$ for $n > 1$, then X_n counts the number of comparisons in classical quicksort. Note however that $g(n)$ grows too rapidly for us to be able to apply the theorem below.

E. If $g(n) = \log_2 n$ for $n > 0$, then X_n is the logarithm base two of the product of all subtree sizes.

F. If $g(n) = \mathbb{1}_{[n=1]} - \mathbb{1}_{[n=2]}$ for $n > 0$, then X_n counts the number of nodes in the tree that have two children, one of which is a leaf.

Example 2: Tree patterns. Fix a tree T. We write $S(u) \approx T$ if the subtree at u defined by the permutation $S(u)$ is equal to T, where equality of trees refers to shape only, not node labeling. Note that at least one, and possibly many permutations with $|S(u)| = |T|$, may give rise to T. If we set

$$X_n = \sum_u \mathbb{1}_{[S(u) \approx T]}$$

then X_n counts the number of subtrees precisely equal to T. Note that these subtrees are necessarily disjoint. We are tempted to call them suffix tree patterns, as they hug the bottom of the binary search tree.

EXAMPLE 3: PREFIX TREE PATTERNS. Fix a tree T. We write $S(u) \supset T$ if the subtree at u defined by the permutation $S(u)$ consists of T (rooted now at u) and possibly other nodes obtained by replacing all external nodes of T by new subtrees. Define

$$X_n = \sum_u \mathbb{1}_{[S(u) \supset T]}$$

For example, if T is a single node, then X_n counts the number of nodes, n. If T is a complete subtree of size $2^{k+1} - 1$ and height k, then X_n counts the number of occurrences of this complete subtree pattern (as if we try and count by sliding the complete tree to all nodes in turn to find a match). Matching complete subtrees can possibly overlap. If T consists of a single node and a right child, then X_n counts the number of nodes in the tree with just one right child.

EXAMPLE 4: IMABALANCE PARAMETERS. If we set $f(S(u))$ equal to 1 if and only if the sizes of the left and right subtrees of u are equal, then X_n counts the number of nodes at which we achieve a complete balance.

EXAMPLE 5: LOCAL COUNTERS. Following notation introduced by Devroye (1991), we may just elect to study indicator functions f with $f(S(u)) = 0$ if $|S(u))| > k$ for a fixed given k. In fact, the setting in Devroye (1991) is more general, as permutations are not necessarily restricted to those that correspond to nodes in the binary search tree.

In this paper, we study X_n. First we derive its mean and variance. This is followed by a weak law of large numbers for X_n/n. Several interesting examples illustrate this universal law. A general central limit theorem with normal limit is obtained for X_n using Stein's method. Several specific laws are obtained for particular choices of f. For example, for toll functions g as in Example 1, with $g(n)$ growing at a rate inferior to $n^{1/3}$, a universal central limit theorem is established in Theorem 6.

Another representation of binary search trees

We replace the sum over all nodes u in a random tree in the definition of X_n by a sum over a deterministic set of index pairs, thereby greatly facilitating systematic analysis. We denote by $\sigma(i,k)$ to subset $(i, U_i), \ldots, (i+k-1, U_{i+k-1})$, so that $|\sigma(i,k)| = k$. We define $\sigma^*(i,k) = \sigma(i-1, k+1)$, with the convention that $(0, U_0) = (0,0)$ and $(n+1, U_{n+1}) = (n+1, 0)$. Define the event

$$A_{i,k} = [\sigma(i,k) \text{ defines a subtree}] \ .$$

This event depends only on $\sigma^*(i,k)$, as $A_{i,k}$ happens if and only if among $U_{i-1},\ldots,U_{i+k}$, U_{i-1} and U_{i+k} are the two smallest values. We set $Y_{i,k} = \mathbb{1}_{[A_{i,k}]}$ and note that it is a function of $U_{i-1},\ldots,U_{i+k}$. Rewrite our tree parameter as follows:

$$X_n = \sum_u f(S(u)) = \sum_{i=1}^{n} \sum_{k=1}^{n-i+1} Y_{i,k} f(\sigma(i,k)) \ .$$

For example, in the toll function example with toll function g, this yields

$$X_n = \sum_u g(|S(u)|) = \sum_{i=1}^{n} \sum_{k=1}^{n-i+1} Y_{i,k} g(k) \ .$$

Mean and variance for toll functions

Let σ be a uniform random permutation of size k. Then define

$$\mu_k = \mathbb{E}\{f(\sigma)\} \ ,$$

$$\tau_k^2 = \mathbb{E}\{f^2(\sigma)\} \ ,$$

and

$$M_k = \sup_{\sigma:|\sigma|=k} |f(\sigma)| \ .$$

Note that $|\mu_k| \le \tau_k \le M_k$. In the toll function example, we have $\mu_k = g(k)$ and $\tau_k = M_k = |g(k)|$. We opt to develop the theory below in terms of these parameters. For some parts, such as the law of large numbers, the second moment approach may be avoided, but this comes at the expense of considerably more intricate computations and proofs.

Lemma 0. *Assume $|\mu_k| < \infty$ for all k, $\mu_k = o(k)$, and*

$$\sum_{k=1}^{\infty} \frac{|\mu_k|}{k^2} < \infty \ .$$

Define

$$\mu = \sum_{k=1}^{\infty} \frac{2\mu_k}{(k+2)(k+1)} \ .$$

Then

$$\lim_{n\to\infty} \frac{\mathbb{E}\{X_n\}}{n} = \mu \ .$$

If also $|\mu_k| = O(\sqrt{k}/\log k)$, then $\mathbb{E}\{X_n\} - \mu n = o(\sqrt{n})$.

Proof. We have

$$\begin{aligned}
\mathbb{E}\{X_n\} &= \sum_{i=1}^{n}\sum_{k=1}^{n-i+1}\mathbb{E}\{Y_{i,k}\}\mu_k \\
&= \sum_{i=2}^{n}\sum_{k=1}^{n-i}\frac{2}{(k+2)(k+1)}\mu_k + \sum_{k=1}^{n-1}\frac{1}{k+1}\mu_k + \sum_{i=1}^{n}\frac{1}{n-i+2}\mu_{n-i+1} + \mu_n
\end{aligned}$$

since

$$\mathbb{E}\{Y_{i,k}\} = \begin{cases} 1 & \text{if } i=1 \text{ and } i+k=n+1; \\ 1/(k+1) & \text{if } i=1 \text{ or } i+k=n+1 \text{ but not both;} \\ 2/(k+2)(k+1) & \text{otherwise.} \end{cases}$$

It is trivial to conclude the first part of Lemma 0. For the last part, we have:

$$\begin{aligned}
&|\mathbb{E}\{X_n - \mu n\}| \\
&\le \sum_{k=1}^{\infty}\frac{2}{(k+2)(k+1)}|\mu_k| + \sum_{i=2}^{n}\sum_{k=n-i+1}^{\infty}\frac{2}{(k+2)(k+1)}|\mu_k| + 2\sum_{k=1}^{n}\frac{|\mu_k|}{k+1} + |\mu_n| \\
&\le O(1) + \sum_{k=1}^{\infty}\frac{2\min(k,n)|\mu_k|}{(k+2)(k+1)} + 2\sum_{k=1}^{n}\frac{|\mu_k|}{k+1} + |\mu_n| \\
&\le O(1) + 4\sum_{k=1}^{n}\frac{|\mu_k|}{k+1} + n\sum_{k=n+1}^{\infty}\frac{|\mu_k|}{(k+2)(k+1)} + |\mu_n| \ . \ \square
\end{aligned}$$

Lemma 1. *Assume that $M_n < \infty$ for all n and that $f \ge 0$. Assume that for some $b \ge c \ge a > 0$, we have $\mu_n = O(n^a)$, $\tau_n = O(n^c)$, and $M_n = O(n^b)$. If $a+b<2$, $c<1$, then $\mathbb{V}\{X_n\} = o(n^2)$. If $a+b<1$, $c<1/2$, then $\mathbb{V}\{X_n\} = O(n)$. If f is a toll function and $M_n = O(n^b)$, then $\mathbb{V}\{X_n\} = o(n^2)$ if $b<1$ and $\mathbb{V}\{X_n\} = O(n)$ if $b < 1/2$.*

