Phase change of limit laws in the quicksort recurrence under varying toll functions

Hsien-Kuei Hwang[1]
Institute of Statistical Science
Academia Sinica
Taipei 115
Taiwan

Ralph Neininger[2]
School of Computer Science
McGill University
3480 University Street
Montreal, H3A 2K6
Canada

April 12, 2002

Abstract

We characterize all limit laws of the quicksort type random variables defined recursively by $X_n \stackrel{d}{=} X_{I_n} + X^*_{n-1-I_n} + T_n$ when the "toll function" T_n varies and satisfies general conditions, where (X_n), (X^*_n), (I_n, T_n) are independent, $X_n \stackrel{d}{=} X^*_n$, and I_n is uniformly distributed over $\{0, \ldots, n-1\}$. When the "toll function" T_n (cost needed to partition the original problem into smaller subproblems) is small (roughly $\limsup_{n\to\infty} \log E(T_n)/\log n \le 1/2$), X_n is asymptotically normally distributed; non-normal limit laws emerge when T_n becomes larger. We give many new examples ranging from the number of exchanges in quicksort to sorting on broadcast communication model, from an in-situ permutation algorithm to tree traversal algorithms, etc.

AMS subject classifications. Primary: 68W40 68Q25; secondary: 60F05 11B37

Key words. Quicksort, binary search trees, analysis of algorithms, limit distribution, method of moments, contraction method

Abbreviated title: Limit laws of quicksort recurrence

1 Quicksort recurrence

Quicksort, invented by Hoare [36], is one of the most widely used general-purpose sorting algorithms and was selected to be among the top ten most influential algorithms in Science and Engineering in the 20th century; see JaJa [40]. For more information on practical implementation and recent development of quicksort, see for example [5, 48, 74]. Assume that the input comes from a sequence of independent and identically distributed random variables with a common continuous distribution, the cost measures, say X_n, on quicksort can generally be described by $X_0 = 0$ and for $n \ge 1$

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-1-I_n} + T_n, \qquad (1)$$

where $(X_n), (X^*_n), (T_n, I_n)$ are independent, $X_n \stackrel{d}{=} X^*_n$, and I_n is uniformly distributed over $\{0, \ldots, n-1\}$. Here the symbol "$\stackrel{d}{=}$" denotes equivalence in distribution and T_n is either a deterministic function

[1]The work of this author was done while he was visiting School of Computer Science, McGill University. He thanks the School for hospitality and support.

[2]Partially supported by the Deutsche Forschungsgemeinschaft.

of n or a random variable depending on I_n or not. Throughout this paper, we call T_n the *toll function*. Note that this description implicitly assumes that the randomness is preserved for each subfile after partitioning, a property enjoyed by many partitioning schemes but easily violated if carelessly implemented; see Sedgewick [70] for a detailed discussion. Our aim in this paper is to develop a distribution theory for X_n based on the stochastic behavior of T_n.

The motivation of such a study is multifold. First, the model is simple yet prototypical of many sophisticated divide-and-conquer schemes. Viewing this recurrence from an equally important binary search tree perspective, a large number of extensions and variants (see Devroye [18] and Gonnet and Baeza-Yates [30]) can be studied. Second, the inherent phase change of the limit laws from normal to non-normal is a new, interesting phenomenon, which should also occur in many other structures. Third, how sensitive is the limit law of the cost with respect to the toll function? For such a simple structure, a certain "robustness" is expected. Also the extent to which the normal law persists is helpful in giving a deeper understanding of the associated algorithms; roughly, the variance is increasing as the toll function grows, and the algorithms become less useful in practice if the variance is too large. Fourth, a complete characterization of the limit law under varying toll functions is still lacking in the literature. Fifth, the diverse examples we collected were the catalysts that stimulated our study.

The most studied special case is $T_n = n + O(1)$, which corresponds to the number of comparisons used by quicksort to sort a random input, or, equivalently, to the total path length of a random binary search tree. It is known that

$$\frac{X_n - E(X_n)}{n} \xrightarrow{d} Y,$$

where " $\xrightarrow{d}$ " denotes convergence in distribution. Here Y satisfies

$$Y \stackrel{d}{=} UY + (1-U)Y^* + 2U\log U + 2(1-U)\log(1-U) + 1, \tag{2}$$

where $Y \stackrel{d}{=} Y^*$, U is a uniform random variable over the unit interval, and Y, Y^*, and U are independent; see Rösler [65], Régnier [63], and Fill and Janson [25].

Other known cases leading to a normal limit law are

- the number of leaves in a random binary search tree for which $T_n = \delta_{n1}$, the Kronecker symbol; see Devroye [17, 19], Flajolet et al. [27];
- the log-product of subtree sizes for which $T_n = \log n$; see Fill [24];
- the number of occurrences of any fixed pattern with T_n equal to the probability of the pattern when n is equal to the size of the pattern; see Flajolet et al. [27];
- the number of occurrence of subtrees of a given fixed size; see Aldous [1], Devroye [17, 19];
- the number of nodes whose subtree sizes are larger than a given page size $b \geq 1$; see Flajolet et al. [27].

The case when $T_n = n^\alpha$, where $\alpha > 1$, was studied by Neininger [52]; this case leads again to non-normal limit laws.

The rough picture reflected by these sporadic examples is that if the toll function is small such as $\log n$ or $O(1)$ then the limit law of the total cost is normal, and that for large toll function such as n it is non-normal. But when does the limit law of the total cost fail to be normal? We show, under general conditions, that $\sqrt{n}$ is roughly the separating line between normal and non-normal limit laws; this is intuitively in accordance with the classical law of errors. For, from a structural point of view, if the toll function is small, then the contribution from each subproblem is not dominating, so that the normal limit law is quite expected (in vivid terms, the situation resembles the democratic system). On the other hand, if the toll function is large, then the main contribution comes from a few subproblems of large size, rendering large variance and thus non-normal law in the limit (totalitarian system?). An even more intuitive guess is that if $\text{Var}(T_n) = o(\text{Var}(X_n))$ then X_n would be asymptotically normally distributed; otherwise, the limit law would be non-normal. This guess, although false in general, is true for the conditions we consider.

These examples will turn out to be special cases of our general results. We will discuss more new examples in Section 6.

We give two different approaches, based, respectively, on the contraction method (see Rösler and Rüschendorf [68]) and the method of moments, to prove the different limit laws for several reasons. First, we propose the two approaches in a consistent and synthetic way so that they are likely to be applied to other algorithmic problems. Indeed, almost all asymptotic properties of the moments are encapsulated into an "asymptotic transfer" lemma, which relates the asymptotic behavior of the toll function T_n to that of the total cost X_n. Such a transfer also clarifies the sensitivity of the total cost with respect to the toll function (see also Devroye [19]). Second, each approach has its own advantages and inconveniences; we give them for more methodological interests. Third, both approaches can more or less be classified as "computational," in contrast to the "probabilistic" approach used by Devroye in the companion paper [19].

The contraction method, first introduced by Rösler [65] for the analysis in distribution of the quicksort algorithm (namely, (1) with $T_n = n - 1$), starts from a recursive equation satisfied by the random variable in question. Then one computes the first or the second moments, scales properly, proves that the scaled recurrence stabilizes in the limit and chooses a suitable probability metric so that the stabilized equation defines a map of measures that is a contraction in this metric and has a unique fixed-point in some space of probability measures. The weak convergence of the scaled random variables to this fixed-point then follows from the contraction properties; see Rösler [66], Rachev and Rüschendorf [61], Rösler [67] for more information and Rösler and Rüschendorf [68] for a survey. This approach is especially simple if the limiting map has contraction properties in the minimal L_2 metric. In this case only knowledge on the first moment is required for the application of the method. This property will become clear in the case of "large" toll functions (very roughly $E(T_n) \gg \sqrt{n}$). For "small" toll functions, the limiting equation necessitates the use of a probability metric that is ideal of order larger than two as well as information on the variance. In either case a feature of the contraction method is that the dependence between T_n and I_n can be succinctly handled. For other applications of the contraction method, see [16, 49, 51, 55].

The method of moments, one of the most classical ways of deriving limit distributions, has been widely applied to problems in diverse fields (see for example Billingsley [8, Section 30], Diaconis [20]). It consists in first computing the mean and variance, scaling properly the random variable, computing by induction the higher moments of the scaled random variable, applying Carleman's criterion to justify the unicity of the limit law, and then concluding the convergence in distribution and of all moments (or convergence in L_p for all $p > 0$) by the Frechet-Shohat moment convergence theorem (see Loève [43]). While the method of moments is usually used as the "last weapon" for proving limit laws, it does have some advantages: first, it provides more information than weak convergence; second, it is more transparent, self-contained, and requires less advanced theory. We systematize the use of this method, so that all major task boils down to the asymptotic transfer from the toll function to the total cost. Previously, this method was applied by Hennequin in his Ph. D. Thesis [35, Sec. IV.4] to characterize the limit laws of his generalized quicksort (covering in particular the quicksort with median-of-$(2k + 1)$). His proof is, however, incomplete in that his Abelian lemma [35, p. 79] gives only an estimate inside the unit circle for the generating function in question, so that his application of the singularity analysis (see Flajolet and Odlyzko [28]) is not fully justified. We use a different approach, more elementary in nature, to link the asymptotics of the toll function and that of the total cost. For recent applications of the method of moments to similar problems, see Fill [24], Flajolet et al. [29], Dobrow and Fill [21], Schachinger [69].

A schematic diagram illustrating the two approaches is given in Figure 1.

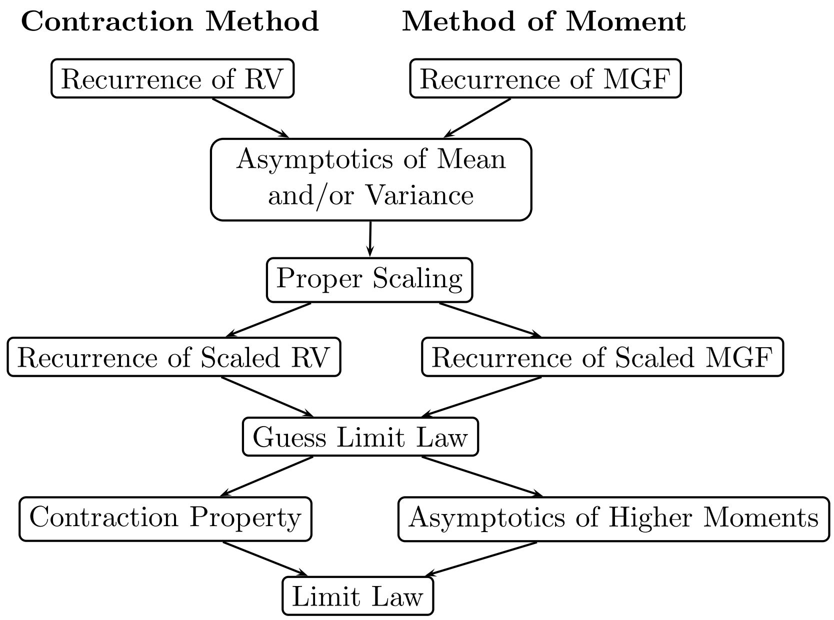

Figure 1: *Main steps used by the contraction method and the method of moments. Here RV denotes "random variable" and MGF denotes "moment generating function".*

Typically, the method of moments requires more assumptions on the moments of the toll function than the contraction method, and the results obtained are stronger. On the other hand, it is also possible to obtain the convergence of all moments by the contraction method based on moment generating functions, see Rösler [65] for details. For another approach to recursive random variables, which we might term "inductive approximation approach", see Pittel [57] and the references therein. See also [13, 55] for an interesting example for which the method of moments applies but the contraction method fails (the space requirement of random m-ary search trees when $m > 26$).

Viewing our results as bridging the transition from normal to "the quicksort law" (2), we can investigate other kind of transitions by looking at different recurrences (or algorithms). A closely related recurrence to (1) is the one-sided quicksort recurrence

$$X_n \stackrel{d}{=} X_{I_n} + T_n, \tag{3}$$

for which we can vary the toll function to bridge the normal law and the Dickman distribution; see [39]. Roughly, our results say that if the toll function is of logarithmic order then the limit law is normal; the limit law is non-normal for larger toll functions; see Section 7 for more precise results and examples.

For the class of problems we study in this paper and many others, an important feature distinguishing normal and non-normal limit laws is the effect of cancellation caused by centering the random variable. Roughly, *the more cancellations of higher moments, the more likely the limit law is normal*. Since our settings cover almost all practical variations of the toll functions, the cancellation effect will be more "visible" in different cases, especially in the method of moments.

We give the main asymptotic transfer results in the next section. Then we prove the phase change of the limit laws in Sections 3 and 4. We first give a more straightforward proof by the method of moments under stronger assumptions; then we apply the contraction method under more general settings. Continuities of the variation of the limit laws are discussed in Section 5. We discuss many examples in Section 6. In particular, the number of exchanges used by quicksort gives an intriguing example of T_n depending on I_n. Section 7 addresses a similar distribution theory for the recurrence (3); this is included because it is closely related.

Notations. Throughout this paper, X_n and T_n are related by (1). We use consistently the following notations: $x_n := E(X_n)$, $t_n := E(T_n)$, $P_n(y) := E(e^{X_n y})$, $Q_n(y) := E(e^{T_n y})$, $H_n := \sum_{1\le k\le n} 1/k$. The symbols U and $N(0,1)$ always represent a uniform $[0,1]$ and a standard normal random variable, respectively. The symbol n_0 denotes a suitable nonnegative integer whose value may vary from one occurrence to another. All unspecified limits (including O, o, $\sim$) are taken to be $n \to \infty$.

Slowly varying functions. A nonnegative function $L(n)$ defined for $n \ge n_0 \ge 0$ and not identically zero, is called *slowly varying* if for all real $\lambda > 0$

$$L(n) \sim L(\lfloor \lambda n \rfloor) \qquad (n \to \infty).$$

If $n_0 > 0$, we define $L(n) = 0$ for $0 \le n < n_0$. Typical slowly varying functions include any powers of $\log n$ and $\log\log n$, $e^{(\log n)^\alpha (\log\log n)^\beta}$, where $0 \le \alpha, \beta < 1$ and $e^{\log n/\log\log n}$.

2 Mean and asymptotic transfers

We develop the main elementary tools in this section that will be used later. While the same results can be obtained via differential equations and suitable analytic tools, we contend ourselves with the elementary approach due to the simplicity of the recurrence. See [14] for more general recurrences of quicksort type.

The mean $x_n := E(X_n)$ satisfies, by (1), $x_0 = 0$ and

$$x_n = \frac{2}{n} \sum_{0\le k<n} x_k + t_n \qquad (n \ge 1), \tag{4}$$

where $t_n := E(T_n)$.