Proof. Let $Z_\alpha, \alpha \in A$, be a finite collection of random variables with finite second moments. Let E denote the collection of all pairs (α,β) from A^2 with $\alpha \ne \beta$ and Z_α not independent of Z_β. If $S = \sum_{\alpha\in A} Z_\alpha$, then

$$\mathbb{V}\{S\} = \sum_{\alpha\in A}\mathbb{V}\{Z_\alpha\} + \sum_{(\alpha,\beta)\in E}\left(\mathbb{E}\{Z_\alpha Z_\beta\} - \mathbb{E}\{Z_\alpha\}\mathbb{E}\{Z_\beta\}\right) \ .$$

We apply this fact with A being the collection of all pairs (i,k), with $1 \le i \le n$ and $1 \le k \le n-i+1$. Let our collection of random variables be the products $Y_{i,k} f(\sigma(i,k))$, $(i,k) \in V$. Note that E consists only of pairs $((i,k),(j,\ell))$ from A^2 with $i+k \ge j-1$ and $j+\ell \ge i$. This means that the intervals $[i, i+k-1]$ and $[j, j+\ell-1]$ correspond to an element of E if and only if they overlap or are disjoint and separated by exactly zero or one integer m. But to bound $\mathbb{V}\{X_n\}$ from above, since $f \geq 0$, we have

$$\mathbb{V}\{X_n\} \leq \sum_{(i,k)\in A} \mathbb{V}\{Y_{i,k} f(\sigma(i,k))\} + \sum_{((i,k),(j,\ell))\in E} \mathbb{E}\{Y_{i,k} f(\sigma(i,k)) Y_{j,\ell} f(\sigma(j,\ell))\} = I + II \ .$$

By the independence of $Y_{i,k}$ and $f(\sigma(i,k))$, we have

$$\begin{aligned}
\mathbb{V}\{Y_{i,k} f(\sigma(i,k))\} &= \mathbb{V}\{Y_{i,k}\}\mathbb{E}\{f^2(\sigma(i,k))\} + (\mathbb{E}\{Y_{i,k}\})^2 \mathbb{V}\{f(\sigma(i,k))\} \\
&= \mathbb{V}\{Y_{i,k}\}\tau_k^2 + (\mathbb{E}\{Y_{i,k}\})^2(\tau_k^2 - \mu_k^2) \\
&\leq \mathbb{E}\{Y_{i,k}\}\tau_k^2
\end{aligned}$$

and thus $I = O(n)$ if $\tau_n^2 = O(n)$, $\sum_{k=1}^n \tau_k^2/k = O(n)$ and $\sum_k \tau_k^2/k^2 < \infty$. These conditions hold if $c < 1/2$. We have $I = o(n^2)$ if $c < 1$.

In II, we have $Y_{i,k}Y_{j,\ell} = 0$ unless the intervals $[i, i+k-1]$ and $[j, j+\ell-1]$ are disjoint and precisely one integer apart, or nested. For disjoint intervals, we note the independence of $Y_{i,k}Y_{j,\ell}$, $f(\sigma(i,k))$ and $f(\sigma(j,\ell))$, so that

$$\mathbb{E}\{Y_{i,k} f(\sigma(i,k)) Y_{j,\ell} f(\sigma(j,\ell))\} = \mathbb{E}\{Y_{i,k}Y_{j,\ell}\}\mu_k\mu_\ell \ .$$

If none of the intervals contains 1 or n, then a brief argument shows that

$$\mathbb{E}\{Y_{i,k}Y_{j,\ell}\} \leq \frac{4}{(k+\ell+3)(k+1)(\ell+1)} \ .$$

If one interval covers 1 and the other n, then $k + \ell = n - 1$, and $\mathbb{E}\{Y_{i,k}Y_{j,\ell}\} = 1/n$. In the other cases, the expected value is bounded by $2/(k+\ell+2)(k+1)$ or $2/(k+\ell+2)(\ell+1)$, depending upon which interval covers 1 or n. Thus, the sum in II limited to disjoint intervals is bounded by

$$\begin{aligned}
&n \sum_{k=1}^n \sum_{\ell=1}^n \frac{4\mu_k\mu_\ell}{(k+\ell+3)(k+1)(\ell+1)} + 1 + \sum_{k=1}^n \sum_{\ell=1}^n \frac{4\mu_k\mu_\ell}{(k+\ell+2)(k+1)} \\
&\leq 2n \sum_{k=1}^n \sum_{\ell=1}^k \frac{4\mu_k\mu_\ell}{(k+3)(k+1)(\ell+1)} + 1 + 2\sum_{k=1}^n \sum_{\ell=1}^k \frac{4\mu_k\mu_\ell}{(k+2)(k+1)} \ .
\end{aligned}$$

If $\mu_n = O(n^a)$ for $a > 0$, then it is easy to see that the three sums taken together are $O(n^{2a})$.

We next consider nested intervals. For properly nested intervals, with $[i, i+k-1]$ being the bigger one, we have

$$\begin{aligned}
\mathbb{E}\{Y_{i,k} f(\sigma(i,k)) Y_{j,\ell} f(\sigma(j,\ell))\} &= \mathbb{E}\{Y_{i,k}\}\mathbb{E}\{f(\sigma(i,k)) Y_{j,\ell} f(\sigma(j,\ell))\} \\
&\leq \frac{2M_k \mathbb{E}\{Y_{i,k}\}\mu_\ell}{(\ell+2)(\ell+1)} \ .
\end{aligned}$$

Summed over all allowable pairs $(i,k),(j,\ell)$ with the outer interval not covering 1 or n, and noting that in all cases considered, $a<1$,this yields a quantity not exceeding

$$n\sum_{k=1}^{n} k\sum_{\ell=1}^{k} \frac{4M_k\mu_\ell}{(\ell+2)(\ell+1)(k+2)(k+1)} \le n\sum_{k=1}^{n} M_k O(k^{a-2}) = \begin{cases} O(n^{b+a}) & \text{if } b+a\neq 1\ , \\ O(n\log n) & \text{if } b+a=1. \end{cases}$$

The contribution of the border effect is of the same order. This is $o(n^2)$ if $a+b<2$. It is $O(n)$ if $a+b\le 1$.

Finally, we consider nested intervals with $i=j$ and $\ell<k$. Then

$$\mathbb{E}\{Y_{i,k}f(\sigma(i,k))Y_{j,\ell}f(\sigma(j,\ell))\} \le \mathbb{E}\{Y_{i,k}\}M_k\frac{\mu_\ell}{\ell+1}\ .$$

Summed over all appropriate (i,k,ℓ) such that the outer interval does not cover 1 or n, we obtain a bound of

$$n\sum_{k=1}^{n}\sum_{\ell=1}^{k} \frac{2M_k\mu_\ell}{(k+2)(k+1)(\ell+1)} = O(n^{a+b}+\mathbb{1}_{[a+b=1]}n\log n)\ .$$

The border cases do not alter this bound. Thus, the contribution to II for these nested intervals is $o(n^2)$ if $a+b<2$ and is $O(n)$ if $a+b<1$. $\square$

A law of large numbers

The estimates of the previous section permit us to obtain a law of large numbers.

THEOREM 1. *Assume that $M_n<\infty$ for all n and that $f\ge 0$. Assume that for some $b\ge c\ge a>0$, we have $\mu_n=O(n^a)$, $\tau_n=O(n^c)$, and $M_n=O(n^b)$. If $a+b<2$, $c<1$, then*

$$\frac{X_n}{n}\to\mu$$

in probability. If f is a toll function and $M_n=O(n^b)$, then $X_n/n\to\mu$ in probability when $b<1$.

PROOF. Note that $a < 1$. By Lemma 0, we have $\mathbb{E}\{X_n\}/n \to \mu$. Choose $\epsilon > 0$. By Chebyshev's inequality and Lemma 1,

$$\mathbb{P}\{|X_n - \mathbb{E}\{X_n\}| > \epsilon n\} \leq \frac{\mathbb{V}\{X_n\}}{\epsilon^2 n^2} = o(1) \ .$$

Thus, $X_n/n - \mathbb{E}\{X_n\}/n \to 0$ in probability. $\square$

Four examples will illustrate this result.

EXAMPLE 1. We let f be the indicator function of anything, and note that the law of large numbers holds. For example, let $\mathcal{T}$ be a possibly infinite collection of possible tree patterns, and let X_n count the number of subtrees in a random binary search tree that match a tree from $\mathcal{T}$. Then, as shown below, the law of large numbers holds. There is inherent limitation to $\mathcal{T}$, which, in fact, might be the collection of all trees whose size is a perfect square and whose height is a prime number at the same time. Let X_n be the number of subtrees in a random binary search tree that match a given prefix tree pattern T, with $|T| = k$ fixed.

THEOREM 2. *For any non-empty tree pattern collection $\mathcal{T}$, we have*

$$\frac{X_n}{n} \to \mu$$

in probability, and $\mathbb{E}\{X_n\}/n \to \mu$, where

$$\mu = \sum_{n=1}^{\infty} \frac{2\mu_n}{(n+2)(n+1)}$$

and μ_n is the probability that a random binary search tree of size n matches an element of $\mathcal{T}$.

PROOF. Theorem 1 applies since f is an indicator function. By Lemma 0, we obtain the limit μ for $\mathbb{E}\{X_n\}/n$. $\square$

Note that Theorem 2 remains valid if we replace the phrase "matches an element of $\mathcal{T}$" by the phrase "matches an element of $\mathcal{T}$ at its root", so that $\mathcal{T}$ is a collection of what we called earlier prefix tree patterns.

EXAMPLE 2. Perhaps more instructive is the example of the sumheight $\mathcal{S}_n$, the sum of the heights of all subtrees in a random binary search tree on n nodes.

Theorem 3. *For a random binary search tree, the sumheight satisfies*

$$\frac{\mathcal{S}_n}{\mathbb{E}\{\mathcal{S}_n\}} \to 1$$

in probability. Here

$$\mathbb{E}\{\mathcal{S}_n\} \sim n \sum_{k=1}^{\infty} \frac{2h_k}{(k+2)(k+1)} ,$$

where h_k is the expected height of a random binary search tree on k nodes.

Proof. The statement about the expected height follows from Lemma 0 without work. As the height of a subtree of size k is at most $k-1$, we see that we may apply Theorem 1 with $M_k = k-1$. By well-known results (Robson, 1977; Pittel, 1984; Devroye, 1986, 1987), we have $\mathbb{E}\{H_n^2\} = O(\log^2 n)$ where H_n is the height of a random binary search tree. Thus, we may formally take a and c arbitrarily small but positive, and $b = 1$. $\square$

Example 3. Define $L(u)$ to be the largest number of full levels below u, and let $C(u) = 2^{L(u)+1} - 1$ be the size of that largest full subtree rooted at u. Define

$$X_n = \sum_u C(u) .$$

This parameter measures to some extent the amount of balance in the tree.

Theorem 4. *For a random binary search tree,*

$$\frac{X_n}{n} \to \mu$$

in probability, and $\mathbb{E}\{X_n\}/n \to \mu$, where

$$\mu = \sum_{n=1}^{\infty} \frac{2\mu_n}{(n+2)(n+1)}$$

and μ_n is the expected size of the largest complete subtree rooted at the root of a random binary search tree of size n.

PROOF. We verify that Theorem 1 and Lemma 0 may be applied with $a = 0.35$, $b = 1$ and $c = 0.35$. Indeed, if H_n is the number of full levels starting at a level below the root in a random binary search tree on n nodes, we know from Devroye (1986) that $L_n/\log n \to \gamma = 0.3711\cdots$ in probability and in the mean. This result does not suffice, as we need to show that $\mathbb{E}\{2^{L_n}\} = O(n^a)$ with $a = 0.35$. But using a representation for tree sizes in terms of products of independent uniform $[0,1]$ random variables $U_1, \ldots, U_n$ (Devroye, 1986) (the tree size for any node at distance k from the root is distributed as $\lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_k \rfloor$), we see that

$$
\begin{aligned}
\mathbb{P}\{L_n \geq k\} &\leq \left(\mathbb{P}\{nU_1 \cdots U_k \geq 1\}\right)^{2^k} \\
&\leq \exp\left(-2^k \mathbb{P}\{ne^{-G_k} < 1\}\right) \\
&\leq \exp\left(-2^k \mathbb{P}\{G_k > \log n\}\right) \\
&\leq \exp\left(-2^k \int_{\log n}^{\infty} y^{k-1}/(k-1)! e^{-y} dy\right) \\
&\leq \exp\left(-2^k (\log n)^{k-1}/(k-1)! n\right) \\
&\leq \exp\left(-2^k (\log n)^{k-1} \sqrt{k}/(k/e)^k e\sqrt{2\pi} n\right) \\
&\qquad \text{(by Stirling's approximation)} \\
&\leq \exp\left(-(2e \log n/k)^k \sqrt{k}/e\sqrt{2\pi} n \log n\right) \\
&= \exp\left(-(2e^{1-1/c}/c)^{c \log n} \sqrt{c}/e\sqrt{2\pi \log n}\right) \\
&\qquad \text{(after setting } k = c \log n\text{)} \\
&\leq \exp\left(-n^{\log \sqrt{4/e}}/e\sqrt{4\pi \log n}\right) \\
&\qquad \text{(by the choice } c = 1/2\text{)}
\end{aligned}
$$

Clearly, then,

$$
\mathbb{E}\{2^{L_n}\} \leq n\mathbb{P}\{L_n \geq (1/2)\log n\} + 2^{(1/2)\log n} = o(1) + n^{\log \sqrt{2}} = o(n^{0.35}) \ .
$$

We also have $\mathbb{E}\{2^{2L_n}\} = o(n^{0.7})$ by the same argument. Thus, both the conditions of Lemma 0 and Theorem 1 are satisfied and the law of large numbers follows. $\square$

EXAMPLE 4. Consider $X_n = \sum_u (N(u))^{0.999}$. Recall that $\sum_u N(u)$ is the number of comparisons in quicksort, plus n. Thus, X_n is a discounted parameter with respect to the number of quicksort comparisons. Clearly, Theorem 1 applies with $a = b = c = 0.999$, and thus, $X_n/n \to \mu$ in probability, and $\mathbb{E}\{X_n\}/n$ tends to the same constant μ. In a sense, this application is near the limit of the range for Theorem 1. For example, it is known that with $X_n = \sum_u (N(u))^{1+\epsilon}$, there is no asymptotic concentration, and thus, $X_n/g(n)$ does not converge to a constant for any choice of $g(n)$. Also, for $X_n = \sum_u (N(u))^1$, the quicksort example, we have $X_n/2n\log n \to 1$ in probability (Sedgewick, 1983), so that once again Theorem 1 is not applicable. Therefore, in a vague sense, the conditions of Theorem 1 are nearly best possible, as the theorem applies to $X_n = \sum_u (N(u))^{1-\epsilon}$. with $\epsilon \in (0,1]$.

Dependency graph

We will require the notion of a dependency graph for a collection of random variables $(Z_\alpha)_{\alpha\in V}$, where V is a set of vertices. Let the edge set E be such that for all disjoint subsets A and B of V, either there is an edge of E between A and B, or there is no edge, and in the latter case, $(Z_\alpha)_{\alpha\in A}$ and $(Z_\alpha)_{\alpha\in B}$ are mutually independent. Clearly, the complete graph is a dependency graph for any set of random variables, but this is useless. One usually takes the minimal graph (V, E) that has the above property, or one tries to keep $|E|$ as small as possible. Note that necessarily, Z_α and Z_β are independent if $(\alpha, \beta) \notin E$, but to have a dependency graph requires much more than just checking pairwise independence. We call the neighborhood of $N(\alpha)$ of vertex $\alpha \in V$ the collection of vertices β such that $(\alpha, \beta) \in E$ or $\alpha = \beta$. We define the neighborhood $N(\alpha_1, \ldots, \alpha_r)$ as $\cup_{j=1}^r N(\alpha_j)$.

A. Consider now for V the pairs (i, k) with $1 \le i \le n$ and $1 \le k \le n-i+1$. Let our collection of random variables be the permutations $\sigma(i, k)$, $(i, k) \in V$. Let us connect (i, k) to (j, ℓ) when $i+k \ge j-1$ and $j+\ell \ge i$. This means that the intervals $(i, i+k-1)$ and $(j, j+\ell-1)$ correspond to an edge in E if and only if they overlap or are disjoint and separated by exactly zero or one integer m. We claim that (V, E) is a dependency graph. Indeed, if we consider disjoint subsets A and B of vertices with no edges between them, then these vertices correspond to intervals that are pairwise separated by at least two integers, and thus, $(\sigma(i, k))_{(i,k)\in A}$ and $(\sigma(j, \ell))_{(j,\ell)\in B}$ are mutually independent.