Lemma 1. *Let $\{b_n\}_{n\ge1}$ be a give sequence and define a_n by $a_0 := 0$, and*

$$a_n = \frac{2}{n} \sum_{0\le k<n} a_k + b_n \qquad (n \ge 1). \tag{5}$$

Then for $n \ge 1$

$$a_n = b_n + 2(n+1) \sum_{1\le k<n} \frac{b_k}{(k+1)(k+2)}. \tag{6}$$

Proof. Take the difference $(n+1)a_n - na_n$ and then iterate the resulting recurrence. ▮

From this lemma, we obtain the exact solution for $E(X_n)$

$$E(X_n) = t_n + 2(n+1) \sum_{1\le k<n} \frac{t_k}{(k+1)(k+2)}, \tag{7}$$

for $n \ge 1$; see Devroye [19] for a concrete interpretation of each term on the right-hand side of (7).

The main tool we need is the following lemma linking the asymptotic behavior of the toll function to that of the total cost.

Lemma 2 (Asymptotic transfers). *Assume that a_n satisfies (5). (i) The conditions $b_n = o(n)$ and $\sum_k b_k/k^2 < \infty$ are both necessary and sufficient for*

$$a_n \sim \Upsilon[b]n, \qquad \Upsilon[b] := 2\sum_{k\ge 1} \frac{b_k}{(k+1)(k+2)};$$

(ii) if $b_n \sim nL(n)$, then

$$a_n \sim \begin{cases} \Upsilon[b]n, & \text{if } \sum_{k\ge 1} L(k)/k < \infty; \\ 2n \sum_{k\le n} \frac{L(k)}{k}, & \text{if } \sum_{k\le n} L(k)/k \to \infty; \end{cases}$$

(iii) if $b_n \sim n^\alpha L(n)$, where $\alpha > 1$, then

$$a_n \sim \frac{\alpha+1}{\alpha-1} n^\alpha L(n).$$

Proof. The sufficiency part of (i) follows directly from the exact solution (6). For the necessary part, assume that $a_n \sim cn$ for some constant c. Then by (4),

$$b_n = a_n - \frac{2}{n} \sum_{0\le k<n} a_k = o(n).$$

From this and (6), we deduce that $c = \Upsilon[b] < \infty$.

Part (ii) also results from (6) and the estimate (see Bingham et al. [9, Proposition 1.5.9a])

$$L(n) = o\left(\sum_{k\le n} \frac{L(k)}{k}\right). \tag{8}$$

For part (iii), we have

$$a_n \sim n^\alpha L(n) + 2n \sum_{1\le k\le n} k^{\alpha-2} L(k).$$

But

$$\sum_{1\le k\le n} k^{\alpha-2}L(k) \sim L(n) \sum_{1\le k\le n} k^{\alpha-2} \sim \frac{n^{\alpha-1}}{\alpha-1} L(n);$$

see Proposition 1.5.8 of Bingham et al. [9, p. 26]. ∎

Remarks. 1. If $b_n = o(\sqrt{n})$, then $a_n = \Upsilon[b]n + o(\sqrt{n})$.

2. If $b_n \sim n^\alpha L(n)$, where $\alpha \ge 1/2$, $\alpha \ne 1$, then

$$a_n = \Upsilon[b]n + \frac{\alpha+1}{\alpha-1} n^\alpha L(n)(1+o(1)). \tag{9}$$

3. If we replace the two $\sim$'s for b_n in the lemma by $O(.)$ (or $o(.)$) in cases (ii) and (iii), then the same results hold by replacing $\sim$ for a_n by $O(.)$ (or $o(.)$).

3 Limit laws. I. Method of moments

We study the limit laws of X_n. Briefly, we derive weak convergence and convergence of all moments of X_n (properly normalized) to some Y when estimates for moments of T_n are available. We consider mainly the case when T_n is independent of I_n. The case when T_n depends on I_n requires a straightforward extension of the method; we will discuss briefly the dependence extension for small toll functions; the large toll functions will be discussed via examples in Section 6.

Let $P_n(y) := E(e^{X_n y})$. Then from (1) and independence

$$P_n(y) = \frac{Q_n(y)}{n} \sum_{0\le k<n} P_k(y)P_{n-1-k}(y) \qquad (n \ge 1),$$

with $P_0(y) := 1$, where $Q_n(y) := E(e^{T_n y})$.

Before going further, we need to discard the special case when $T_n = c$ for $n \ge 1$, which yields $X_n = cn$ for $n \ge 1$.

Lemma 3. *Assume that I_n and T_n are independent for $n \ge 1$. The variance of X_n is zero for $n \ge 1$ iff $T_n = c$ for $n \ge 1$ for some constant c.*

Proof. Let $\phi_n(y) := e^{-x_n y}P_n(y)$. Then $\phi_n'(0) = 0$ and

$$\phi_n''(0) = \mathrm{Var}(X_n) = \frac{2}{n}\sum_{0\le k<n} \phi_k''(0) + \psi_n \qquad (n \ge 2),$$

where, defining $\Delta_{n,k} = x_k + x_{n-1-k} - x_n$,

$$\begin{aligned}\psi_n &= Q_n''(0) + \frac{2}{n}t_n \sum_{0\le k<n} \Delta_{nk} + \frac{1}{n}\sum_{0\le k<n} \Delta_{nk}^2 \\ &= E(T_n^2) - (E(T_n))^2 + \frac{1}{n}\sum_{0\le k<n} \Delta_{nk}^2 - \left(\frac{1}{n}\sum_{0\le k<n}\Delta_{nk}\right)^2. \end{aligned} \tag{10}$$

The assertion of the lemma follows from the Cauchy-Schwarz inequality and induction. ∎

Define $Y_\alpha = Y_\alpha(T)$ by

$$Y_\alpha \stackrel{d}{=} \begin{cases} U^\alpha Y_\alpha + (1-U)^\alpha Y_\alpha^* + T, & \text{if } \alpha > 1/2, \alpha \ne 1; \\ UY + (1-U)Y^* + 2U\log U + 2(1-U)\log(1-U) + T, & \text{if } \alpha = 1, \end{cases} \tag{11}$$

where $Y \stackrel{d}{=} Y^*$ and Y, Y^*, T, U are independent. Here T is essentially the limit distribution of T_n/t_n. It will turn out that Y_α is the limit law of X_n, after properly normalized. From this defining equation, it follows that the m-th moment of Y_α, denoted by η_m, satisfies, if it exists, the recurrence $\eta_0 = 1$ and for $m \ge 1$

$$\eta_m = \begin{cases} \displaystyle\sum_{a+b+c=m} \binom{m}{a,b,c} \tau_a \eta_b \eta_c B(b\alpha+1, c\alpha+1), & \text{if } \alpha > 1/2, \alpha \ne 1; \\ \displaystyle\sum_{a+b+c+d=m} \binom{m}{a,b,c,d} \tau_a \eta_b \eta_c \int_0^1 x^{b\alpha}(1-x)^{c\alpha}\Lambda(x)^d \,\mathrm{d}x, & \text{if } \alpha = 1, \end{cases} \tag{12}$$

where $\tau_m = E(T^m)$, $\Lambda(x) := 2x\log x + 2(1-x)\log(1-x)$ and $B(u,v)$ denotes the beta integral:

$$B(u,v) := \int_0^1 x^{u-1}(1-x)^{v-1}\,\mathrm{d}x = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v+1)} \qquad (u,v>0),$$

Γ being the Gamma function. Also the moment generating function $\eta(z) := E(e^{Y_\alpha z})$ satisfies

$$\eta(z) = \begin{cases} \tau(z)\int_0^1 \eta(x^\alpha z)\eta((1-x)^\alpha z)\,\mathrm{d}x, & \text{if } \alpha > 1/2, \alpha \ne 1; \\ \tau(z)\int_0^1 \eta(x^\alpha z)\eta((1-x)^\alpha z)e^{\Lambda(x)z}\,\mathrm{d}x, & \text{if } \alpha = 1, \end{cases}$$

provided that both $\eta(z)$ and $\tau(z) := E(e^{Tz})$ exist.

Lemma 4. *Assume $\alpha > 0$. If the moment generating function of T exists, then $\{\eta_m\}_m$ characterizes uniquely the distribution $\mathcal{L}(Y_\alpha)$.*

Proof. Assume that the series $\sum_m \tau_m z^m/m!$ converges for $|z| \le \delta$ for some $\delta > 0$. We show that $|\eta_m| \le m!K^m$ for a sufficiently large K. By induction using (12), we have

$$
\begin{aligned}
\frac{|\eta_m|}{m!} &\le \frac{m\alpha+1}{m\alpha-1} \sum_{0\le b<m} \int_0^1 x^{b\alpha} \sum_{0\le a\le m-b} \frac{|\tau_a|}{a!} K^{m-a} (1-x)^{(m-a-b)\alpha}\,\mathrm{d}x \\
&\le K^m \left(\sum_{a\ge 0} \frac{|\tau_a|}{a!} K^{-a}\right) \frac{m\alpha+1}{m\alpha-1} \sum_{0\le b<m} \int_0^1 x^{b\alpha}(1-x)^{(m-b)\alpha}\,\mathrm{d}x \\
&\le K^m \left(\sum_{a\ge 0} \frac{|\tau_a|}{a!} K^{-a}\right) \frac{m\alpha+1}{m\alpha-1} \left(\frac{1}{m\alpha+1} + (m-1)\int_0^1 x^{\alpha}(1-x)^{(m-1)\alpha}\,\mathrm{d}x\right) \\
&= K^m \left(\sum_{a\ge 0} \frac{|\tau_a|}{a!} K^{-a}\right) \left[\frac{1}{m\alpha-1} + \frac{(m\alpha+1)(m-1)\Gamma(\alpha+1)\Gamma(m\alpha+1-\alpha)}{(m\alpha-1)(m\alpha+1)\Gamma(m\alpha+1)}\right].
\end{aligned}
$$

Take first K so large that the series $\sum_a |\tau_a|K^{-a}/a!$ converges. Then since the terms in brackets tends to zero as $m \to \infty$, there exists an $m_0 > 0$ such that

$$
\left(\sum_{a\ge 0} \frac{|\tau_a|}{a!} K^{-a}\right) \left[\frac{1}{m\alpha-1} + \frac{(m\alpha+1)(m-1)\Gamma(\alpha+1)\Gamma(m\alpha+1-\alpha)}{(m\alpha-1)(m\alpha+1)\Gamma(m\alpha+1)}\right] < 1,
$$

for $m > m_0$. On the other hand, $|\eta_m| \le m!K^m$ for $m \le m_0$ if K was chosen sufficiently large. We conclude that $|\eta_m| \le m!K^m$ and the required assertion then follows from Carleman's criterion, stating that *the moment sequence $\{\eta_m\}_m$ uniquely characterizes a distribution if $\sum_m \eta_{2m}^{-1/(2m)} = \infty$.* ∎

The condition we impose (that $\tau(z)$ exists) is certainly far from optimal but is sufficient for practical applications.

Define

$$
\Upsilon[t] := 2\sum_{k\ge 1} \frac{t_k}{(k+1)(k+2)}.
$$

Theorem 1 (Large toll functions). *Let (X_n) be given by (1), where T_n is independent of I_n. If*

$$
E(T_n) \sim n^\alpha L(n), \quad \text{and} \quad E\left(\frac{T_n}{t_n}\right)^m \to \tau_m \qquad (m = 2, 3, \dots), \tag{13}
$$

where $\alpha > 1/2$, and $\tau(z) := \sum_m \tau_m z^m/m!$ exists, then

$$
\frac{X_n - \xi_n}{n^\alpha L(n)} \xrightarrow{d} Y_\alpha,
$$

with convergence of all moments, where $Y_\alpha = Y_\alpha(T)$ is defined as above and

$$
\xi_n = \begin{cases} \Upsilon[t]n, & \text{if } 1/2 < \alpha < 1; \\ E(X_n), & \text{if } \alpha = 1; \\ 0, & \text{if } \alpha > 1. \end{cases}
$$

For small toll functions, we distinguish two overlapping cases: (i)

$$
t_n = O(\sqrt{n}/(\log n)^{1/2+\varepsilon}), \quad \text{and} \quad E(T_n^m) = O(t_n^m), \tag{14}
$$

for $m = 2, 3, \ldots$; and (ii)

$$t_n \sim \sqrt{n} L(n), \quad E(T_n^2) \sim \tau_2 t_n^2, \quad \text{and} \quad E(T_n^m) = O(t_n^m), \tag{15}$$

for $m = 3, 4, \ldots$. More general conditions can be studied, but we content ourselves with these two for simplicity of presentation.

In the first case, define

$$s^2(n) := \sigma^2 n, \qquad \sigma^2 := \Upsilon[\psi] = 2 \sum_{k \geq 1} \frac{\psi_k}{(k+1)(k+2)}, \tag{16}$$

where ψ_k is given in (10). In the second case, define $s(n)$ as in (16) if $\sum_{k\geq 1} L^2(k)/k < \infty$, and

$$s^2(n) := \left(\frac{9}{2}\pi - 16 + 2\tau_2 \right) n \sum_{k \leq n} \frac{L^2(k)}{k}, \tag{17}$$

if $\sum_{k\leq n} L^2(k)/k \to \infty$. Note that $\sigma > 0$ by (10) and (16) and the leading constant $\frac{9}{2}\pi - 16 + 2\tau_2$ is positive since $\frac{9}{2}\pi > 14$ and $\tau_2 \geq 1$ by $E(T_n^2) \geq t_n^2$.

Theorem 2 (Small toll functions). *Let (X_n) be given by (1), where T_n is independent of I_n. If T_n satisfies either (14) or (15), then*

$$\frac{X_n - \Upsilon[t]n}{s(n)} \xrightarrow{d} N(0, 1),$$

with mean and variance satisfying $E(X_n) \sim \Upsilon[t]n$ and $\mathrm{Var}(X_n) \sim s^2(n)$. The limit holds with convergence of all moments.

The proof uses the method of moments and the asymptotic transfer lemma.

3.1 Large toll functions

For simplicity of presentation, we split the proof of Theorem 1 into three cases: $1/2 < \alpha < 1$, $\alpha = 1$ and $\alpha > 1$, although we can easily encapsulate them into one.