B. Consider next the collection of random variables $Y_{i,k}g(k)$. For this collection, we can make a smaller dependency graph. Eliminate all edges from the graph of the previous paragraph if the intervals defined by the endpoints of the edges are properly nested. For example, if $i < j < j+\ell-1 < i+k$, then the edge between (i, k) and (j, ℓ) is removed. The graph thus obtained is still a dependency graph. This observation repeatedly uses the fact that if one considers a sequence $Z_1, \ldots, Z_n$ of i.i.d. random variables with a uniform $[0, 1]$ distribution, then Z_1, Z_n and the permutation of $Z_2, \ldots, Z_{n-1}$ are all independent. Thus, for properly nested intervals as above, $Y_{i,k}g(k)$ is independent of $Y_{j,\ell}g(\ell)$.

C. A third dependency graph that will be useful is constructed as above when V is restricted to those pairs (i, k) with $1 \le i \le n$ and $1 \le k \le n-i+1$, and, additionally, $k \le K$. Typically, $K = o(n)$, so this will restrict the degree of each vertex in the dependency graph. For example, given any vertex (i, k) in this graph, its neighborhood $N((i, k))$ has cardinality bounded by $(2K+2)K$, because the starting point for a connected interval has at most $2K+2$ choices, and the length at most K.

Stein's method

Stein's method (Stein, 1972) allows one to deduce a normal limit law for certain sums of random variables while only computing first and second order moments and verifying a certain dependence condition. Many variants have seen the light of day in recent years, and we will simply employ the following version derived in Janson, Łuczak and Ruciński (2000, Theorem 6.33):

LEMMA 2. *Suppose that $(S_n)_1^\infty$ is a sequence of random variables such that $S_n = \sum_{\alpha\in V_n} Z_{n\alpha}$, where for each n, $\{Z_{n\alpha}\}_\alpha$ is a family of random variables with dependency graph (V_n, E_n). Let $N(.)$ denote the neighborhood of a vertex or vertices. Suppose further that there exist numbers M_n and Q_n such that*

$$\sum_{\alpha\in V_n} \mathbb{E}\{|Z_{n\alpha}|\} \le M_n$$

and for every $\alpha, \alpha' \in V_n$:

$$\sum_{\beta\in N(\alpha,\alpha')} \mathbb{E}\{|Z_{n\beta}||Z_{n\alpha}, Z_{n\alpha'}\} \le Q_n \ .$$

Let $\sigma_n^2 = \mathbb{V}\{S_n\}$. Then

$$\frac{S_n - \mathbb{E}\{S_n\}}{\sqrt{\mathbb{V}\{S_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

if

$$\lim_{n\to\infty} \frac{M_n Q_n^2}{\sigma_n^3} = 0 \ .$$

Sums of functions of subtrees

In this section, we apply Stein's method to the random variable

$$X_{n,K} = \sum_{u:|S(u)|\le K} f(S(u)) \ ,$$

where $K = K(n)$ is a sequence of positive numbers. By verifying the conditions of Lemma 1, we obtain

LEMMA 3. *Let g be a nondecreasing positive function such that $|f(\sigma)| \le g(|\sigma|)$ for all permutations σ. Assume that $\mathbb{V}\{X_{n,K}\} = \Omega(n)$, where $K = K(n)$ be a sequence such that $K \ge 1$, yet*

$$K^2 g(K) = o(n^{1/4}) \ .$$

Then

$$\frac{X_{n,K} - \mathbb{E}\{X_{n,K}\}}{\sqrt{\mathbb{V}\{X_{n,K}\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

PROOF. We apply Lemma 2 with basic collection of random variables $Y_{i,k} f(\sigma(i,k))$, $(i,k) \in V_n$, where V_n is the collection $\{(i,k) : 1 \le i \le n, 1 \le k \le \min(K, n-i+1)\}$. Let E_n be the edges in the dependency graph L_n defined by connecting (i,k) to (j,ℓ) if the respective intervals are overlapping or if the respective intervals are disjoint with zero or one integers separating them. We note that

$$\sum_{(i,k)\in V_n} \mathbb{E}\{Y_{i,k} g(k)\} \le (\nu + o(1))n$$

by computations not unlike those for the mean done earlier, where

$$\nu = \sum_{n=1}^{\infty} \frac{2g(n)}{(n+2)(n+1)} \ .$$

To apply Lemma 2, we note that we may thus take $M_n = O(n)$. We also note that $\sigma_n^2 = \Omega(n)$, by assumption. Define

$$Q_n = \sup_{(i,k),(j,\ell)\in V_n} \sum_{(p,r)\in N((i,k),(j,\ell))} \mathbb{E}\{Y_{p,r} g(r) | Y_{i,k} f(\sigma(i,k)), Y_{j,\ell} f(\sigma(j,\ell))\} \ .$$

Indeed, as g bounds $|f|$, this is all we need to bound. The technical condition in Lemma 2 is satisfied if $Q_n = o(n^{1/4})$. To compute an upper bound for Q_n, we bound as follows:

$$Q_n \le \sup_{(i,k),(j,\ell)\in V_n} |N((i,k),(j,\ell))| g(K)$$

Each of the intervals represented by (i,k) and (j,ℓ) has length at most K. Clearly, $(p,r) \in N((i,k),(j,\ell))$ means that both p and $p+r-1$ must be in these intervals or within the $K+1$ neighbors of them. Each of p and r has thus at most $3K+2$ choices, so that $|N((i,k),(j,\ell))| \le (3K+2)^2$. Thus,

$$Q_n \le (3K+2)^2 g(K)$$

from which Lemma 3 follows without further work. □

A simple example counts the number of subtrees in a random binary search tree that match one of a given collection of tree patterns (these are "terminal matches" at the bottom of the tree), where each pattern is of size $\le K$, where K may depend upon n. As f is an indicator function, we may take $g \equiv 1$. Let X_n denote the number of matches.

Lemma 4. *Let X_n be the number of matches of a tree pattern in a collection of tree patterns depending arbitrarily on n as long as within the collection, the maximal tree size is $K = K(n)$. If*

$$\lim_{n\to\infty} \frac{\mathbb{E}\{X_n\}K^2}{(\mathbb{V}\{X_n\})^{3/2}} = 0 ,$$

then

$$\frac{X_n - \mathbb{E}\{X_n\}}{\sqrt{\mathbb{V}\{X_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) .$$

Proof. Follow the proof of Lemma 3, but do not use the estimate $\sigma_n^2 = \Omega(n)$. $\square$

Note that the above result remains true even if the collection of patterns itself is a function of n, changing in cardinality and in membership with n, within the condition imposed on K. This result extends the central limit laws of Devroye (1991), where K had to remain fixed. Indeed, the technical condition of Lemma 4 becomes $\mathbb{V}\{X_n\}/\mathbb{E}^{2/3}\{X_n\} \to \infty$. Note in this respect that for K fixed, and the collection of tree patterns non-empty for all n, we have $\mathbb{V}\{X_n\} = \Theta(n)$, and $\mathbb{E}\{X_n\} = \Theta(n)$, facts that are easy to verify.

Sums of indicator functions

In this section, we take a simple example, in which

$$X_n = \sum_u \mathbb{1}_{[S(u)\in A_n]} ,$$

where A_n is a non-empty collection of permutations of length k, with k possibly depending upon n. We denote $p_{n,k} = |A_n|/k!$, the probability that a randomly picked permutation of length k is in the collection A_n. Particular examples include sets A_n that correspond to a particular tree pattern, in which case X_n counts the number of occurrences of a given tree pattern of size k (a "terminal pattern") in a random binary search tree. The interest here is in the case of varying k. As we will see below, for a central limit law, k has to be severely restricted.