Case L1. $1/2 < \alpha < 1$**.** In this case, $\xi_n = \Upsilon[t]n$ and by (7)

$$E(X_n) \sim \Upsilon[t]n + \frac{\alpha+1}{\alpha-1} n^\alpha L(n).$$

Shift the mean by $\Upsilon[t]n$ by defining $\Pi_n(y) := P_n(y) e^{-\Upsilon[t]ny}$. Then $\Pi_0(y) = 1$ and

$$\Pi_n(y) = \frac{Q_n(y) e^{-\Upsilon[t]y}}{n} \sum_{0 \leq k < n} \Pi_k(y) \Pi_{n-1-k}(y) \qquad (n \geq 1). \tag{18}$$

Taking m times derivatives with respect to y on both sides and then substituting $y = 0$, we have, by defining $\Pi_{n,m} := \Pi_n^{(m)}(0) = E((X_n - \Upsilon[t]n)^m)$,

$$\Pi_{n,m} = \frac{2}{n} \sum_{0 \leq k < n} \Pi_{k,m} + R_{n,m} \qquad (n \geq 1), \tag{19}$$

with $\Pi_{0,m} = 0$, where

$$R_{n,m} := \sum_{\substack{a+b+c+d=m \\ b,c<m}} \binom{m}{a,b,c,d} Q_n^{(a)}(0) \frac{(-\Upsilon[t])^d}{n} \sum_{0 \leq k < n} \Pi_{k,b} \Pi_{n-1-k,c}.$$

By assumption (13),

$$Q_n^{(m)}(0) = E(T_n^m) \sim \tau_m n^{m\alpha} L^m(n) \qquad (m \geq 1).$$

For convenience, define $\tau_0 = \tau_1 = 1$. We proceed by induction. Assume

$$\Pi_{n,m} \sim g_m n^{m\alpha} L^m(n). \tag{20}$$

This holds true for $m = 1$ by (9) with $g_1 = (\alpha+1)/(\alpha-1)$. By induction and slow variation of $L(n)$, we deduce that for $m \geq 2$

$$\begin{aligned} R_{n,m} &\sim \sum_{\substack{a+b+c=m \\ b,c<m}} \binom{m}{a,b,c} \tau_a g_b g_c n^{a\alpha-1} L^a(n) \sum_{0\leq k<n} k^{b\alpha} L^b(k)(n-1-k)^{c\alpha} L^c(n-1-k) \\ &\sim L^m(n) \sum_{\substack{a+b+c=m \\ b,c<m}} \binom{m}{a,b,c} \tau_a g_b g_c n^{a\alpha-1} \sum_{0\leq k<n} k^{b\alpha}(n-1-k)^{c\alpha} \\ &\sim n^{m\alpha} L^m(n) \sum_{\substack{a+b+c=m \\ b,c<m}} \binom{m}{a,b,c} \tau_a g_b g_c B(b\alpha+1, c\alpha+1). \end{aligned}$$

It follows, by the asymptotic transfer lemma, that

$$\Pi_{n,m} \sim \frac{m\alpha+1}{m\alpha-1} n^{m\alpha} L^m(n) \sum_{\substack{a+b+c=m \\ b,c<m}} \binom{m}{a,b,c} \tau_a g_b g_c B(b\alpha+1, c\alpha+1) \qquad (m \geq 2).$$

Thus if we define g_m recursively by

$$g_m = \frac{m\alpha+1}{m\alpha-1} \sum_{\substack{a+b+c=m \\ b,c<m}} \binom{m}{a,b,c} \tau_a g_b g_c B(b\alpha+1, c\alpha+1) \qquad (m \geq 2),$$

then (20) holds for all $m \geq 1$. Note that $g_m = \eta_m$ for $m \geq 1$; see (12). We conclude, by the Frechet-Shohat moment convergence theorem (see [43]) and Lemma 4, that $\{\eta_m\}$ is the sequence of moments of some distribution function and that $(X_n - \Upsilon[t]n)/(n^\alpha L(n))$ converges in distribution to Y_α.

Case L2. $\alpha = 1$. Define this time $\Pi_n(y) := P_n(y)e^{-x_n y}$, where $x_n = E(X_n)$. Then $\Pi_0(y) = 1$ and

$$\Pi_n(y) = \frac{Q_n(y)}{n} \sum_{0\leq k<n} \Pi_k(y)\Pi_{n-1-k}(y)e^{\Delta_{n,k} y} \qquad (n \geq 1),$$

where $\Delta_{n,k} = x_k + x_{n-1-k} - x_n$. Observe first, by (7), that

$$\Delta_{n,k} = t_k + t_{n-1-k} - t_n - 2(k+1) \sum_{k\leq j<n} \frac{t_j}{(j+1)(j+2)} - 2(n-k) \sum_{n-k<j<n} \frac{t_j}{(j+1)(j+2)};$$

from this and $t_n \sim nL(n)$ we deduce that for $k = \lfloor xn \rfloor$

$$\Delta_{n,k} \sim \Lambda(x) n L(n), \qquad \Lambda(x) := 2x\log x + 2(1-x)\log(1-x), \tag{21}$$

uniformly for $0 \leq x \leq 1$.

Write as above $\Pi_{n,m} = \Pi_{n,m}(0)$. Then $\Pi_{n,m}$ satisfies (19), with

$$R_{n,m} := \sum_{\substack{a+b+c+d=m \\ b,c<m}} \binom{m}{a,b,c,d} \frac{Q_n^{(a)}(0)}{n} \sum_{0\leq k<n} \Pi_{k,b}\Pi_{n-1-k,c}\Delta_{n,k}^d.$$

Note that $\Pi_{n,1} = 0$. We prove by induction that

$$\Pi_{n,m} \sim g_m n^m L^m(n). \tag{22}$$

The case $m = 1$ is true with $g_1 = 0$. For $m \geq 2$, we have, similarly as above,

$$\begin{aligned} R_{n,m} &\sim \sum_{\substack{a+b+c+d=m \\ b,c<m}} \binom{m}{a,b,c,d} \tau_a g_b g_c n^{a-1} L^a(n) \sum_{0\leq k<n} k^b L^b(k)(n-1-k)^c L^c(n-1-k)\Delta_{n,k}^d \\ &\sim n^m L^m(n) \sum_{\substack{a+b+c+d=m \\ b,c<m}} \binom{m}{a,b,c,d} \tau_a g_b g_c \int_0^1 x^b(1-x)^c \Lambda^d(x)\,\mathrm{d}x. \end{aligned}$$

It follows, by Lemma 2, that (22) holds with

$$g_m = \frac{m+1}{m-1} \sum_{\substack{a+b+c+d=m \\ b,c<m}} \binom{m}{a,b,c,d} \tau_a g_b g_c \int_0^1 x^b(1-x)^c \Lambda^d(x)\,\mathrm{d}x \qquad (m \geq 2).$$

Thus convergence in distribution follows as in case **L1**.

Note that $\mathrm{Var}(X_n) \sim g_2 n^2 L^2(n)$, where

$$g_2 = 7 - \frac{2}{3}\pi^2 + 3\mathrm{Var}(T). \tag{23}$$

Case L3. $\alpha > 1$**.** In this case, no centering is needed since $\xi_n = 0$. We apply *mutatis mutandis* the same argument in Case **L1** for $P_n(y)$. The proof is similar and is omitted here.

3.2 Small toll functions

When t_n is small, namely $t_n = O(\sqrt{n}L(n))$, $E(X_n)$ is linear, so we center X_n as in Case **L1** above by defining $\Pi_n(y) := P_n(y)e^{-\Upsilon[t]ny}$. Write again $\Pi_{n,m} := \Pi_n^{(m)}(0)$. Then $\Pi_n(y)$ satisfies (18) and

$$\Pi_{n,1} = \Pi_n'(0) = t_n - 2n\sum_{j\geq n} t_j j^{-2} + O(1).$$

Variance. By (18), the sequence $\Pi_{n,2}$ satisfies (19) with

$$R_{n,2} = Q_n''(0) - t_n^2 + (t_n - \Upsilon[t])(2\Pi_{n,1} - t_n + \Upsilon[t]) + \frac{2}{n}\sum_{0\leq k<n} \Pi_{k,1}\Pi_{n-1-k,1},$$

(the recurrence of $\Pi_{n,1}$ being used to simplify).

Thus, by the asymptotic transfer lemma, if $\sum_k R_{k,2}/k^2 < \infty$, then

$$\mathrm{Var}(X_n) \sim \sigma^2 n, \qquad \sigma^2 := 2\sum_{k\geq 1} \frac{R_{k,2}}{(k+1)(k+2)}. \tag{24}$$

But the condition $\sum_k R_{k,2}/k^2 < \infty$ is not so transparent. We thus consider two simple, overlapping cases.

Case S1. $t_n = O(\sqrt{n}(\log n)^{-1/2-\varepsilon})$**.** In this case, we have

$$\Pi_{n,1} = O(\sqrt{n}(\log n)^{-1/2-\varepsilon}),$$

and by (14)

$$R_{n,2} = O(n(\log n)^{-1-2\varepsilon});$$

thus

$$\mathrm{Var}(X_n) \sim \sigma^2 n.$$

Case S2. $t_n \sim \sqrt{n}L(n)$**.** By (9),

$$\Pi_{n,1} \sim -3\sqrt{n}L(n),$$

from which we deduce, using (15), that

$$R_{n,2} \sim \left(\frac{9}{4}\pi - 8 + \tau_2\right) nL^2(n).$$

Applying Lemma 2 yields

$$\mathrm{Var}(X_n) \sim \begin{cases} \sigma^2 n, & \text{if } \sum_k L^2(k)/k < \infty; \\ \left(\frac{9}{2}\pi - 16 + 2\tau_2\right) n \sum_{k\le n} \frac{L^2(k)}{k}, & \text{if } \sum_k L^2(k)/k = \infty. \end{cases}$$

Note that the definition of σ^2 in (24) can be shown to be identical to (16).

Asymptotic normality. For higher moments, we use again (19) but split $R_{n,m}$ into two parts: $R_{n,m} = R^{(1)}_{n,m} + R^{(2)}_{n,m}$, where

$$\begin{aligned} R^{(1)}_{n,m} &:= \sum_{1\le b<m} \binom{m}{b} \frac{1}{n} \sum_{0\le k<n} \Pi_{k,b}\Pi_{n-1-k,m-b}, \\ R^{(2)}_{n,m} &:= \sum_{\substack{a+b+c+d=m \\ a+d\ge 1 \\ b,c<m}} \binom{m}{a,b,c,d} Q_n^{(a)} \frac{(-\Upsilon[t])^d}{n} \sum_{0\le k<n} \Pi_{k,b}\Pi_{n-1-k,c}. \end{aligned}$$

Case S1. $t_n = O(\sqrt{n}(\log n)^{-1/2-\varepsilon})$**.** Assume that for $m \ge 1$

$$\begin{cases} \Pi_{n,2m} \sim g_m n^m, \\ \Pi_{n,2m-1} = o(n^{m-1/2}). \end{cases}$$

This is true for $m = 1$ with $g_1 = \sigma^2$. By induction, we have

$$R^{(2)}_{n,m} = O\left(n^{m/2}(\log n)^{-1/2-\varepsilon}\right).$$

The main contribution for even moments comes from $R^{(1)}_{n,m}$.

$$\begin{aligned} R^{(1)}_{n,2m} &\sim \sum_{1\le j<m} \binom{2m}{2j} \frac{1}{n} \sum_{0\le k<n} \Pi_{k,2j}\Pi_{n-1-k,2m-2j} \\ &\sim \sum_{1\le j<m} \binom{2m}{2j} g_j g_{m-j} \frac{1}{n} \sum_{0\le k<n} k^j (n-1-k)^{m-j} \\ &\sim \frac{n^m}{m+1} \sum_{1\le j<m} \frac{\binom{2m}{2j}}{\binom{m}{j}} g_j g_{m-j}. \end{aligned}$$

By Lemma 2,

$$\Pi_{n,2m} \sim \frac{n^m}{m-1} \sum_{1\le j<m} \frac{\binom{2m}{2j}}{\binom{m}{j}} g_j g_{m-j} \qquad (m \ge 2).$$

Thus we take g_m so that $g_1 = \sigma^2$ and

$$g_m = \frac{1}{m-1} \sum_{1\le j<m} \frac{\binom{2m}{2j}}{\binom{m}{j}} g_j g_{m-j} \qquad (m \ge 2).$$

The solution is given by

$$g_m = \frac{(2m)!}{2^m m!}\,\sigma^{2m} \qquad (m \geq 1),$$

which equals the $2m$th moment of the normal distribution with mean zero and variance σ^2.

Similarly, for $m \geq 2$,

$$R^{(1)}_{n,2m-1} = o(m^{m-1/2}),$$

and thus by the o-version of Lemma 2

$$\Pi_{n,2m-1} = o(n^{m-1/2}).$$

The asymptotic normality follows.

Case S2. $t_n \sim \sqrt{n}L(n)$**.** In this case, noting that $L^2(n)$ is also slowly varying, we have (see (8))

$$Q''_n(0) \sim \tau_2 t_n^2 \sim \tau_2 n L^2(n) = o\left(n \sum_{k\leq n} \frac{L^2(k)}{k}\right).$$

In particular, if $\sum_k L^2(k)/k < \infty$, then $t_n = o(\sqrt{n})$. The proof then follows the same line of arguments as in Case **S1**.

3.3 T_n depends on I_n

In this case, we have $P_0(y) = 1$ and

$$P_n(y) = \frac{1}{n} \sum_{0\leq k<n} P_k(y)P_{n-1-k}(y)Q_{n,k}(y) \qquad (n \geq 1),$$

where $Q_{n,k}(y)$ is the moment generating function of T_n conditioned on $I_n = k$.

First, Lemma 3 still holds since ψ_n satisfies

$$\psi_n = \frac{1}{n} \sum_{0\leq k<n} \left(Q''_{n,k}(0) - Q'_{n,k}(0)^2\right) + \frac{1}{n} \sum_{0\leq k<n} \left(Q'_{n,k}(0) + \Delta_{n,k}\right)^2, \tag{25}$$

and the same argument applies.

A full extension of the limit laws of X_n to this case requires more assumptions on the asymptotic behavior of $Q_{n,k}(y)$. There is, however, a special case for which the extension is trivial: the Case **S1**, namely, when T_n satisfies (14). The asymptotic normality holds without any additional assumptions. Intuitively, this is the case when each toll summand has only limited contribution to the total cost, thus whether T_n depends on I_n or not does not change the “democratic” nature of the problem, rendering the same law of errors to take effect. The case **S2** needs one more condition (26) and the extension is also straightforward.

Define $s(n)$ as in (16) with ψ_k there replaced by (25) when T_n satisfies (14). In the case when $t_n \sim \sqrt{n}L(n)$, we need in addition to (15) the following estimate

$$E(T_n\sqrt{I_n}L(I_n)) + E(T_n\sqrt{n-1-I_n}L(n-1-I_n)) \sim \tau'_2 n L^2(n). \tag{26}$$

Define $s(n)$ by

$$s^2(n) := \left(2\tau_2 - 6\tau'_2 + \frac{9}{2}\pi\right) n \sum_{k\leq n} \frac{L^2(k)}{k},$$

if $\sum_k L^2(k)/k$ diverges, and $s^2(n) := \sigma^2 n$ otherwise.

Theorem 2′ (Small toll functions—T_n dependent on I_n). *Let (X_n) be given by (1). If T_n satisfies either (14) or the two estimates (15) and (26), then*

$$\frac{X_n - \Upsilon[t]n}{s(n)} \xrightarrow{d} N(0,1),$$

with mean and variance satisfying $E(X_n) \sim \Upsilon[t]n$ and $\mathrm{Var}(X_n) \sim s^2(n)$. In either case, convergence of all moments holds.