Our main result is this:

THEOREM 5. *We have*

$$\mathbb{E}\{X_n\} = \frac{2np_{n,k}}{(k+2)(k+1)} + O(1)$$

regardless of how k varies with n. If $k = o(\log n / \log\log n)$, then $\mathbb{E}\{X_n\} \to \infty$, $X_n/\mathbb{E}\{X_n\} \to 1$ in probability, and

$$\frac{X_n - \mathbb{E}\{X_n\}}{\sqrt{\mathbb{V}\{X_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

PROOF. Observe that

$$X_n = \sum_{i=1}^{n-k+1} Y_{i,k} Z_i \ .$$

where $Z_i = \mathbb{1}_{[\sigma(i,k)\in A_n]}$, Thus,

$$\mathbb{E}\{X_n\} = \sum_{i=2}^{n-k} \frac{2}{(k+2)(k+1)} \mathbb{E}\{Z_1\} + 2 \times \frac{1}{k+1} \mathbb{E}\{Z_1\} = \frac{2(n-k-1)p_{n,k}}{(k+2)(k+1)} + \frac{2p_{n,k}}{k+1} \ .$$

This proves the first part of the Theorem.

The computation of the variance is slightly more involved. However, it is simplified by considering the variance of

$$Y_n = \sum_{i=2}^{n-k} Y_{i,k} Z_i \ ,$$

and noting that $|X_n - Y_n| \le 2$. This eliminates the border effect. We note that $Y_{i,k}Y_{j,k} = 0$ if $i < j \le i+k$. Thus,

$$\begin{aligned}
\mathbb{E}\{Y_n^2\} &= \sum_{i=2}^{n-k} \mathbb{E}\{Y_{i,k}Z_i\} + 2 \sum_{2\le i<j\le n-k} \mathbb{E}\{Y_{i,k}Z_iY_{j,k}Z_j\} \\
&= \mathbb{E}\{Y_n\} + 2 \sum_{2\le i, i+k+1 \le n-k} \mathbb{E}\{Y_{i,k}Z_iY_{i+k+1,k}Z_{i+k+1}\} + 2 \sum_{2\le i, i+k+1 < j \le n-k} \mathbb{E}\{Y_{i,k}Z_i\}\mathbb{E}\{Y_{j,k}Z_j\} \\
&= (n-k-1)\beta + 2(n-2k-2)\alpha + (n-2k)^2\beta^2 + (10k+6-5n)\beta^2
\end{aligned}$$

where $\alpha = \mathbb{E}\{Y_{2,k}Z_2Y_{3+k,k}Z_{3+k}\}$, and $\beta = \mathbb{E}\{Y_{2,k}Z_2\}$. Also,

$$(\mathbb{E}\{Y_n\})^2 = ((n-k-1)\beta)^2 \ .$$

Thus,

$$\begin{aligned}
\mathbb{V}\{Y_n\} &= 2(n-2k-2)\alpha + (n-k-1)\beta + \left((n-2k)^2 - (n-k-1)^2 + (10k+6-5n)\right)\beta^2 \\
&= n\left(2\alpha + \beta - (2k+3)\beta^2\right) + O(k\alpha + k\beta + k^2\beta^2) \ .
\end{aligned}$$

We note that $\beta = 2p_{n,k}/(k+2)(k+1)$. To compute α, let A, B, C be the minimal values among $U_1, \ldots, U_{k+1}$; U_{k+2}, and $U_{k+3}, \ldots, U_{2k+3}$, respectively. Clearly,

$$\alpha = p_{n,k}^2 \mathbb{E}\{Y_{2,k}Y_{3+k,k}\}.$$

Considering all six permutations of A, B, C separately, one may compute the latter expected value as

$$\frac{2}{2k+3}\frac{1}{2k+2}\frac{1}{k+1} + \frac{1}{2k+3}\frac{1}{k+1}\frac{1}{k+1} + \frac{2}{2k+3}\frac{1}{2k+2}\frac{1}{2k+1} = \frac{5k+3}{(2k+3)(2k+1)(k+1)^2} \ .$$

Thus,

$$\alpha = \frac{(5k+3)p_{n,k}^2}{(2k+3)(2k+1)(k+1)^2} \ .$$

We have

$$\begin{aligned}\mathbb{V}\{Y_n\} &= \\ &= n\left(p_{n,k}^2\frac{10k+6}{(2k+3)(2k+1)(k+1)^2} - p_{n,k}^2\frac{8k+12}{(k+2)^2(k+1)^2} + p_{n,k}\frac{2}{(k+2)(k+1)}\right) + O(p_{n,k}/k) \ .\end{aligned}$$

Note that regardless of the value of $p_{n,k}$, the coefficient of n is strictly positive. Indeed, the coefficient is at least

$$\begin{aligned}& p_{n,k}^2\left(\frac{(10k+6)(k+2)^2 - (8k-8)(2k+3)(2k+1) + 2(2k+3)(2k+1)(k+2)(k+1)}{(k+2)^2(k+1)^2(2k+3)(2k+1)}\right) \\ &= p_{n,k}^2\left(\frac{8k^4+18k^3+4k^2-6k}{(k+2)^2(k+1)^2(2k+3)(2k+1)}\right) \ .\end{aligned}$$

Thus, there exist universal constants $c_1, c_2, c_3 > 0$ such that

$$\mathbb{V}\{Y_n\} \geq c_1 n p_{n,k}^2/k^2 - c_2 p_{n,k}/k \ ,$$

and

$$\mathbb{V}\{Y_n\} \leq c_3 n p_{n,k}/k^2 \ .$$

We have $Y_n/\mathbb{E}\{Y_n\} \to 1$ in probability if $\mathbb{V}\{Y_n\} = o(\mathbb{E}^2\{Y_n\})$, that is, if $k = o(\sqrt{np_{n,k}})$. Using $p_{n,k} \geq 1/k!$, we note that this condition holds if $k = o(\log n/\log\log n)$.

Finally, we turn to the normal limit law and note that

$$\frac{Y_n - \mathbb{E}\{Y_n\}}{\sqrt{\mathbb{V}\{Y_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

if (see Lemma 4)

$$\lim_{n\to\infty} \frac{k^2\mathbb{E}\{Y_n\}}{(\mathbb{V}\{Y_n\})^{3/2}} = 0 \ .$$

This holds if

$$\lim_{n\to\infty} \frac{np_{n,k}}{n^{3/2}p_{n,k}^3/k^3} = 0$$

and $np_{n,k}/k \to \infty$. Both conditions are satisfied if $k = o(\log n/\log\log n)$. □

The limitation on k in Theorem 5 cannot be lifted without further conditions: indeed, if we consider as a tree pattern the tree that consists of a right branch of length k only, then $\mathbb{E}\{X_n\} \to 0$ if $k > (1+\epsilon)\log n/\log\log n$ for any given fixed $\epsilon > 0$. As X_n is integer-valued, no meaningul limit laws can exist in such cases.

Notes on the variance

For toll functions, that is, functions f such that $f(\sigma) = g(|\sigma|)$ for some function g, we need the following Lemma.

LEMMA 5. *Define* $X_n = \sum_u g(|S(u)|)$ *and* $X_{n,K} = \sum_u g(|S(u)|)\mathbb{1}_{[|S(u)|\le K]}$ *for a random binary search tree. The following statements are equivalent:*