The proof of Theorem 2 requires only minor modifications and the $R_{n,2}$ there should be replaced by

$$R_{n,2} = \frac{1}{n}\sum_{0\le k<n} Q''_{n,k}(0) - 2\Upsilon[t]\Pi_{n,1} - \Upsilon[t]^2 + \frac{2}{n}\sum_{0\le k<n} Q'_{n,k}(0)\left(\Pi_{k,1} + \Pi_{n-1-k,1}\right) + \frac{2}{n}\sum_{0\le k<n} \Pi_{k,1}\Pi_{n-1-k,1}.$$

We leave aside the discussions of large toll functions since (i) such cases can be succinctly incorporated in the settings by the contraction method, (ii) Theorem 2′ covers most practical applications, and (iii) we will describe one such example in Section 6.

4 Limit laws. II. Contraction method

We consider the limit laws of X_n using the contraction method in this section. An advantage of this approach is that dependence of T_n on I_n can be easily handled.

4.1 Outline of the method

According to our discussions in the previous section, we first introduce the standardized versions (Y_n) of (X_n) by $Y_n := 0$ for $0 \le n \le n_0$ and

$$Y_n := \frac{X_n - x_n}{s(n)} \qquad (n > n_0),$$

where $s(n) > 0$ is an appropriate scaling to be defined later and $n_0 > 0$ is suitably chosen so that $s(n) > 0$ for all $n > n_0$.

We first sketch the method of proof. The first step is to transform the original recurrence (1) into a modified recurrence for the scaled quantities (Y_n) by defining $Y_n = 0$ for $n \le n_0$ and

$$Y_n \stackrel{d}{=} A_1^{(n)} Y_{I_n} + A_2^{(n)} Y^*_{n-1-I_n} + b_n \qquad (n > n_0), \tag{27}$$

where, for $n > n_0$, $A_1^{(n)} = s(I_n)/s(n)$, $A_2^{(n)} = s(n-1-I_n)/s(n)$ and

$$b_n = \frac{1}{s(n)}\left(x_{I_n} + x_{n-1-I_n} - x_n + T_n\right) =: h_n(T_n, I_n). \tag{28}$$

According to (1), (Y_n), (Y_n^*), and $(A_1^{(n)}, A_2^{(n)}, b_n)$ are independent and $Y_n \stackrel{d}{=} Y_n^*$ for all $n \ge 0$.

If the coefficients $A_1^{(n)}$, $A_2^{(n)}$, and the additive term b_n stabilize as $n \to \infty$, say to A_1, A_2, and b, respectively, and we expect that (Y_n) converges in distribution, then the weak limit Y of (Y_n) should satisfy the limiting equation corresponding to (27):

$$Y \stackrel{d}{=} A_1 Y + A_2 Y^* + b, \tag{29}$$

where Y, Y^*, (A_1, A_2, b) are independent and $Y \stackrel{d}{=} Y^*$.

The contraction method then proceeds by showing that a fixed-point equation as (29) has exactly one solution in a certain space of probability measures and that the scaled random variables under consideration converge in distribution to this fixed-point.

Usually, the existence and uniqueness of a fixed-point of the limiting equation in a subspace of probability distributions is shown by endowing the subspace with a metric and by proving that the limiting equation defines a contraction map on this space. Then the existence of a unique fixed-point is implied by Banach's fixed-point theorem in the case of a complete metric or an appropriate substitute in the incomplete case.

In particular, the minimal L_2-metric ℓ_2 is often used, where ℓ_r-metrics are defined on the spaces $\mathcal{M}_r$ of probability measures on the Borel σ-algebra of $\mathbb{R}$ with finite absolute rth moment by

$$\ell_r(\nu, \varrho) := \inf\{\|X - Y\|_r \,:\, X \stackrel{d}{=} \nu, Y \stackrel{d}{=} \varrho\} \qquad (\nu, \varrho \in \mathcal{M}_r),$$

for $r \geq 1$. We denote by $\mathcal{M}_r(0) \subset \mathcal{M}_r$ the subspace of the centered probability measures in $\mathcal{M}_r$. The metric spaces $(\mathcal{M}_r, \ell_r)$ and $(\mathcal{M}_r(0), \ell_r)$ are complete, and convergence in the ℓ_r-metric is equivalent to weak convergence and convergence of the rth moment. For simplicity, we write $\ell_r(X, Y) := \ell_r(\mathcal{L}(X), \mathcal{L}(Y))$. The infimum in the definition of ℓ_r is attained for all $\nu, \varrho \in \mathcal{M}_r$ and (X, Y) are called optimal couplings of ν, ϱ if $\ell_r(\nu, \varrho) = \|X - Y\|_r$; see Bickel and Freedman [11], Rachev [59], and Rachev and Rüschendorf [62] for more properties of the minimal L_r-metric.

The existence of a unique fixed-point $\mathcal{L}(Y)$ in $\mathcal{M}_2(0)$ for (29) and the convergence in ℓ_2 of (Y_n) given by (27) to Y holds in particular if the following properties are satisfied (see Rösler [67])

(a) $E(b_n) = E(b) = 0, \quad E(b^2) < \infty;$

(b) $\|(A_1^{(n)}, A_2^{(n)}, b_n) - (A_1, A_2, b)\|_2 \to 0;$

(c) $E(A_1^2) + E(A_2^2) < 1;$

(d) For all $n_1 \in \mathbb{N}$, $E[\mathbf{1}_{\{I_n \leq n_1\}}(A_1^{(n)})^2] + E[\mathbf{1}_{\{n-1-I_n \leq n_1\}}(A_2^{(n)})^2] \to 0.$

This is the line we will follow for large toll functions T_n. In the case of small toll functions we will end up with a well-known limiting equation that is not a contraction in ℓ_2 and has the normal distributions as solutions. In this case asymptotic normality will be derived by a change of the metric as used in Rachev and Rüschendorf [61]. The metric used later is ideal of order larger than two, which implies the contraction properties of the limiting equation with respect to this metric on the appropriate space.

4.2 Large toll functions

Assume that

$$E(T_n) \sim n^\alpha L(n) \quad \text{and} \quad \left(\frac{T_n}{E(T_n)}, \frac{I_n}{n}\right) \xrightarrow{L_2} (T, U),$$

where $\alpha > 1/2$, $L(n)$ is slowly varying, and T is square-integrable. In particular, T_n may depend on I_n and T may depend on U. For our applications to quicksort and binary search trees, this U comes up (essentially) as the first partitioning element of quicksort (or the root of the associated binary search tree). Therefore, I_n has, conditioned on $U = u$, the binomial $B(n-1, u)$ distribution and $I_n/n \to U$ holds in L_p for all $p > 0$.

For the scaling factor, we assume at the moment that the variance of X_n admits an expansion of the form

$$\mathrm{Var}(X_n) \sim \sigma^2 \, n^{2\alpha} L^2(n),$$

where $\sigma = \sigma(\alpha, (T, U))$ is a positive constant given later in Corollary 1. This will later turn out to be true (up to degenerate cases). Therefore, we use the scaling $s(n) := n^\alpha L(n)$ and define $Y_n := 0$ for $0 \le n \le n_0$ and

$$Y_n := \frac{X_n - E(X_n)}{n^\alpha L(n)} \qquad (n > n_0);$$

so that for $n > n_0$

$$Y_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\alpha \frac{L(I_n)}{L(n)} Y_{I_n} + \left(\frac{n-1-I_n}{n}\right)^\alpha \frac{L(n-1-I_n)}{L(n)} Y^*_{n-1-I_n} + h_n(\alpha, (T_n, I_n)), \tag{30}$$

where $h_n(\alpha, (T_n, I_n)) := (x_{I_n} + x_{n-1-I_n} - x_n + T_n)/(n^\alpha L(n))$.

Observe that our formal L_2-convergence assumption on I_n/n is equivalent to $I_n/n \to U$ in L_p for all $p \ge 0$. Using this and the estimate

$$\frac{L(I_n)}{L(n)} \xrightarrow{L_2} 1,$$

we obtain

$$\begin{aligned} \left\| \left(\frac{I_n}{n}\right)^\alpha \frac{L(I_n)}{L(n)} - U^\alpha \right\|_2 &\le \left\| \left(\frac{I_n}{n}\right)^\alpha - U^\alpha \right\|_2 + \left\| \left(\frac{I_n}{n}\right)^\alpha \left(\frac{L(I_n)}{L(n)} - 1\right) \right\|_2 \\ &= o(1) + \left\| \frac{L(I_n)}{L(n)} - 1 \right\|_2 \to 0. \end{aligned} \tag{31}$$

Analogously,

$$\left\| \left(\frac{n-1-I_n}{n}\right)^\alpha \frac{L(n-1-I_n)}{L(n)} - (1-U)^\alpha \right\|_2 \to 0.$$

Finally, $h_n(\alpha, (T_n, I_n))$ also stabilizes:

$$h_n(\alpha, (T_n, I_n)) \xrightarrow{L_2} h(\alpha, (T, U)), \tag{32}$$

where for $\alpha > 1/2$

$$h(\alpha, (T, U)) := \begin{cases} \dfrac{\alpha+1}{\alpha-1}\Big(U^\alpha + (1-U)^\alpha - 1\Big) + T, & \text{if} \quad \alpha \ne 1, \\ 2U \log U + 2(1-U)\log(1-U) + T, & \text{if} \quad \alpha = 1. \end{cases} \tag{33}$$

For $\alpha = 1$, (32) is proved by the relation (see (21))

$$\frac{x_{I_n} + x_{n-1-I_n} - x_n}{nL(n)} \xrightarrow{L_2} 2U \log U + 2(1-U)\log(1-U),$$

and our assumption $T_n/t_n \to T$ in L_2. The case $\alpha \ne 1$ is established by using the asymptotic expansions (9) for $1/2 < \alpha < 1$ and Lemma 2 for $\alpha > 1$, respectively:

$$\begin{aligned} &\frac{1}{n^\alpha L(n)}\Big(x_{I_n} + x_{n-1-I_n} - x_n + T_n\Big) \\ &= \frac{1}{n^\alpha L(n)} \frac{\alpha+1}{\alpha-1}\Big(I_n^\alpha L(I_n) + (n-1-I_n)^\alpha L(n-1-I_n) - n^\alpha L(n) + T_n\Big) + o(1) \\ &\to \frac{\alpha+1}{\alpha-1}\Big(U^\alpha + (1-U)^\alpha - 1\Big) + T \quad \text{in } L_2, \end{aligned}$$

where the $o(1)$ depends on the randomness but the convergence is uniform. This establishes the stabilization of the modified recursion (30) to the limiting equation

$$Y \stackrel{d}{=} U^\alpha Y + (1-U)^\alpha Y^* + h(\alpha, (T, U)). \tag{34}$$

Note that this equation coincides for independent T, U with (11) in the case $\alpha = 1$; for $\alpha > 1/2, \alpha \neq 1$, (34) a translated version of (11) in the sense that Y is a fixed-point of (34) if and only if $Y + (\alpha + 1)/(\alpha - 1)$ is a fixed point of (11). This is because in Theorem 1, the random variable is not centered for $\alpha \neq 1$ by the exact mean, so that the mean of Y_α there equals $(\alpha+1)/(\alpha-1)$ while our Y has mean zero.

The limiting equation (34) defines a map $S_{\alpha,(T,U)}$ on $\mathcal{M}_2$:

$$S_{\alpha,(T,U)} : \mathcal{M}_2 \to \mathcal{M}_2, \quad \nu \mapsto \mathcal{L}\Big(U^\alpha Z + (1-U)^\alpha Z^* + h(\alpha, (T,U))\Big), \tag{35}$$

where $Z, Z^*, (T,U)$ are independent, $Z \stackrel{d}{=} Z^* \stackrel{d}{=} \nu$, and $h(\alpha, (T,U))$ is given by (33).

Theorem 3. *Let (X_n) be given by (1). Assume that*

$$E(T_n) \sim n^\alpha L(n) \quad \textit{and} \quad \left(\frac{T_n}{E(T_n)}, \frac{I_n}{n}\right) \xrightarrow{L_2} (T,U),$$

where $\alpha > 1/2$, and that T is square-integrable. Then

$$\ell_2\left(\frac{X_n - E(X_n)}{n^\alpha L(n)}, Y_{\alpha,(T,U)}\right) \to 0,$$

where $\mathcal{L}(Y_{\alpha,(T,U)})$ is the unique fixed-point in $\mathcal{M}_2(0)$ of the map $S_{\alpha,(T,U)}$ defined in (35).

Proof. First we show that the restriction of $S_{\alpha,(T,U)}$ to $\mathcal{M}_2(0)$ is a map into $\mathcal{M}_2(0)$. Let $\nu \in \mathcal{M}_2(0)$. Then $S_{\alpha,(T,U)}(\nu)$ has finite second moment because of independence and the same property of the coefficients. The assumption $T_n/E(T_n) \to T$ in L_2 implies that $E(T) = 1$ and therefore $E(h(\alpha,(T,U))) = 0$ for all $\alpha > 1/2$. This implies $E(S_{\alpha,(T,U)}(\nu)) = 0$, and thus $S_{\alpha,(T,U)}(\nu) \in \mathcal{M}_2(0)$.

By Theorem 3 in Rösler [66] or Lemma 1 in Rösler and Rüschendorf [68] $S_{\alpha,(T,U)}$ is Lipschitz continuous on $(\mathcal{M}_2(0), \ell_2)$ where the Lipschitz constant $\mathrm{lip}(S_{\alpha,(T,U)})$ satisfies

$$\mathrm{lip}(S_{\alpha,(T,U)}) \leq \Big(E(U^{2\alpha}) + E((1-U)^{2\alpha})\Big)^{1/2}.$$

Since $\alpha > 1/2$ we have $\mathrm{lip}(S_{\alpha,(T,U)}) \leq \sqrt{2/(2\alpha+1)} < 1$; thus $S_{\alpha,(T,U)}$ is a contraction on $\mathcal{M}_2(0)$. By Banach's fixed-point theorem $S_{\alpha,(T,U)}$ has a unique fixed-point $\mathcal{L}(Y_{\alpha,(T,U)})$ in $\mathcal{M}_2(0)$.

By (30) and (34) the standardized variables $Y_n = (X_n - E(X_n))/n^\alpha L(n)$ and $Y_{\alpha,(T,U)}$ satisfy, respectively,

$$Y_n \stackrel{d}{=} A_1^{(n)} Y_{I_n} + A_2^{(n)} Y^*_{n-1-I_n} + b_n,$$

and

$$Y_{\alpha,(T,U)} \stackrel{d}{=} A_1 Y_{\alpha,(T,U)} + A_2 Y^*_{\alpha,(T,U)} + b.$$

It remains to check the conditions **(a)**–**(d)**.

First, by taking expectations in (30) and (34), respectively, we obtain $E(b_n) = E(b) = 0$; also $E(b^2) < \infty$ since T is square-integrable. Thus **(a)** is satisfied. Condition **(b)** is established in (31) and (32) and condition **(c)** is the contraction property of $S_{\alpha,(T,U)}$. Finally, condition **(d)** follows from $|s(I_n)/s(n)|, |s(n-1-I_n)/s(n)| < 1$ since

$$\begin{aligned} E(\mathbf{1}_{\{I_n \leq n_1\}}(A_1^{(n)})^2) + E(\mathbf{1}_{\{n-1-I_n \leq n_1\}}(A_2^{(n)})^2) &\leq P(I_n \leq n_1) + P(n-1-I_n \leq n_1) \\ &= \frac{2n_1}{n} \to 0, \end{aligned}$$

for all $n_1 \in \mathbb{N}$. We complete the proof by applying Rösler's theorem [67]. ∎

Note that if $h(\alpha,(T,U)) = 0$, then the limit distribution $\mathcal{L}(Y_{\alpha,(T,U)})$ is degenerate, namely, $Y_{\alpha,(T,U)} = 0$ almost surely. In this case more knowledge on the asymptotics of T_n is necessary and a scaling other than $n^\alpha L(n)$ should be used (our limit law yields merely $\mathrm{Var}(X_n) = o(n^\alpha L(n))$).

Corollary 1. *If $h(\alpha, (T, U)) \neq 0$ (see (33)), then the sequence (X_n) of Theorem 3 satisfies*

$$\mathrm{Var}(X_n) \sim \sigma^2 n^{2\alpha} L^2(n),$$

where $\sigma^2 = \sigma^2(\alpha, (T, U))$ is defined by

$$\sigma^2 = \begin{cases} \dfrac{\alpha(\alpha+1)^2 B(\alpha,\alpha) + 2(\alpha^2 - 2\alpha - 1)}{(2\alpha-1)(\alpha-1)^2} + C, & \text{if } \alpha \neq 1; \\ 7 - \dfrac{2\pi^2}{3} + C, & \text{if } \alpha = 1, \end{cases}$$

with $C = C(\alpha, (T, U))$ given by

$$C = \begin{cases} \dfrac{2\alpha+1}{2\alpha-1}\left(\mathrm{Var}(T) + 2\dfrac{\alpha+1}{\alpha-1} E\left[T(U^\alpha + (1-U)^\alpha)\right] - \dfrac{4}{\alpha-1}\right), & \text{if } \alpha \neq 1; \\ 3\Big(\mathrm{Var}(T) + 4E[T(U\log U + (1-U)\log(1-U))] + 2\Big), & \text{if } \alpha = 1. \end{cases}$$

Proof. By Theorem 3, $\mathrm{Var}(X_n) = \mathrm{Var}(n^\alpha L(n) Y_n) \sim E(Y^2_{\alpha,(T,U)}) n^{2\alpha} L^2(n)$, thus $\sigma^2 = E(Y^2_{\alpha,(T,U)})$. Since $Y_{\alpha,(T,U)}$ solves the equation (34), we deduce, by taking squares and expectations, that

$$E\left(Y^2_{\alpha,(T,U)}\right) = \frac{2\alpha+1}{2\alpha-1} E(h^2(\alpha, (T, U))),$$

which leads to the expressions in the corollary. ∎

If T is independent of U, then $C = (2\alpha+1)\mathrm{Var}(T)/(2\alpha-1)$, which coincides with (23) for $\alpha = 1$. Moreover, $C = 0$ if $T = 1$, which holds in particular if the toll functions (T_n) are all deterministic.

4.3 Small toll functions

In this section we consider small toll functions by the contraction method, assuming again that T_n and I_n may be dependent. Write $s(n)^2 := \mathrm{Var}(X_n)$. As in the analysis by the method of moments, we consider two cases:

$$t_n = O(\sqrt{n}/(\log n)^{1/2+\epsilon}), \quad E(T_n^2) = O(t_n^2), \quad \text{and} \quad E\left(\frac{T_n}{s(n)}\right)^{2+\delta} \to 0, \tag{36}$$

where $0 < \delta < 1$; and

$$\begin{cases} t_n \sim \sqrt{n} L(n), \quad E(T_n^2) \sim \tau_2 n L^2(n), \quad E\left(\dfrac{T_n}{s(n)}\right)^{2+\delta} \to 0, \quad \text{and} \\ E(T_n \sqrt{I_n} L(I_n)) + E(T_n \sqrt{n-1-I_n} L(n-1-I_n)) \sim \tau_2' n L^2(n). \end{cases} \tag{37}$$

In particular, if we assume (14) or (15), then (36) or (37) hold, respectively.

We first look for stabilization in (27) in order to derive a limiting equation. In the case (36), we have (see (16)), $s(n)^2 \sim \sigma^2 n$; thus

$$A_1^{(n)} = \frac{s(I_n)}{s(n)} \to U^{1/2} \quad \text{in } L_{2+\delta}. \tag{38}$$

Similarly,

$$A_2^{(n)} = \frac{s(n-1-I_n)}{s(n)} \to (1-U)^{1/2} \quad \text{in } L_{2+\delta}. \tag{39}$$

For the additive term in (27), we obtain $(x_{I_n} + x_{n-1-I_n} - x_n)/s(n) \to 0$ in $L_{2+\delta}$ by the expansion $E(X_n) = \Upsilon[t]n + o(\sqrt{n})$. This together with $T_n/s(n) \to 0$ gives

$$\frac{x_{I_n} + x_{n-1-I_n} - x_n + T_n}{s(n)} \to 0 \quad \text{in } L_{2+\delta}. \tag{40}$$

The recursion (27) for $Y_n = (X_n - E(X_n))/s(n)$ and $Y_n = 0$ if $s(n) = 0$ now lead to the limiting equation

$$Y \stackrel{d}{=} U^{1/2}Y + (1-U)^{1/2}Y^*. \tag{41}$$

The conditions (38)–(40) are also satisfied in the case (37) using the corresponding expansions for $s(n)$. Briefly, (38) and (39) are proved by $(\sum_{k\le I_n} L^2(k)/k)/(\sum_{k\le n} L^2(k)/k) \to 1$. For (40), if $\sum_k L^2(k)/k < \infty$, then $L(k) \to 0$, implying that $(x_{I_n} + x_{n-1-I_n} - x_n)/s(n) \to 0$ in $L_{2+\delta}$. If $\sum_k L^2(k)/k = \infty$, the same $L_{2+\delta}$ convergence follows from $L^2(n) = o(\sum_{k\le n} L^2(k)/k)$; see (8).

In all cases we obtain the limiting equation (41). Therefore, we cannot follow the line as for large toll functions since (41) has no contraction properties in ℓ_2 and is not a contraction for any ℓ_r-metric. This is well-known and discussed in Rachev and Rüschendorf [61] and Rösler and Rüschendorf [68]. Thus we have to choose a metric that is $(r,+)$-ideal, where $r > 2$, and to refine the work space $\mathcal{M}_2(0)$ in order to obtain contraction properties for equation (41).

The situation here is similar to the size of random tries discussed in Rachev and Rüschendorf [61]. We obtain weak convergence of (Y_n) to a normal distribution by applying similar arguments; see also Rösler and Rüschendorf [68].

Following Rachev and Rüschendorf, define, for $r = m + 1/p$ with $m \in \mathbb{N}$ and $p \in [1,\infty)$,

$$\mathcal{F}_r := \{f \in C^{m+1} \, : \, \|f^{(m+1)}\|_q \le 1\},$$

where $1/p + 1/q = 1$ and $f^{(m+1)}$ denotes the $m+1$st derivative of the function $f : \mathbb{R} \to \mathbb{R}$. Then we will use the metric

$$\mu_r(X,Y) := \sup_{f\in\mathcal{F}_r} |E[f(X) - f(Y)]|,$$

which was introduced and studied in Maejima and Rachev [45]; see also Rachev and Rüschendorf [60].

We briefly state the properties of μ_r, which are used subsequently. The metric μ_r is $(r,+)$-ideal, i.e., $\mu_r(cX,cY) = c^r\mu_r(X,Y)$ for $c > 0$ and $\mu_r(X+Z,Y+Z) \le \mu_r(X,Y)$ if Z is independent of X,Y. An upper estimate for μ_r in Zolotarev's metric ζ_r and corresponding properties for the metric ζ_r (see Zolotarev [75]) imply that $\mu_r(X,Y) < \infty$ if $E(X^j) = E(Y^j)$ for all $j = 1,\dots,m$ and $E(|X|^r), E(|Y|^r) < \infty$. Convergence in μ_r implies convergence in distribution, since a lower estimate in Levy's metric L is valid: $(L(X,Y))^{r+1} \le C(r)\mu_r(X,Y)$ for some constant $C(r) < \infty$. We will also use the fact that convergence in ℓ_r implies convergence in μ_r. This follows from the upper estimate $\mu_r(X,Y) \le C'(r)\kappa_r(X,Y)$ with some constant $C'(r) < \infty$ and the difference pseudomoment κ_r and the fact that κ_r and ℓ_r are topologically equivalent (see Rachev [59, p. 301]).

The following proof of asymptotic normality is based on the approach used in Rachev and Rüschendorf [61] mentioned above. The differences here are that we derive convergence in $\mu_{2+\delta}$ rather than only weak convergence, and that the estimate of the additive term $h_n(T_n, I_n)$ is simplified. These improvements are due to the fact that more information on the moments is known in our case.

Theorem 4. *Let (X_n) be given by (1). If T_n satisfies either (36) or (37), then*

$$\mu_{2+\delta}\left(\frac{X_n - E(X_n)}{\sqrt{\mathrm{Var}(X_n)}}, N(0,1)\right) \to 0.$$

Proof. Let $r := 2+\delta$. The key idea of the proof is to introduce a mixed quantity that combines the structure of the modified recursion with the normal distribution; see [61] and [68, Section 6]. We denote by N, N^* two independent standard normal random variables that are also independent of all other quantities.

Then we define the distributions of our mixtures M_n by $M_n := 0$ for $0 \le n \le n_0$ and for $n > n_0$

$$\begin{aligned} M_n &\stackrel{d}{:=} \frac{s(I_n)}{s(n)} N + \frac{s(n-1-I_n)}{s(n)} N^* + h_n(T_n, I_n) \\ &\stackrel{d}{=} \left[\left(\frac{s(I_n)}{s(n)}\right)^2 + \left(\frac{s(n-1-I_n)}{s(n)}\right)^2 \right]^{1/2} N + h_n(T_n, I_n), \end{aligned} \tag{42}$$

with $h_n(T_n, I_n)$ given in (28). A comparison with (27) shows that $E(M_n) = 0$, $E(M_n^2) = 1$, and $E|M_n|^r < \infty$ for $n \ge n_0$, thus μ_r-distances between Y_n, M_n and $N(0,1)$ are finite. We convent all μ_r-distances for these quantities with indices $\le n_0$ to be zero. We may estimate

$$\mu_r(Y_n, N(0,1)) \le \mu_r(Y_n, M_n) + \mu_r(M_n, N(0,1)).$$

By (38) and (39), the factor between the brackets in (42) converges to 1 in L_r; this together with the L_r-convergence of $h_n(T_n, I_n)$ to 0 yields $\ell_r(M_n, N(0,1)) \to 0$ and, therefore, $\mu_r(M_n, N(0,1)) \to 0$. Here we used the estimates (38)–(40).

Denote by λ_n the joint distribution of (T_n, I_n). By the $(r,+)$-ideality of μ_r, we have

$$\begin{aligned} \mu_r(Y_n, M_n) &= \sup_{f \in \mathcal{F}_r} |E(f(Y_n) - f(M_n))| \\ &= \sup_{f \in \mathcal{F}_r} \left| \int E\left[f\left(\frac{s(k)}{s(n)} Y_k + \frac{s(n-1-k)}{s(n)} Y^*_{n-1-k} + h_n(t,k) \right) \right.\right. \\ &\qquad \left.\left. - f\left(\frac{s(k)}{s(n)} N + \frac{s(n-1-k)}{s(n)} N^* + h_n(t,k) \right) \right] \mathrm{d}\lambda_n(t,k) \right| \\ &\le \int \mu_r\left(\frac{s(k)}{s(n)} Y_k + \frac{s(n-1-k)}{s(n)} Y^*_{n-1-k} + h_n(t,k), \right. \\ &\qquad \left. \frac{s(k)}{s(n)} N + \frac{s(n-1-k)}{s(n)} N^* + h_n(t,k) \right) \mathrm{d}\lambda_n(t,k) \\ &\le \frac{1}{n} \sum_{k=0}^{n-1} \left(\mu_r\left(\frac{s(k)}{s(n)} Y_k, \frac{s(k)}{s(n)} N \right) + \mu_r\left(\frac{s(n-1-k)}{s(n)} Y^*_{n-1-k}, \frac{s(n-1-k)}{s(n)} N^* \right) \right) \\ &\le \frac{2}{n} \sum_{k=0}^{n-1} \left(\frac{s(k)}{s(n)} \right)^r \mu_r(Y_k, N). \end{aligned} \tag{43}$$

Thus, we obtain the reduction inequality

$$\mu_r(Y_n, N(0,1)) \le \frac{2}{n} \sum_{k=0}^{n-1} \left(\frac{s(k)}{s(n)} \right)^r \mu_r(Y_k, N(0,1)) + o(1).$$

By (38) and (39)

$$\frac{2}{n} \sum_{k=0}^{n-1} \left(\frac{s(k)}{s(n)} \right)^r = 2E\left(\frac{s(I_n)}{s(n)} \right)^r \to 2E\left(U^{r/2} \right) = \frac{2}{r/2+1} < 1.$$

From this and the reduction inequality, we deduce by a bootstrapping argument (see Rösler [65, p. 94] or Rachev and Rüschendorf [61, p. 786]) that $\mu_r(Y_n, N(0,1)) \to 0$. [First prove that $\mu_r(Y_n, N(0,1))$ remains bounded; then refine the approximation.] ∎

5 Continuous change of limits

We prove that the limit distributions $\mathcal{L}(Y_{\alpha,(T,U)})$ in Theorem 3 are continuous in the parameters $(\alpha,(T,U))$, where $\alpha > 1/2$. The property still holds as $\alpha \downarrow 1/2$ in the case of deterministic toll functions and in the random case under appropriate assumptions.