A. $\mathbb{V}\{X_{n,K}\} = \Omega(n)$ *for any* K *with* $K \to \infty$ *as* $n \to \infty$.

B. $\mathbb{V}\{X_n\} = \Omega(n)$.

C. $\mathbb{V}\{X_k\} > 0$ *for some* $k > 0$.

D. The function g *is not constant on* $\{1, 2, \ldots\}$.

PROOF. D implies C. Indeed, let k be the first integer at least equal to 2 such that

$$g(k) \neq g(k-1) = \cdots = g(1) \ .$$

For the integers up to k, we have the representation $g(x) = c + d\mathbb{1}_{[x=k]}$ with $d \neq 0$. We have $X_i = ic$, $i < k$, $X_k = kc + d$, and $X_{k+1} = f(k+1) + kc + dN_k$, where N_k is the number of nodes for which $|S(u)| = k$. Note that $N_k = 1$ with probability $2/(k+1)$ and 0 otherwise. Thus,

$$\mathbb{V}\{X_{k+1}\} = d^2 \times \frac{2(k-1)}{(k+1)^2} > 0 \ .$$

C implies A. For two random variables W, Y, we have $\mathbb{V}\{W\} = \mathbb{E}\{\mathbb{V}\{W|Y\}\} + \mathbb{V}\{\mathbb{E}\{W|Y\}\}$. Thus,

$$\mathbb{V}\{X_{n,K}\} \geq \mathbb{E}\{\mathbb{V}\{X_{n,K}|\mathcal{F}_k\}\}$$

where $\mathcal{F}_k$ is defined as follows. Identify in the permutation that defines the tree all nodes u for which $|S(u)| = k$. By construction, each $S(u)$ corresponds to an interval of the original random permutation (of length n). Let $\mathcal{F}_k$ be all elements of the original random permutation except those corresponding to the intervals representing $S(u)$ with $|S(u)| = k$. By the conditional independence of the various $S(u)$'s of size k (none can overlap), and defining $N_k = \sum_u \mathbb{1}_{[|S(u)|=k]}$, we have for n so large that $K \geq k$,

$$\begin{aligned} \mathbb{V}\{X_{n,K}\} &\geq \mathbb{E}\left\{ \sum_{u:|S(u)|=k} \mathbb{V}\{X_{k,K}|\mathcal{F}_k\} \right\} \\ &= \mathbb{E}\left\{N_k \mathbb{V}\{X_{k,K}\}\right\} \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\{N_k\}\mathbb{V}\{X_{k,K}\} \\
&= \mathbb{E}\{N_k\}\mathbb{V}\{X_k\} \\
&\sim \frac{2n\mathbb{V}\{X_k\}}{(k+2)(k+1)} \ .
\end{aligned}
$$

Thus, the lower bound follows if $\mathbb{V}\{X_k\} > 0$ for some k.

A implies B. Just take $K = n$.

B implies D. If g is constant on the positive integers, then $\mathbb{V}\{X_k\} = 0$ for all k. □

For general functions f on the set of all permutations, it is always possble to have $X_n = 0$ (and thus $\mathbb{V}\{X_n\} = 0$) along a subsequence for n. Assume that all values for $f(\sigma)$ are given, with $|\sigma| < n$. Let σ be of size n. Define $f(\sigma)$ such that $X_n = \sum_u f(S(u)) = 0$. One can even construct examples in which f is integer-valued and $f(\sigma) = |\sigma| = n$ while for smaller permutations σ', the values $f(\sigma')$ are quite arbitrary, as long as they are taken from $\{0, 1, 2, \ldots, |\sigma'|\}$. In the latter case, we have $X_n \equiv n$ for that particular n. So, faced with these pesky examples, we will develop the sequel with the condition $\mathbb{V}\{X_n\} = \Omega(n)$ thrown in. The reader should be warned that this condition needs rigorous verification in each application that does not deal with a toll function.

Sums of functions of sizes of subtrees

In this section, we consider two types of random variables,

$$X_n = \sum_u g(|S(u)|)$$

and

$$X_{n,K} = \sum_{u:|S(u)| \le K} g(|S(u)|) \ ,$$

where $K = K(n) \le n$ is a sequence of positive numbers. Define $G(n) = \max_{1 \le i \le n} |g(i)|$.

Lemma 6. *Assume that g is not constant on $\{1, 2, \ldots\}$. If $K \to \infty$, and $G(K) \log^2 K = o(n^{1/4})$, then*

$$\frac{X_{n,K} - \mathbb{E}\{X_{n,K}\}}{\sqrt{\mathbb{V}\{X_{n,K}\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

proof of lemma 6. We apply Lemma 2 with basic collection of random variables $Y_{i,k} g(k)$, $(i, k) \in V_n$, where V_n is the collection $\{(i, k) : 1 \le i \le n, 1 \le k \le \min(K, n - i + 1)\}$. Let E_n be the edges in the dependency graph L_n defined by connecting (i, k) to (j, ℓ) if the respective intervals are overlapping without being properly nested, or if the respective intervals are disjoint with zero or one integers separating them. (Note that the dependency graph is thus considerably smaller than in the proof of Lemma 3.) We note that

$$\sum_{(i,k)\in V_n} \mathbb{E}\{Y_{i,k}|g(k)|\} \leq (\nu + o(1))n$$

by computations not unlike those for the mean done earlier, where

$$\nu = \sum_{n=1}^{\infty} \frac{2|g(n)|}{(n+2)(n+1)} \ .$$

To apply Lemma 2, we note that we may thus take $M_n = O(n)$. We also note that $\sigma_n^2 = \Omega(n)$, by Lemma 5, since $K \to \infty$ and g is not constant on $\{1, 2, \ldots\}$. It suffices thus to show that $Q_n = o(n^{1/4})$. Note that conditioning on $Y_{i,k}g(k)$ is equivalent to conditioning on $Y_{i,k}$. Thus, we may bound Q_n by

$$Q_n \leq G(K) \sup_{(i,k),(j,\ell)\in V_n} \sum_{(p,r)\in N((i,k),(j,\ell))} \mathbb{E}\{Y_{p,r}|Y_{i,k}, Y_{j,\ell}\} \ .$$

We show that sum above is uniformly bounded over all choices of $(i,k),(j,\ell)$ by $O(\log^2 K)$.

Consider the set $S = \{0, 1, \ldots, n, n+1\}$ and mark $0, n+1, i-1, i+k, j-1, j+\ell$ (where duplications may occur). The last four marked points are neighbors of the intervals represented by (i,k) and (j,ℓ). Mark also all integers in S that are neighbors of these marked numbers. The total number of marked places does not exceed $3 \times 4 + 2 \times 2 = 16$. The set S, when traversed from small to large, can be described by consecutive intervals of marked and unmarked integers. The number of unmarked integer intervals is at most five. We call these intervals $H_1, \ldots, H_5$, from left to right, with some of these possibly empty. Set $H = \cup_i H_i$. Define $H^c = S - H$. Consider $Y_{p,r}$ for $r \leq K$ fixed. Let $s = p + r - 1$ be the endpoint of the interval on which $Y_{p,r}$ sits. Note that $Y_{p,r}$ depends upon $\{U_i\}_{p-1\leq i\leq s+1}$. We note four situations:

A. If $p, s \in H_i$ for a given i, then $Y_{p,r}$ is clearly independent of $Y_{i,k}, Y_{j,\ell}$. In fact, then, $(p,r) \notin N((i,k),(j,\ell))$.

B. If $p, s \in H^c$, then we bound $\mathbb{E}\{Y_{p,r}|Y_{i,k}, Y_{j,\ell}\}$ by one.

C. If p or s is in H^c and the other endpoint is in H_i, then we bound as follows:

$$\mathbb{E}\{Y_{p,r}|Y_{i,k}, Y_{j,\ell}\} \leq \frac{1}{1 + |H_i \cap \{p, \ldots, s\}|}$$

because we can only be sure about the i.i.d. nature of the U_i's in $H_i \cap \{p, \ldots, s\}$ together with the two immediate neighbors of this set.

D. If $p \in H_i$, $s \in H_j$, $i < j$, then we argue as in case C twice, and obtain the following bound:

$$\mathbb{E}\{Y_{p,r}|Y_{i,k}, Y_{j,\ell}\} \leq \frac{1}{1 + |H_i \cap \{p, \ldots, s\}|} \times \frac{1}{1 + |H_j \cap \{p, \ldots, s\}|} \ .$$

The above considerations permit us to obtain a bound for Q_n by summing over all $(p,r) \in N((i,k),(j,\ell))$. The sum for all cases (A) is zero. The sum for case (B) is at most $16^2 = 256$. The sum over all (p,r) as in (C) is at most

$$2 \times 16 \times 5 \times \sum_{r=1}^{K} \frac{1}{r+1} \le 160 \log(K+1) \ .$$

Finally, the sum over all (p,r) described by (D) is at most

$$\binom{5}{2} \left(\sum_{r=1}^{K} \frac{1}{r+1} \right)^2 \le 10 \log^2(K+1) \ .$$

The grand total is $O(\log^2 K)$, as required. This concludes the proof of Lemma 6. $\square$

Corollary 1. As $K \le n$, we deduce that for $G(n) = o(n^{1/4}/\log^2 n)$,

$$\frac{X_n - \mathbb{E}\{X_n\}}{\sqrt{\mathbb{V}\{X_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

This result will be slightly improved in Theorem 6 below.