Theorem 5. *Let $\alpha \to \beta > 1/2$ and $T = T(\alpha) \to V$ in L_2 for a square-integrable V. Then*

$$\ell_2(Y_{\alpha,(T,U)}, Y_{\beta,(V,U)}) \to 0.$$

Let $\alpha \downarrow 1/2$ and $T = T(\alpha)$ satisfy $\|T\|_{2+\delta} = o(\sigma(\alpha,(T,U)))$ as $\alpha \downarrow 1/2$. If T is independent of U, then

$$\mu_{2+\delta}\left(\frac{Y_{\alpha,(T,U)}}{\sigma(\alpha,(T,U))}, N(0,1)\right) \to 0.$$

The property still holds if T, U are dependent, provided that (i) $h(\alpha,(T,U)) \neq 0$ for h given in (33) and (ii) $\sigma(\alpha,(T,U))$ in Corollary 1 is properly divergent.

Proof. (Sketch) Consider the special case $\beta = 1$. For $\alpha > 1/2$, we have, by definition,

$$\begin{aligned} Y_{\alpha,(T,U)} &\stackrel{d}{=} U^\alpha Y_{\alpha,(T,U)} + (1-U)^\alpha Y^*_{\alpha,(T,U)} + \frac{\alpha+1}{\alpha-1}\Big(U^\alpha + (1-U)^\alpha - 1\Big) + T, \\ Y_{1,(V,U)} &\stackrel{d}{=} UY_{1,(V,U)} + (1-U)Y^*_{1,(V,U)} + 2U\log U + 2(1-U)\log(1-U) + V, \end{aligned}$$

where $(Y_{\alpha,(T,U)}, Y_{1,(V,U)})$, $(Y^*_{\alpha,(T,U)}, Y^*_{1,(V,U)})$, (T,U,V) are independent, and optimal couplings of $\mathcal{L}(Y_{\alpha,(T,U)})$ and $\mathcal{L}(Y_{1,(V,U)})$ are formed by $(Y_{\alpha,(T,U)}, Y_{1,(V,U)})$, $(Y^*_{\alpha,(T,U)}, Y^*_{1,(V,U)})$. To match these two fixed-point equations we use the Taylor expansion

$$x^\alpha = x + (\alpha-1)x\log x + \int_1^\alpha (\alpha - y)\Big(x^{y-1} + x^y(\log x)^2\Big)\,\mathrm{d}y \qquad (x \in (0,1), \alpha > 0). \tag{44}$$

Using the representations of $Y_{\alpha,(T,U)}, Y_{1,(V,U)}$ given in the coupled fixed-point equations in the estimate $\ell_2(Y_{\alpha,(T,U)}, Y_{1,(V,U)}) \le \|Y_{\alpha,(T,U)} - Y_{1,(V,U)}\|_2$, we obtain, after tedious calculations, that

$$\ell_2(Y_{\alpha,(T,U)}, Y_{1,(V,U)}) \ll \max\Big\{|\alpha-1|, \sqrt{|\alpha-1|\|T-V\|_2}, \|T-V\|_2\Big\}, \tag{45}$$

as $\alpha \to 1$ and $T \to V$ in L_2. In particular, we used the expansion

$$B(\alpha,\alpha) = 1 - 2(\alpha-1) + (4 - \pi^2/6)(\alpha-1)^2 + O((\alpha-1)^3) \tag{46}$$

to derive $\sigma(\alpha,(T,U)) \to \sigma(1,(V,U))$ as $\alpha \to 1$ and $T \to V$ in L_2. This implies the assertion for $\beta = 1$. The general case $\beta > 1/2$ can be treated by the same approach and is indeed simpler since the expansions (44) and (46) are not needed.

For the second part we denote $r := 2 + \delta$ and $Z_{\alpha,(T,U)} := Y_{\alpha,(T,U)}/\sigma(\alpha,(T,U))$. These rescaled quantities satisfy the fixed-point equation

$$Z_{\alpha,(T,U)} \stackrel{d}{=} U^\alpha Z_{\alpha,(T,U)} + (1-U)^\alpha Z^*_{\alpha,(T,U)} + \frac{1}{\sigma(\alpha,(T,U))}\left[\frac{\alpha+1}{\alpha-1}\Big(U^\alpha + (1-U)^\alpha - 1\Big) + T\right],$$

where $Z_{\alpha,(T,U)}$, $Z^*_{\alpha,(T,U)}$, (U,T) being independent and $Z_{\alpha,(T,U)} \stackrel{d}{=} Z^*_{\alpha,(T,U)}$. We denote by N, N^* two independent standard normal distributed random variables being independent of the other quantities. Then

$$N(0,1) \stackrel{d}{=} U^{1/2}N + (1-U)^{1/2}N^*.$$

Moreover, we define, similarly to (42), the mixtures

$$M_{\alpha,(T,U)} \stackrel{d}{=} U^\alpha N + (1-U)^\alpha N^* + \frac{1}{\sigma(\alpha,(T,U))}\left[\frac{\alpha+1}{\alpha-1}\Big(U^\alpha + (1-U)^\alpha - 1\Big) + T\right].$$

Then $E(M_{\alpha,(T,U)}) = 0$, $E(M^2_{\alpha,(T,U)}) = 1$, and $E|M_{\alpha,(T,U)}|^r < \infty$; thus the μ_r distances between $Z_{\alpha,(T,U)}$, $N(0,1)$, and $M_{\alpha,(T,U)}$ are finite. It follows that

$$\mu_r(Z_{\alpha,(T,U)}, N(0,1)) \leq \mu_r(Z_{\alpha,(T,U)}, M_{\alpha,(T,U)}) + \mu_r(M_{\alpha,(T,U)}, N(0,1)).$$

A calculation similar to (43) implies, for $\alpha > 1/2$, that

$$\begin{aligned}\mu_r(Z_{\alpha,(T,U)}, M_{\alpha,(T,U)}) &\leq 2E(U^{r\alpha})\mu_r(Z_{\alpha,(T,U)}, N(0,1)) \\ &\leq \frac{2}{1+r/2}\mu_r(Z_{\alpha,(T,U)}, N(0,1)).\end{aligned}$$

Note that the assumptions (i) and (ii) for the dependent case are also satisfied in the case when T and U are independent (see Corollary 1). The asymptotic normality for dependent and independent cases can be derived under conditions (i) and (ii) by proving $\mu_r(M_{\alpha,(T,U)}, N(0,1)) = o(1)$ as $\alpha \downarrow 1/2$. This follows from the convergence in ℓ_r, which is obtained using the fixed-point equations for $N(0,1)$, $M_{\alpha,(T,U)}$, $\|T\|_r = o(\sigma(\alpha,(T,U)))$ and that $\sigma(\alpha,(T,U))$ is properly divergent, giving

$$\begin{aligned}\ell_r(M_{\alpha,(T,U)}, N(0,1)) &\leq 2\,\|U^{1/2} - U^\alpha\|_r \|N\|_r \\ &\quad + \frac{1}{\sigma(\alpha,(T,U))}\left[\left\|\frac{\alpha+1}{\alpha-1}\Big(U^\alpha + (1-U)^\alpha - 1\Big)\right\|_r + \|T\|_r\right],\end{aligned}$$

which tends to zero for $\alpha \downarrow 1/2$ under our assumptions. It follows that

$$\mu_r(Z_{\alpha,(T,U)}, N(0,1)) \leq \frac{2}{1+r/2}\mu_r(Z_{\alpha,(T,U)}, N(0,1)) + o(1),$$

thus $2/(1+r/2) < 1$ implies $\mu_r(Z_{\alpha,(T,U)}, N(0,1)) \to 0$. ▮

Note that in the case $\alpha \downarrow 1/2$ and $T = 1$, which holds especially for deterministic toll functions, all conditions of the theorem are satisfied.

We may endow $(1/2,\infty) \times L_2$ with the metric $d((\alpha,T),(\beta,V)) := |\alpha-\beta| + \|T-V\|_2$. Then, for fixed U, the map $Y : (1/2,\infty) \times L_2 \to \mathcal{M}_2(0)$, $(\alpha,T) \mapsto \mathcal{L}(Y_{\alpha,(T,U)})$ is locally Lipschitz continuous with respect to d and ℓ_2. This follows by making all the constants explicit in the estimate (45) and in the corresponding one for general $\beta > 1/2$.

6 Examples

In this section, we discuss many examples, most of them being new.

The number of exchanges of quicksort. The number of exchanges used by quicksort satisfies (1) with T_n dependent on I_n. While Theorem 1 does not apply, its proof does. The starting point is the recurrence $P_0(y) = 1$ and for $n \geq 1$

$$P_n(y) = \frac{1}{n}\sum_{0\leq k<n} P_k(y)P_{n-1-k}(y) \sum_{0\leq j\leq \min\{k,n-1-k\}} \pi_{n,k,j}e^{jy},$$

where $\pi_{n,k,j}$ denotes the probability that there are exactly j exchanges when the rank of the pivot element is $k+1$; so that (see Sedgewick [70, p. 55])

$$\pi_{n,k,j} = \frac{\binom{k}{j}\binom{n-1-k}{j}}{\binom{n-1}{k}}. \tag{47}$$

Note that the exact number of exchanges used depends on implementation details and we count only the essential random part.

Using the identity

$$\sum_{j\geq 1} \pi_{n,k,j} j(j-1)\cdots(j-v+1) = \frac{(n-v-1)!k!(n-1-k)!}{(n-1)!(k-v)!(n-k-1-v)!} \qquad (v=0,1,2,\dots), \tag{48}$$

and (7), we easily obtain

$$E(X_n) = \frac{n+1}{3} H_n - \frac{7}{9} n + \frac{1}{18} \qquad (n \geq 2).$$

For higher moments, we proceed as in Section 3 ($\alpha = 1$) by defining $\Pi_n(y) := P_n(y) e^{-x_n y}$ and $\Pi_{n,m} := \Pi_n^{(m)}(0)$. Then by the same approach, we deduce that

$$\Pi_{n,m} \sim g_m n^m \qquad (n \geq 2),$$

where $g_0 = 1$, $g_1 = 0$ and for $n \geq 2$

$$g_m = \sum_{a+b+c=m} \binom{m}{a,b,c} g_a g_b \int_0^1 x^a (1-x)^b \left(\frac{x}{3} \log x + \frac{1-x}{3} \log(1-x) + x(1-x) \right)^c \mathrm{d}x.$$

Thus

$$\frac{X_n - x_n}{n} \xrightarrow{d} Y,$$

as well as convergence of all moments, where

$$Y \stackrel{d}{=} UY + (1-U)Y^* + \frac{U}{3} \log U + \frac{1-U}{3} \log(1-U) + U(1-U),$$

with $Y \stackrel{d}{=} Y^*$ and Y, Y^*, U independent.

On the other hand, Theorem 3 applies by establishing

$$\frac{T_n}{n/6} \xrightarrow{L_2} 6U(1-U).$$

This follows from (48).

In particular, by the recurrence of g_m or by Corollary 1, $\mathrm{Var}(X_n) \sim \left(\frac{11}{60} - \frac{\pi^2}{54} \right) n^2$.

Note that by (48)

$$E(T_n^k) \sim E(T_n(T_n-1)\cdots(T_n-k+1)) = \frac{k!k!(n-k-1)!}{(2k+1)!(n-2k-1)!} \sim \frac{k!k!}{(2k+1)!} n^k,$$

for $k \geq 1$. Thus T_n/n has in the limit a beta distribution:

$$P\left(\frac{T_n}{n} < x \right) \to 1 - \sqrt{1-4x} \qquad (0 < x < 1/4).$$

Unlike the number of comparisons, which has quadratic worst-case behavior, the number of exchanges is at most of order $n \log n$. Also it is interesting to note that the histograms of $P(X_n = i)$ are very close to normal curves for n small; see Figure 2. An explanation of this phenomenon is that the leading constant of the variance (as well as g_3) is very small $\frac{11}{60} - \frac{\pi^2}{54} \approx 0.00056288$. The "non-normality character" of Y will emerge for large enough n.

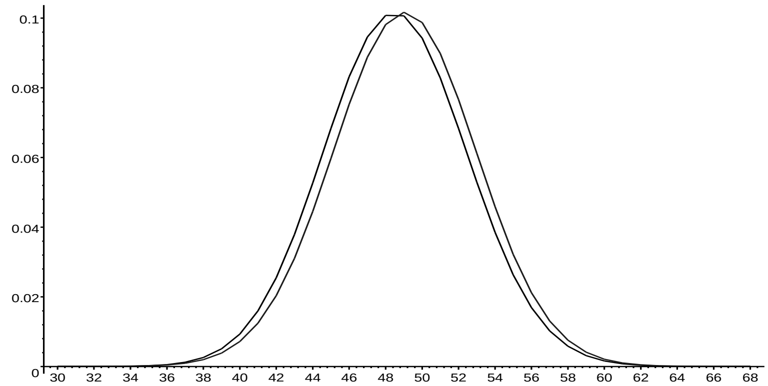

Figure 2: *The histogram of* $P(X_{60} = k)$ *for* k *from 30 to 68 and the corresponding normal curve* $e^{-(k-E(X_{60})-1/2)^2/(2\mathrm{Var}(X_{60}))}/\sqrt{2\pi \mathrm{Var}(X_{60})}$. *We shifted the normal density by* $1/2$*; for otherwise the two curves will be almost indistinguishable.*

The limit law (49) is different from that of the number of comparisons (2); however, the limit distributions are related by their defining fixed-point equations. Indeed, the correlation of the number of comparisons and the number of exchanges is asymptotic to

$$\frac{\sqrt{5}(39-4\pi^2)}{2\sqrt{(21-2\pi^2)(99-10\pi^2)}} \approx -0.864042...$$

This can be proved by the bivariate limit law of both variates that can be derived by a multivariate extension of the contraction method (see Neininger [53] for details). Thus *the number of comparisons and the number of exchanges are highly negatively correlated.* Intuitively, when the shape of the corresponding binary search tree is very skewed, few key exchanges are needed; on the other hand, the number of exchanges reaches its maximum when the pivot element is around $n/2$ (see (47)). Roughly, the more "balanced" the permutation, the more number of exchanges is needed. The situation here is more or less the same when one uses the median-of-$(2t+1)$ quicksort: while the number of comparisons decreases with t, the number of exchanges increases. We might say that we trade off the number of exchanges for the number of comparisons.

Note that the same limit law (2) for $T_n = n + O(1)$ persists for $T_n = n + \omega(n)$, where $\omega(n) = o(n)$ and $\sum_n \omega(n)/n^2 < \infty$; this reflects the "robustness" of the limit laws.

Paged trees. Fix a page (or bucket) size $b \ge 1$. Cut all nodes with subtree sizes $\le b$. The resulting tree is called the b-index of the tree; see Flajolet et al. [27] and Mahmoud [47]. What is the size of a random b-index? And what is the total path length? Obviously, both random variables satisfy (1) (with different initial conditions). The asymptotic normality of the size was established for fixed b by Flajolet et al. [27] with mean equal to $2(n+1)/(b+2) - 1$. The variance is equal to (the expression given in [27] being wrong)

$$2\left(4H_{2b+2} - 4H_{b+1} - \frac{(b+1)(5b+2)}{(2b+3)(b+2)}\right)\frac{n+1}{b+2} \qquad (n \ge 2b+2).$$

Indeed, we can prove that the asymptotic normality holds for $2 \le b = o(n)$. This does not follow directly from our results but easily amended by truncating the first b terms in our exact and asymptotic expressions (6) and by applying the same arguments.