Corollary 2. A sufficient condition for Lemma 6 is $K = O(n^a)$ for some $0 \le a \le 1$, and $G(n) = o(n^{1/4a}/\log^2 n)$. Other sufficient conditions include either $K = O(1)$, or $K = O(\log n)$, $G(n) = O(\exp(\epsilon n))$ for all $\epsilon > 0$.

Theorem 6. *Assume that g is not constant on $\{1, 2, \ldots\}$. If $G(n) = o(n^{1/3}/\log^2 n)$, then*

$$\frac{X_n - \mathbb{E}\{X_n\}}{\sqrt{\mathbb{V}\{X_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

Proof. Theorem 6 follows directly from Lemma 6 if we can prove the following facts: with $K = \lfloor n^{3/4} \rfloor$,

A. For all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\{X_n - X_{n,K} \geq \epsilon\sqrt{n}\} = 0$. A sufficient condition is $\mathbb{E}\{X_n - X_{n,K}\} = o(\sqrt{n})$.

B. $\liminf_{n\to\infty} \mathbb{V}\{X_n\}/n > 0$.

C. $\mathbb{V}\{X_{n,K}\} \sim \mathbb{V}\{X_n\}$.

For part A, note the following:

$$\begin{aligned}
|\mathbb{E}\{X_n - X_{n,K}\}| &= \left| \sum_{(i,k):1\leq i\leq n, K<k\leq n-i+1} \mathbb{E}\{Y_{i,k}\}g(k) \right| \\
&\leq G(n) + 2 \sum_{K<k\leq n-1} \mathbb{E}\{Y_{1,k}\}G(k) + \sum_{(i,k):2\leq i\leq n, K<k\leq n-i} \mathbb{E}\{Y_{i,k}\}G(k) \\
&\leq G(n) + 2 \sum_{K<k\leq n-1} \frac{G(k)}{k+1} + \sum_{(i,k):2\leq i\leq n, K<k\leq n-i} \frac{2G(k)}{(k+2)(k+1)} \\
&\leq o(n^{1/3}) + \sum_{k=K+1}^{n} \sum_{i=2}^{n-k} \frac{2G(k)}{(k+2)(k+1)} \\
&\leq o(n^{1/3}) + n \sum_{k=K+1}^{n} \frac{2G(k)}{(k+2)(k+1)} \\
&= o(n^{1/2})
\end{aligned}$$

by our choice of K.

Part B is immediate from Lemma 5.

For part C, set

$$X_n = X_{n,K} + W_{n,K}$$

and note that

$$\mathbb{V}\{X_n\} = \mathbb{V}\{X_{n,K}\} + \mathbb{V}\{W_{n,K}\} + 2\mathbb{E}\{(X_{n,K} - \mathbb{E}\{X_{n,K}\})(W_{n,K} - \mathbb{E}\{W_{n,K}\})\} \ .$$

We have from Lemmas 5 and 1, $\mathbb{V}\{X_{n,K}\} = \Theta(n)$. We will show that $\mathbb{V}\{W_{n,K}\} = o(n)$. By the Cauchy-Schwarz inequality,

$$\mathbb{E}\{(X_{n,K} - \mathbb{E}\{X_{n,K}\})(W_{n,K} - \mathbb{E}\{W_{n,K}\})\} = o\left(\sqrt{n}\sqrt{\mathbb{V}\{X_{n,K}\}}\right)$$

so that

$$\frac{\mathbb{V}\{X_n\}}{\mathbb{V}\{X_{n,K}\}} = 1 + o(1) + o\left(\frac{\sqrt{n}}{\sqrt{\mathbb{V}\{X_{n,K}\}}}\right) = 1 + o(1) \ .$$

We now show $\mathbb{V}\{W_{n,K}\} = o(n)$. Let $V_n = \{(i,k) : 1 \le i \le n, 1 \le k \le \min(K, n-i+1)\}$ be the vertex set and let E_n be the edge set for the dependency graph for the random variables $Y_{i,k}g(k)$. That is, two pairs are connected by an edge if their intervals are not properly nested and are either overlapping or disjoint with at most one integer between them. Then:

$$\mathbb{V}\{W_{n,K}\} = \sum_{(i,k)\in V_n} \mathbb{V}\{Y_{i,k}g(k)\} + \sum_{((i,k),(j,\ell))\in E_n} g(k)g(\ell)\left(\mathbb{E}\{Y_{i,k}Y_{j,\ell}\} - \mathbb{E}\{Y_{i,k}\}\mathbb{E}\{Y_{j,\ell}\}\right) .$$

The first sum on the right hand side is bounded by

$$\sum_{(i,k)\in V_n} g^2(k)\mathbb{E}\{Y_{i,k}\} = o(n^{2/3}) + n\sum_{k=K}^{\infty} g^2(k)/k^2 = o(n^{2/3}) + n \times o(K^{-1/3}) = o(n^{3/4}) .$$

Consider a fixed edge $((i,k),(j,\ell)) \in E_n$. Note that if the intervals for (i,k) and (j,ℓ) properly overlap without being nested, then $Y_{i,k}Y_{j,\ell} = 0$. The same is true if they are directly adjacent. So, if $((i,k),(j,\ell)) \in E_n$, the product $Y_{i,k}Y_{j,\ell}$ is nonzero only if they are nested and have one coinciding endpoint, or if the intervals are separated by precisely one integer. Thus, for the latter intervals, with, say, $1 < i \le i+k-1 = j-2 < j \le j+\ell-1 < n$,

$$\mathbb{E}\{Y_{i,k}Y_{j,\ell}\} \le \frac{2}{(\ell+2)(\ell+1)} \times \frac{1}{k+1} .$$

For the intervals aligned at i, with, say, $1 < i = j \le i+k-1 < j+\ell-1 < n$, we have

$$\mathbb{E}\{Y_{i,k}Y_{j,\ell}\} \le \frac{2}{(\ell+2)(\ell+1)} \times \frac{1}{k+1} .$$

The last two bounds are also valid with ℓ and k interchanged. Thus,

$$\begin{aligned}
&\sum_{((i,k),(j,\ell))\in E_n} g(k)g(\ell)\mathbb{E}\{Y_{i,k}Y_{j,\ell}\} \\
&\le 4G(n)\sum_{(j,\ell)\in V_n} G(\ell)\mathbb{E}\{Y_{j,\ell}\} + \sum_{((i,k),(j,\ell))\in E_n} \frac{2G(k)G(\ell)}{\ell^2 k} \\
&\le o(n^{1/3})\sum_{(j,\ell)\in V_n} G(\ell)\mathbb{E}\{Y_{j,\ell}\} + \sum_{((i,k),(i,\ell))\in E_n} \frac{2G(k)G(\ell)}{\ell^2 k} \\
&\quad + \sum_{((i,k),(i+k+1,\ell))\in E_n} \frac{2G(k)G(\ell)}{\ell^2 k} + \sum_{((j+\ell+1,k),(j,\ell))\in E_n} \frac{2G(k)G(\ell)}{\ell^2 k} \\
&= I + II + III + IV .
\end{aligned}$$

First of all,

$$\begin{aligned}
I &\le o(n^{1/3})n\sum_{\ell=K}^{n} \frac{2G(\ell)}{(\ell+2)(\ell+1)} + o(n^{1/3})\sum_{\ell=K}^{n} \frac{G(\ell)}{\ell+1} \\
&\le \frac{o(n^{4/3}G(n))}{K+1} + o(n^{2/3}) \\
&= o(n^{11/12}).
\end{aligned}$$

Now, $\sum_{k\le\ell} G(k)/k = o(\ell^{1/3})$. Thus, using a symmetry argument,

$$II = 2 \sum_{((i,k),(i,\ell))\in E_n, k<\ell} \frac{2G(k)G(\ell)}{\ell^2 k} \le O(n) \sum_{\ell=K}^{n} \frac{o(\ell^{1/3})G(\ell)}{\ell^2} = O(n) \times o(K^{-1/3}) = o(n^{3/4}) \ .$$

Also,

$$III \le O(n) \sum_{k=K}^{n} G(k)/k \times \sum_{\ell=K}^{n} G(\ell)/\ell^2 = O(n)o(n^{1/3})o(K^{-2/3}) = o(n^{5/6}) \ .$$

Similarly, $IV = o(n^{5/6})$. We conclude that $\mathbb{V}\{W_{n,K}\} = o(n^{11/12})$, which is more than was needed. This concludes the proof of Theorem 6. $\square$

Remark. Using the contraction method and the method of moments, Hwang and Neininger (2001) showed that the central limit result of Theorem 6 holds with $G(n) = O(n^a)$, $a \le 1/2$. Our result is weaker, but requires fewer analytic computations. $\square$

Bibliographic remarks.