If we vary b such that $2 \le b = o(n)$, then the path length of the b-index gives an interesting example with mean of order n/b, which varies from linear to any function tending to infinity. Thus the limit laws change from non-normal to normal when b increases.

This variation of path length suggests in turn a variation of quicksort: stop subfiles of size less than or equal to b, where b can vary with n. We can show that the limit law of the total number of comparisons used in the quicksort partitioning stages does not change as long as $b = o(n)$. This images another "robustness" of the limit laws.

Leaves and patterns in binary search trees. Our Theorem 2 can be applied to the number of times a given pattern appears in a random binary search tree; see Devroye [17] and Flajolet et al. [27]. The number of times a subtree of size k appears is also asymptotically normally distributed; see Aldous [1] and Devroye [19]. By the correspondence between increasing trees (or binary recursive trees) and permutations, some patterns on trees like the number of leaves also lead to well-known distributions in random permutations; see Bergeron et al. [6] and Flajolet et al. [27].

Analysis of tree traversal algorithms. Binary search trees can be implemented in several different ways: two-pointers, threaded with or without flag, triply linked (with a pointer to parent), etc.; and the nodes can be traversed in different orders: inorder, preorder, postorder, breath-first, depth-first, etc.; see [2, 10, 11, 12, 22, 23, 31, 50, 64] and [30]. The analysis of the cost of these algorithms then reduces to the calculation of certain parameters on trees such as the number of nodes with null (or non-null) left (or right) branches, the number of nodes with both non-null left and right branches, and the number of nodes that are a left child and whose right branch is not empty. All these quantities can be systematically analyzed by applying our results; see Brinck and Foo [11] and Brinck [10] for analysis of the mean of some cost measures.

For example, the major cost (number of pointer operations) needed to traverse a threaded binary search tree in preorder and in inorder is essentially given by (neglecting minor parameters)

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-1-I_n} + T_{I_n},$$

where

$$Q_n(y) := E(e^{T_n y}) = \prod_{1\le k\le n} \frac{k-1+e^y}{k} \qquad (n \ge 1), \tag{49}$$

essentially the Stirling numbers of the first kind (enumerating the number of records in iid sequences, the number of cycles in random permutations, etc.). The distribution of X_n is asymptotically normal. Likewise, the moment generating function $P_n(y)$ of the cost for postorder traversal satisfies

$$P_n(y) = \frac{e^y}{n} \sum_{0\le k\le n-2} P_k(y)P_{n-1-k}(y)Q_k^2(y)V_k(y) + \frac{e^y}{n} P_{n-1}Q_{n-1}^2(y),$$

where $Q_n(y)$ is defined as in (49) and $V_n(y)$ denotes the moment generating function for the depth of the first node in postorder; see (53). The mean was derived by Brinck [10]. Indeed the exact forms of these generating functions are immaterial because our results are strong enough to prove the asymptotic normality of the cost within a large range of variation for the toll function; see also Section 7 for the asymptotic normality of the depth of the first node in postorder.

Secondary parameters of quicksort. If we always sort smaller files first, then the number of stack pushes and pops used to sort a random input satisfies $P_n(y) = 1$ for $n \le 4$ and

$$P_n(y) = \frac{e^y}{n} \sum_{0\le k<n} P_k(y)P_{n-1-k}(y) + \frac{2}{n}(1-e^y)\left(P_{n-1}(y) + P_{n-2}(y)\right) \qquad (n \ge 5).$$

Our results apply and the number of stack pushes is asymptotically normally distributed. If we stop sorting subfiles of sizes less than a certain given value and then use a final insertionsort to complete the sorting, then the number of comparisons and exchanges used by the insertionsort is again normal in the limit. For more information on analysis of quicksort, see Sedgewick [70], Chern and Hwang [13].

Sorting on a broadcast communication model. The model consists of n processors sharing a common channel for communications, allowing one processor to broadcast at each time epoch. To each processor a certain number is attached (the numbers being distinct). The sorting problem is to order these numbers in increasing order. The algorithm proposed in Shiau and Yang [71] is as follows. Select first a loser (a preferable term being "a leader") by the coin-flipping procedure in Prodinger [58]. Split the processors into two subsets containing, respectively, smaller and larger numbers; then sort recursively by the same approach; see Shiau and Yang [71] for details. The number of rounds of coin-tossings (in order to resolve the conflict for using the channel) satisfies (1) with T_n given by $(Q_n(y) := E(e^{T_n y}))$

$$Q_n(y) = \frac{e^y}{2^n} \sum_{1\le k\le n} \binom{n}{k} Q_k(y) + \frac{e^y}{2^n} Q_n(y) \qquad (n \ge 2),$$

with $Q_1(y) = 1$. The mean of X_n is studied by Grabner and Prodinger [32]. By the results of Fill et al. [26], our results apply and X_n is asymptotically normal.

In-situ permutation algorithm. The problem in question is: given a sequence of numbers $\{a_1, \ldots, a_n\}$ and a permutation $\{\pi_1, \ldots \pi_n\}$, output $\{a_{\pi_1}, \ldots, a_{\pi_n}\}$ using at most $O(1)$ space. An algorithm was given by MacLeod [44] and analyzed by Knuth [42]. Kirschenhofer et al. [41] showed that the major cost X_n of the algorithm satisfies the quicksort recurrence (1) with $T_n = I_n$. They extended Knuth's analysis of the first two moments by computing the asymptotics of all moments (non-centered).

Theorem 3 applies and we obtain

$$\frac{X_n - E(X_n)}{n} \xrightarrow{d} Y, \tag{50}$$

where $Y \overset{d}{=} UY + (1-U)Y^* + U \log U + (1-U)\log(1-U) + U$. Note that

$$E(X_n) \sim n \log n, \qquad \mathrm{Var}(X_n) \sim \sigma^2(1, (2U, U))\, n^2 = \left(2 - \frac{\pi^2}{6}\right) n^2.$$

We can indeed prove convergence of all moments using the same approach in Section 3 starting from $P_0(y) = 1$ and

$$P_n(y) = \frac{1}{n} \sum_{0\le k<n} e^{ky} P_k(y) P_{n-1-k}(y) \qquad (n \ge 1). \tag{51}$$

Note that X_n can also be viewed as the *left path length* of random binary search trees (by counting only left branches). In general, one may consider *weighted path length* by assigning weight α to each left branch and β to each right branch in a random binary search tree; our tools apply.

Recursive trees. Interestingly, the limit distribution (50) also appears as the limit distribution of the total path length of random recursive trees; see Dobrow and Fill [21], Mahmoud [46]. This can be explained in two ways. First, by a well-known transformation from multiway trees to binary trees (see Corman et al. [15]), we can actually prove a bijection between the total path length of a recursive tree of n nodes and the left path length of a random binary search tree of $n-1$ nodes, the latter having the same distribution as the major cost of the in-situ permutation algorithm.

Second, the underlying recurrence for total path length of recursive trees is almost identical to (51)

$$X_n \overset{d}{=} X_{J_n} + X_{n-J_n} + J_n,$$

where J_n is uniformly distributed over $\{1, 2, \ldots, n-1\}$.

This connection makes it possible to derive the limit laws of other parameters on recursive trees by our approaches (up to minor modifications) like the number of leaves, the number of nodes with a specified degree, etc.; see Smythe and Mahmoud [72] for a survey of recursive trees. Note that the number of leaves satisfies $P_0(y) = 1$, $P_1(y) = e^y$ and

$$P_n(y) = \frac{1}{n-1} \sum_{1 \le k \le n-2} P_k(y) P_{n-1-k}(y) + \frac{P_{n-1}(y)}{n-1} \qquad (n \ge 2),$$

the underlying distribution being essentially the Eulerian numbers; see Bergeron et al. [6].

Superlinear toll functions. The Wiener index of a graph is defined as the sum of the distances between all pairs of nodes. This index plays an important role in connection with physico-chemical properties (like boiling point, heat of information, crystal defects) of chemical structures; see Gutman et al. [33] and Trinajstić [73]. The Wiener index of a random binary search tree satisfies, neglecting the independence assumptions, (1) with

$$T_n = 2I_n(n-1-I_n) + Z_n + I_n Z^*_{n-1-I_n} + (n-1-I_n) Z'_{I_n},$$

where Z_n denotes the total path length, which satisfies (1) with $T_n = n-1$. The mean is easily seen to be

$$E(X_n) = 2n^2 H_n - 6n^2 + 8nH_n - 10n + 6H_n \qquad (n \ge 1).$$

But our results fail since Z_n and X_n are not independent. The variance satisfies $\mathrm{Var}(X_n) \sim (\frac{20}{3} - \frac{2}{3}\pi^2) n^4$ and the characterization of the limit law of X_n necessitates a multivariate extension of our approach, see Neininger [54] for details.

Other examples. For other examples of the quicksort type leading to an asymptotically normal distribution, see Fill [24], Hofri and Shachnai [37], Panholzer and Prodinger [56], Chern et al. [14].

7 One-sided quicksort recurrence

In this section, we briefly discuss the recurrence (3). Assume that T_n is independent of I_n. Then the moment generating function of X_n satisfies $P_0(y) = 1$ and for $n \ge 1$

$$P_n(y) = \frac{Q_n(y)}{n} \sum_{0 \le k < n} P_k(y),$$

which can be easily solved, by considering the difference $nP_n(y) - (n-1)P_{n-1}(y)Q_n(y)/Q_{n-1}(y)$, giving

$$P_n(y) = Q_n(y) \prod_{0 \le k < n} \frac{k + Q_k(y)}{k+1} \qquad (n \ge 1).$$

Thus $X_n - T_n$ is the sum of independent mixed random variables. The asymptotic transfer from the toll function to the total cost in this case is much simpler.

Lemma 5. *Define* $a_0 = 0$ *and for* $n \ge 1$

$$a_n = b_n + \frac{1}{n} \sum_{0 \le k < n} a_k. \qquad (52)$$

Then

$$a_n = b_n + \sum_{1 \le k < n} \frac{b_k}{k+1} \qquad (n \ge 1).$$

Proof. Omitted. ∎

Lemma 6 (Asymptotic transfer). *Assume a_n satisfies (52). If $b_n \sim n^\alpha L(n)$, where $L(n)$ is slowly varying, then*

$$a_n \sim \begin{cases} \displaystyle\sum_{1\le k<n} \frac{L(k)}{k+1}, & \text{if } \alpha = 0; \\ \displaystyle\frac{\alpha+1}{\alpha} n^\alpha L(n), & \text{if } \alpha > 0. \end{cases}$$

Proof. Omitted. ∎

For the limit laws, we have roughly

$$\begin{aligned} \frac{P_n(y)}{Q_n(y)} &= \prod_{1\le k<n} \left(1 + \frac{Q_k(y)-1}{k+1}\right) \\ &\approx \exp\left(\sum_{1\le k<n} \frac{Q_k(y)-1}{k+1}\right) \\ &\approx \exp\left(y \sum_{1\le k<n} \frac{Q_k'(0)}{k+1} + \frac{y^2}{2} \sum_{1\le k<n} \frac{Q_k''(0)}{k+1} + O\left(|y|^3 \sum_{1\le k<n} \frac{|Q_k'''(0)|}{k+1}\right)\right). \end{aligned}$$

Thus for small toll functions, if

$$\sum_{1\le k<n} \frac{Q_k''(0)}{k+1} \to \infty$$

and

$$\left(\sum_{1\le k<n} \frac{Q_k''(0)}{k+1}\right)^{-3/2} \sum_{1\le k<n} \frac{|Q_k'''(0)|}{k+1} \to 0,$$

then X_n is asymptotically normally distributed.

On the other hand, for larger toll functions, if $T_n/t_n \xrightarrow{d} T$, then roughly

$$\begin{aligned} \frac{P_n(y)}{Q_n(y)} &\approx \exp\left(\sum_{1\le k\le n} \frac{Q_k(y)-1}{k+1}\right) \\ &\approx \exp\left(\int_0^x \frac{Q(v)-1}{v}\, dv\right), \end{aligned}$$

where $Q(y)$ denotes the moment generating function of T.

Instead of making these heuristics rigorous, we state a simpler result, describing mainly the phase change from normal to non-normal laws.

Theorem 6. *Let X_n satisfy (3), where T_n is independent of I_n. Assume that*

$$E(T_n) \sim n^\alpha L(n) \quad \text{and} \quad E\left(\frac{T_n}{t_n}\right)^m \to \tau_m \qquad (m \ge 1),$$

where $\alpha > 0$, and that $Q(z) := \sum_{m\ge 0} \tau_m z^m/m!$ has a nonzero radius of convergence. Then

$$\frac{X_n}{n^\alpha L(n)} \xrightarrow{d} X,$$

with convergence of all moments, where $G(z) := E(e^{zX})$ *satisfies*

$$G(z) = \int_0^1 \exp\left(\frac{1}{\alpha}\int_0^{w^\alpha z} \frac{Q(v)-1}{v}\,\mathrm{d}v\right)\mathrm{d}w,$$

for sufficiently small z*. On the other hand, if*

$$t_n \sim L(n), \quad \text{and} \quad E(|T_n|^m) = O(t_n^m) \qquad (m = 2, 3),$$

and

$$s^2(n) := \sum_{1\le k<n} \frac{Q_k''(0)}{k+1} \to \infty,$$

then

$$\frac{X_n - \sum_{1\le k<n} Q_k'(0)/(k+1)}{s(n)} \xrightarrow{d} N(0,1).$$

Proof. (Sketch) The proof of the asymptotic normality follows from the above argument using moment generating functions and Curtiss's continuity theorem. For large toll functions, we use the method of moments as above by proving

$$P_n^{(m)}(0) \sim g_m n^{m\alpha} L^m(n),$$

where $g_0 = 1$ and for $m \ge 1$

$$g_m = \sum_{0\le j\le m} \binom{m}{j} \frac{g_j}{j\alpha+1}\tau_{m-j}.$$

The required result follows from the same arguments we used for (1). ▮

When $\alpha > 0$, the contraction method gives another access to the limit law, where T_n may depend on I_n.