Central limit theorems for slightly dependent random variables have been obtained by Brown (1971), Dvoretzky (1972), McLeish (1974), Ibragimov (1975), Chen (1978), Hall and Heyde (1980) and Bradley (1981), to name just a few. Stein's method (our Lemma 2, essentially) is one of the central limit theorems that is better equipped to deal with cases of considerable dependence.

Stein's method offers short and intuitive proofs, but other methods may offer attractive alternatives. Hwang and Neininger (2001) are tackling the analysis of random variables of our type by the moment and contraction methods. The ranges of application of the results are not nested—in some situations, Stein's method is more useful, while in others the moment and contraction methods are preferable.

The limit law for L_n, the number of leaves in a random binary search tree, was obtained by Devroye (1991) (see also Mahmoud, 1986):

$$\frac{L_n - \mathbb{E}\{L_n\}}{\sqrt{\mathbb{V}\{L_n\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \ .$$

Equivalently, $(L_n - n/3)/\sqrt{n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2/45)$. That paper deals with general sums $\sum_i f(\sigma(i,k))$ for k fixed and finite, and without regarding the fact that permutations correspond to subtrees of the random binary search tree. As L_n corresponds in our setting to the toll function $\mathbb{1}_{[|\sigma|=1]}$, the above limit law follows easily from Lemma 4, Theorem 5, or Theorem 6. If W_n is the number of nodes with just a right subtree (which occurs at the i-th node if and only if $U_i < U_{i+1}$), then it is easy to see that $f(\sigma(i,k)) = \mathbb{1}_{[U_i < U_{i+1}]}$ for all values of k, and this is clearly not covered by Lemma 4. A relatively easy extension would handle it, but we will not be concerned with that here. The result $(W_n - n/2)/\sqrt{n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/12)$ (Devroye, 1991) is thus not an immediate corollary of the present results.

Aldous (1991) showed that the number $V_{k,n}$ of subtrees of size precisely k in a random binary search tree is in probability asymptotic to $2/(k+2)(k+1)$. Devroye showed that $(V_{k,n} - 2n/(k+2)(k+1))/\sqrt{n}$ tends in law to a normal $(0, c_k)$ random variable where c_k is explicitly defined. The latter result follows also from the present paper if we take as toll function $f(\sigma) = \mathbb{1}_{[|\sigma|=k]}$.

Recently, there has been some interest in the logarithmic toll function $f(\sigma) = \log|\sigma|$ (Grabner and Prodinger, 2001) and the harmonic toll function $f(\sigma) = \sum_{i=1}^{|\sigma|} 1/i$ (Panholzer and Prodinger, 2001). The authors in these papers are mainly concerned with precise first and second moment asymptotics. Fill (1996) obtained the central limit theorem for the case $f(\sigma) = \log|\sigma|$. Clearly, these examples fall entirely within the conditions of Lemma 4 or Theorem 6, with some room to spare.

Flajolet, Gourdon and Martinez (1997) obtained a normal limit law for the number of subtrees in a random binary search tree with fixed finite tree pattern. Clearly, this is a case in which (the indicator function) $f(\sigma)$ depends on σ in an intricate way, but $f = 0$ unless $|\sigma|$ equals the size of the tree pattern. The situation is covered by the law of large numbers of Theorem 2 and the central limit result of Theorem 5. Theorem 5 even allows tree patterns that change with n.

Acknowledgment

We appreciate the comments and improvements suggested to us by Hsien-Kuei Hwang, Ralph Neininger, and Ebrahim Mal-Alla.

References

D. Aldous, "Asymptotic fringe distributions for general families of random trees," *The Annals of Applied Probability*, vol. 1, pp. 228–266, 1991.

R. C. Bradley, "Central limit theorems under weak dependence," *Journal of Multivariate Analysis*, vol. 11, pp. 1–16, 1981.

B. M. Brown, "Martingale central limit theorems," *Annals of Mathematical Statistics*, vol. 42, pp. 59–66, 1971.

L. H. Y. Chen, “Two central limit problems for dependent random variables,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 43, pp. 223–243, 1978.

L. Devroye, “A note on the height of binary search trees,” *Journal of the ACM*, vol. 33, pp. 489–498, 1986.

L. Devroye, “Branching processes in the analysis of the heights of trees,” *Acta Informatica*, vol. 24, pp. 277–298, 1987.

L. Devroye, “Applications of the theory of records in the study of random trees,” *Acta Informatica*, vol. 26, pp. 123–130, 1988.

L. Devroye, “Limit laws for local counters in random binary search trees,” *Random Structures and Algorithms*, vol. 2, pp. 303–316, 1991.

A. Dvoretzky, “Central limit theorems for dependent random variables,” in: *Proceedings of the Sizth Berkeley symposium on Mathematical statistics and Probability*, pp. 513–535, University of California Press, 1972.

J. A. Fill, “On the distribution of binary search trees under the random permutation model,” *Random Structures and Algorithms*, vol. 8(1) , pp. 1–25, 1996.

P. Flajolet, X. Gourdon, and C. Martinez, “Patterns in random binary search trees,” *Random Structures and Algorithms*, vol. 11(3), pp. 223–244, 1997.

G. H. Gonnet, *Handbook of Algorithms and Data Structures*, Addison-Wesley, Reading, Mass., 1984.

P. Grabner and H. Prodinger, “Sorting algorithms for broadcast communications: mathematical analysis,” *Theoretical Computer Science*, 2001. To appear.

P. Hall and C. C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press, New York, 1980.

W. Hoeffding and H. Robbins, “The central limit theorem for dependent random variables,” *Annals of Mathematical Statistics*, vol. 20, pp. 773–780, 1949.

H.-K. Hwang and R. Neininger, “Phase change of limit laws in quicksort recurrence under varying toll functions,” Technical Report, McGill University, Montreal, 2001.

I. A. Ibragimov, “A note on the central limit theorem for dependent random variables,” *Theory of Probability and its Applications*, vol. 20, pp. 135–141, 1975.

S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, John Wiley, New York, 2000.

D. E. Knuth and A. Schonhage, "The expected linearity of a simple equivalence algorithm," *Theoretical Computer Science*, vol. 6, pp. 281–315, 1978.

W. C. Lynch, "More combinatorial problems on certain trees," *Computer Journal*, vol. 7, pp. 299–302, 1965.

W. C. Lynch, "More combinatorial problems on certain trees," *Computer Journal*, vol. 7, pp. 299–302, 1965.

H. Mahmoud and B. Pittel, "On the most probable shape of a search tree grown from a random permutation," *SIAM Journal on Algebraic and Discrete Methods*, vol. 5, pp. 69–81, 1984.

H. M. Mahmoud, "The expected distribution of degrees in random binary search trees," *The Computer Journal*, vol. 29, pp. 36–37, 1986.

H. M. Mahmoud, *Evolution of Random Search Trees*, John Wiley, New York, 1992.

D. L. McLeish, "Dependent central limit theorems and invariance principles," *Annals of Probability*, vol. 2, pp. 620–628, 1974.

A. Panholzer and H. Prodinger, "Binary search tree recursions with harmonic toll functions," *Journal of Computational and Applied Mathematics*, 2001. To appear.

B. Pittel, "On growing random binary trees," *Journal of Mathematical Analysis and Applications*, vol. 103, pp. 461–480, 1984.

J. M. Robson, "The height of binary search trees," *The Australian Computer Journal*, vol. 11, pp. 151–153, 1979.

R. Sedgewick, "Mathematical analysis of combinatorial algorithms," in: *Probability Theory and Computer Science*, edited by G. Louchard and G. Latouche, pp. 123–205, Academic Press, London, 1983.

C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 583–602, 1972.

J. S. Vitter and P. Flajolet, "Average-case analysis of algorithms and data structures," in: *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity*, edited by J. van Leeuwen, pp. 431–524, MIT Press, Amsterdam, 1990.