Theorem 7. *Let* (X_n) *be given by (3). Assume that*

$$E(T_n) \sim n^\alpha L(n) \quad \text{and} \quad \left(\frac{T_n}{E(T_n)}, \frac{I_n}{n}\right) \xrightarrow{L_2} (T, U),$$

where $\alpha > 0$ *and* T *is square-integrable. Then*

$$\ell_2\left(\frac{X_n}{n^\alpha L(n)}, X_{\alpha,(T,U)}\right) \to 0,$$

where $\mathcal{L}(X_{\alpha,(T,U)})$ *is the unique fixed-point of the map*

$$S_{\alpha,(T,U)} : \mathcal{M}_2 \to \mathcal{M}_2, \quad \nu \mapsto \mathcal{L}(U^\alpha Z + T),$$

with $Z, (T, U)$ *being independent and* $\mathcal{L}(Z) = \nu$*.*

Proof. Omitted. ▮

If $T \ne (\alpha+1)(1-U^\alpha)/\alpha$, then

$$\mathrm{Var}(X_n) \sim \sigma^2 n^{2\alpha} L^2(n),$$

where $\sigma = \sigma(\alpha, (T, U))$ is defined by

$$\sigma^2 = \frac{1}{2\alpha} + \frac{2\alpha+1}{2\alpha}\left(\mathrm{Var}(T) + \frac{2(\alpha+1)}{\alpha}E(TU^\alpha) - \frac{2}{\alpha}\right).$$

When $T = (\alpha+1)(1-U^\alpha)/\alpha$, then $\mathrm{Var}(X_n) = o(n^{2\alpha}L^2(n))$.

Tree traversals. The simplest example is when $T_n = 1$ for $n \geq 1$. The distribution is essentially the Stirling numbers of the first kind; see (49). This classical example also appears in a large number of problems; see Bai et al. [3] for some examples. This distribution also has another concrete interpretation: the depth of the first node in inorder traversal.

Interestingly, the depth of the first node in postorder traversal of a random binary search tree satisfies a slightly different recurrence: $P_0(y) = P_1(y) = 1$ and for $n \geq 2$

$$P_n(y) = \frac{e^y}{n} \sum_{1 \leq k < n} P_k(y) + \frac{e^y}{n} P_{n-1}(y), \tag{53}$$

which can be asymptotically solved as

$$P_n(y) = \frac{n^{e^y - 1}}{\Gamma(y)} \varpi(e^y) \left(1 + O(n^{-1})\right) + O(n^{-1}),$$

uniformly for $|y| \leq \delta$, where

$$\varpi(y) = e^y + \int_0^1 w^y e^{yw} \left(1 - y - yw^{-1}\right) \mathrm{d}w.$$

This is derived by applying singularity analysis (see [28]) to the generating function $P(z, e^y) = \sum_n P_n(y) z^n$, which satisfies

$$P(z, y) = (1-z)^{-y} e^z + (1-z)^{-y} e^{-y(1-z)} \int_{1-z}^1 w^y e^{yw} \left(1 - y - yw^{-1}\right) \mathrm{d}w.$$

Therefore the distribution of X_n is asymptotically Poisson with parameter $\log n$ and thus asymptotically normal; see [38]. The mean was discussed by Brinck [10].

Quickselect. The number of comparisons and exchanges used by quickselect to find the smallest (or the largest) elements satisfies (3) with toll functions of linear mean. Our theorems apply and, in particular, the limit law of the number of comparisons is Dickman. The same limit law actually persists for selecting the m-th smallest (or largest) element when $m = o(n)$; see Hwang and Tsai [39] for more details.

The Stirling distribution also naturally appears as the number of partitioning stages used by quickselect to find the smallest or the largest element. This gives yet another addition to the large list of concrete interpretations of the Stirling numbers of the first kind.

Logarithmic product of cycle sizes in random permutation. Permutations can be decomposed into a set of cycles. Given a random permutation of n elements, let $\sigma_1 \leq \cdots \leq \sigma_k$ denote the cycle sizes. Define $X_n := \sum_{1 \leq j \leq k} \log \sigma_j$, which appeared as a good approximation to the logarithmic order of a random permutation. Then X_n satisfies (1) with $T_n = \log n$ and

$$E(e^{X_n y}) = \prod_{1 \leq k \leq n} \left(1 + \frac{y^k - 1}{k}\right).$$

Our result gives the well-known asymptotic normality of X_n with mean $\frac{1}{2} \log^2 n$ and variance $\frac{1}{3} \log^3 n$; see Barbour and Tavaré [4] for further information.

Acknowledgements

We thank Uwe Rösler for helpful suggestions.

References

[1] D. Aldous, Asymptotic fringe distributions for general families of random trees, *Annals of Applied Probability*, **1**, 228–266 (1991).

[2] A. Andersson, A Note on the expected behaviour of binary tree traversals, *Computer Journal*, **33**, 471-472 (1990).

[3] Z.-D. Bai, H.-K. Hwang and W.-Q. Liang, Normal approximations of the number of records in geometrically distributed random variables, *Random Structures and Algorithms*, **13**, 319–334 (1998).

[4] A. D. Barbour and S. Tavaré, A rate for the Erdős-Turán law, *Combinatorics, Probability and Computing*, **3**, 167–176 (1994).

[5] J. L. Bentley and M. D. McIlroy, Engineering a sort function, *Software–Practice and Experience*, **23**, 1249–1265 (1993).

[6] F. Bergeron, P. Flajolet and B. Salvy, Varieties of increasing trees, in *Proceedings of CAAP'92* (Rennes, 1992), Lecture Notes in Computer Science, vol. 581, pp. 24–48, Springer, Berlin, 1992.

[7] P. J. Bickel and P. A. Freedman, Some asymptotic theory for the bootstrap, *Annals of Statistics*, **9**, 1196–1217 (1981).

[8] P. Billingsley, *Probability and Measure*, Third Edition, John Wiley & Sons, New York, 1995.

[9] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular Variation*, Cambridge University Press, Cambridge-New York, 1989.

[10] K. Brinck The expected performance of traversal algorithms in binary trees, *Computer Journal*, **28**, 426–432 (1985).

[11] K. Brinck and N. Y. Foo, Analysis of algorithms on threaded trees, *Computer Journal*, **24**, 148–155 (1981).

[12] W. A. Burkhard, Nonrecursive tree traversal algorithms, *Computer Journal*, **18**, 327–330 (1975); Corrigendum: *Computer Journal*, **20**, 352 (1977).

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

[14] H.-H. Chern, H.-K. Hwang and T.-H. Tsai, An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms, *Journal of Algorithms*, accepted for publication (2002).

[15] T. H. Cormen, C. E. Leiserson, and R. L.Rivest, *Introduction to Algorithms*, MIT Press, 1990.

[16] M. Cramer and L. Rüschendorf, Analysis of recursive algorithms by the contraction method, in *Athens Conference on Applied Probability and Time Series Analysis*, Athens, Greece, 1995, Ed. by C. C. Heyde et al., Lecture Notes in Statistics, **114**, 18-33 (1996).

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

[18] L. Devroye, Universal limit laws for depths in random trees, *SIAM Journal on Computing*, **28**, 409–432 (1999).

[19] L. Devroye, Limit laws for sums of functions of subtrees of random binary search trees, preprint, (2001).

[20] P. Diaconis, Application of the method of moments in probability and statistics, in "Moments in mathematics" (San Antonio, Texas, 1987), pp. 125–142, AMS, Providence, RI (1987).

[21] R. P. Dobrow and J. A. Fill, Total path length for random recursive trees, *Combinatorics, Probability and Computing*, **8**, 317–333 (1999).

[22] B. Dwyer, Simple algorithms for traversing a tree without an auxiliary stack, *Information Processing Letters*, **2**, 143–145 (1974).

[23] T. I. Fenner and G. Loizou, A Note on traversal algorithms for triply linked trees, *BIT*, **21**, 153–156 (1981).

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

[25] J. A. Fill and S. Janson, Smoothness and decay properties of the limiting quicksort density function, in *Mathematics and Computer Science: Algorithms, Trees, Combinatorics, and Probabilities*, Edited by D. Gardy and A. Mokkadem, Birkäuser, Basel, pp. 53–64 (2000).

[26] J. A. Fill, H. Mahmoud, and W. Szpankowski, On the distribution for the duration of a randomized leader election algorithm, *Annals of Applied Probability*, **6**, 1260–1283 (1996).

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

[28] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, *SIAM Journal on Discrete Mathematics*, **3**, 216–240 (1990).

[29] P. Flajolet, P. Poblete and A. Viola, On the analysis of linear probing hashing, *Algorithmica*, **22**, 490-515 (1998).

[30] G. H. Gonnet and R. Baeza-Yates, *Handbook of Algorithms and Data Structures*, Addison-Wesley, Workingham, 1991.

[31] D. Gordon, Eliminating the flag in threaded binary search trees, *Information Processing Letters*, **23**, 209-214 (1986).

[32] P. Grabner and H. Prodinger, Sorting algorithms for broadcast communications: Mathematical analysis, *Theoretical Computer Science*, accepted for publication.

[33] I. Gutman, S. Klavzar, and B. Mohar (Editors), *Fifty Years of the Wiener Index*, volume 35 of *Match* (1997).

[34] P. Hennequin, Combinatorial analysis of quicksort algorithm, *RAIRO Informatique Théorique et Applications*, **23**, 317–333 (1989).

[35] P. Hennequin, *Analyse en moyenne d'algorithme, tri rapide et arbres de recherche*, Ph.D. Thesis, Ecole Polytechnique, 1991.

[36] C. A. R. Hoare, Quicksort, *Computer Journal*, **5**, 10–15, 1962.

[37] M. Hofri and H. Shachnai, Efficient reorganization of binary search trees, *Lecture Notes in Computer Science*, **778**, 152–166 (1994).

[38] H.-K. Hwang, Asymptotics of Poisson approximation to random discrete distributions: an analytic approach, *Advances in Applied Probability*, **31**, 448–491 (1999).

[39] H.-K. Hwang and T.-H. Tsai, Quickselect and Dickman function, *Combinatorics, Probability and Computing*, to appear (2002).

[40] J. JaJa, A perspective on quicksort, *Computing in Science & Engineering*, January/February, 43–49 (2000).

[41] P. Kirschenhofer, H. Prodinger and R. F. Tichy, A contribution to the analysis of in situ permutation, *Glasnik Matematički. Serija III*, **22**, 269–278 (1987).

[42] Knuth, D. E. (1972) Mathematical analysis of algorithms. In *Information Processing '71* (Proc. IFIP Congress, Ljubljana, 1971), Vol. 1: Foundations and systems, pp. 19–27. North-Holland, Amsterdam, 1972.

[43] M. Loève, *Probability Theory. I*, Fourth Edition, Springer-Verlag, New York, 1977.

[44] I. D. G. MacLeod, An algorithm for in-situ permutation, *Australian Computer Journal*, **2**, 16–19 (1970).

[45] M. Maejima and S. T. Rachev, An ideal metric and the rate of convergence to a self similar process, *Annals of Probability*, **15**, 708–727 (1987).

[46] H. M. Mahmoud, Limiting distributions for path lengths in recursive trees, *Probability in Engineering and Information Science*, **5**, 53–59 (1991).

[47] H. M. Mahmoud, *Evolution of Random Search Trees*, John Wiley & Sons, New York, 1992.

[48] H. M. Mahmoud, *Sorting. A Distribution Theory*, Wiley-Interscience, New York, 2000.

[49] H. M. Mahmoud, R. Modarres, and R. T. Smythe, Analysis of quickselect: An algorithm for order statistics, *RAIRO Informatique Théorique et Applications* **29**, 255-276 (1995).

[50] J. M. Morris, Traversing binary trees simply and cheaply, *Information Processing Letters*, **9**, 197–200 (1979).

[51] R. Neininger, *Limit Laws for Random Recursive Structures and Algorithms*, Ph.D. Dissertation, Institut für Mathematische Stochastik, Universität Freiburg, 1999; available via the link `http://www.stochastik.uni-freiburg.de/homepages/neininger/`.

[52] R. Neininger, On binary search tree recursions with monomials as toll functions, *Journal of Computational and Applied Mathematics*, accepted for publication (2001); available via the link `http://www.stochastik.uni-freiburg.de/homepages/neininger/`.

[53] R. Neininger, On a multivariate contraction method for random recursive structures with applications to quadtrees, preprint (2000); available via the link `http://www.stochastik.uni-freiburg.de/homepages/neininger/`.

[54] R. Neininger, Wiener index of random trees, *Combinatorics, Probability and Computing*, accepted for publication (2002); available via the link `http://www.stochastik.uni-freiburg.de/homepages/neininger/`.

[55] R. Neininger and L. Rüschendorf, A general contraction theorem and asymptotic normality in combinatorial structures, preprint (2001); available via the link `http://www.stochastik.uni-freiburg.de/homepages/neininger`.

[56] A. Panholzer and H. Prodinger, Binary search tree recursions with harmonic toll functions, *Journal of Computational and Applied Mathematics*, accepted for publication.

[57] B. Pittel, Normal convergence problem? Two moments and a recurrence may be the clues, *Annals of Applied Probability*, **9**, 1260–1302 (1999).

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

[59] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, John Wiley, New York, (1991).

[60] S. T. Rachev and L. Rüschendorf, A new ideal metric with applications to multivariate stable limit theorems, *Probability Theory and Related Fields*, **94**, 163–187 (1992).

[61] S. T. Rachev and L. Rüschendorf, Probability metrics and recursive algorithms, *Advances in Applied Probability*, **27**, 770–799 (1995).

[62] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems. Vol. 1: Theory*, Springer, New York, (1998).

[63] M. Régnier, A limiting distribution for quicksort, *RAIRO Informatique Théorique et Application*, **23**, 335–343 (1989).

[64] J. M. Robson, An improved algorithm for traversing binary trees without auxiliary stack, *Information Processing Letters*, **2**, 12–14 (1973).

[65] U. Rösler, A limit theorem for "Quicksort," *RAIRO Informatique Théorique et Applications*, **25**, 85–100 (1991).

[66] U. Rösler, A fixed point theorem for distributions, *Stochastic Processes and their Applications*, **42**, 195–214 (1992).

[67] U. Rösler, The analysis of stochastic divide and conquer algorithms, *Algorithmica*, **29**, 238–261 (2001).

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

[69] W. Schachinger, Limiting distribution of the costs of partial match retrievals in multidimensional tries, *Random Structures and Algorithms*, **17**, 428–459 (2000).

[70] R. Sedgewick, *Quicksort*, Gurland Publishing, NY, 1980.

[71] S.-H. Shiau and C.-B. Yang, A fast sorting algorithm and its generalization on broadcast communications, in Proceedings COCOON 2000, *Lecture Notes in Computer Science*, **1858**, 252–261 (2000).

[72] R. T. Smythe and H. M. Mahmoud, A survey of recursive trees, *Theory of Probability and Mathematical Statistics*, **51**, 1–27 (1996).

[73] N. Trinajstić, *Chemical Graph Theory*, Volume II, CRC Press, Boca Raton, Fla., 1983.

[74] J. D. Valois, Introspective sorting and selection revisited, *Software–Practice and Experience*, **30**, 617–638 (2000).

[75] V. M. Zolotarev, Ideal metrics in the problem of approximating distributions of sums of independent random variables, *Theory of Probability and Applications*, **22**, 433–449 (1977).