行政院國家科學委員會專題研究計畫 成果報告

隨機樹節點的外出數度
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 97-2628-M-009-008-
執 行 期 間 ：97 年 08 月 01 日至 98 年 07 月 31 日
執 行 單 位 ：國立交通大學應用數學系（所）

計畫主持人：符麥克

計畫參與人員：此計畫無其他參與人員

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 98 年 09 月 24 日

Project: The out-degree of nodes in random trees

by

Michael Fuchs

1 General

This is the final report on the National Science Council Project entitled "The out-degree of nodes in random trees" with number 97-2628-M-009-008 and time period from August 1st, 2008 to July 31st, 2009.

Before presenting our results in more details, we give an overview of the main outcomes of the project.

- The preprint [2] was written within this project. It contains the main findings of the project (the preprint is enclosed).
- We gave an invited talk at the 9th International Conference on Finite Fields and Their Applications, Dublin, Ireland, July 13-19, 2009.

2 Results

The original purpose of this project was to apply our recent variant of the moment-transfer approach[1] [1], [5] to the analysis of the out-degree of nodes in certain classes of random trees. However, this task turned out to be too complicated. Hence, we slightly changed the focus of this project. More precisely, we investigated further the applicability of the moment-transfer approach (in its standard form) to some problems for which the approach was known to fail. Since this somehow deviates from our original proposal, we start by explaining the new problem and then present our findings.

The moment-transfer approach was used in several recent papers for deriving limit laws of sequences of random variables that satisfy a distributional recurrence. The approach consist of the following steps: first, one observes that all (centered and non-centered) moments satisfy the same type of recurrence; second, one studies this underlying recurrence and obtains transfer theorems; third, one uses the transfer theorems to derive an asymptotic expansion of the mean value; fourth, one shifts the mean; fifth, one uses induction together with the asymptotic of the mean value and the transfer theorems to derive the first order asymptotic of all higher moments; finally, one identifies the limiting distribution via the moment sequence. It was well-known that this approach does not work for some sequence of random variables satisfying particular easy recurrences such as

$$X_n \stackrel{d}{=} X_{I_n} + 1, \qquad (n \geq 1), \tag{1}$$

where $X_0 = 0, I_n =$ Uniform$\{0, \dots, n-1\}$, and $(I_n)_{n\geq 1}$ and $(X_n)_{n\geq 0}$ are independent.

In this project, we proposed a variant of the above scheme which can be applied to (1). Roughly speaking, the main observation was that all higher moments satisfy an expansion of a certain shape which is proved by induction. The same expansion also holds for the central moments. Then, step five above just becomes a claim concerning the leading term of this expansion.

Overall, the new variant of the moment-transfer approach which can be applied to (1) consists of two induction steps instead of only one, where the first step serves as input for the second step. Our approach yields the following result.

[1] This approach has been called *method of moments* before. However, since this name is slightly misleading, we will use *moment-transfer approach* instead.

Theorem 1. *As $n \to \infty$,*

$$\frac{X_n - \log n}{\sqrt{\log n}} \xrightarrow{d} \mathcal{N}(0,1).$$

It should be mentioned that our method of proof of the result is not the only one. More precisely, it is not complicated to compute the moment-generating function from (1). Then, the central limit theorem can also be proved by classical tools. Yet another approach is based on the contraction method and can be found in [10]. However, the important feature of our work is that our method is of some generality and can be applied to several related problems. We will exemplify some of the results which can be achieved by our approach in the sequel (for more details the reader is referred to [2]).

Analysis of Priority Trees. Priority trees have been analyzed in [8], [11], [12]. All limit laws proved in these papers can be re-proved by using our approach. For instance, let X_n denote the number of key comparisons when inserting a random key into a random priority tree of size n. Then, the underlying recurrence (satisfied by the centered and non-centered moment) is given by

$$a_n = \sum_{j=0}^{n-1} \left(\sum_{l=0}^{2} c_l \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} \right) a_j + b_n,$$

where c_0, c_1, c_2 are suitable constants and b_n is a given sequence. Transfer theorems for this recurrence were already obtained in [4]. Hence, by an (almost automatic) application of our method the following result can be proved.

Theorem 2. *As $n \to \infty$,*

$$\frac{X_n - \log^2 n/3}{\sqrt{10 \log^3 n/81}} \xrightarrow{d} \mathcal{N}(0,1).$$

It should be mentioned that our method is easier than the method suggested by the authors in [8]. Actually, both methods are similar in the sense that they both work with the moments. However, the authors in [8] do not shift the mean. Consequently, they need a very precise knowledge of the moments in order to handle the massive cancellations which are completely avoided in our approach.

Successful and Unsuccessful Search in Binary Search Trees. These are classical quantities whose analysis can for instance be found in [9]. The limit laws can be re-derived with our approach. The important tools are again the transfer theorems which have already been obtained in [6].

Depth in Variants of Binary Search Trees. For simplicity we just concentrate on m-ary search trees (other variants of binary search trees are discussed in [2]). So, let X_n denote the depth in a random m-ary search tree build from n records. Again our approach applies, where transfer theorems for the underlying recurrence can be found in [3]. Then, the following result follows.

Theorem 3. *As $n \to \infty$,*

$$\frac{X_n - \log n/(H_m - 1)}{\sqrt{\left(H_m^{(2)} - 1\right) \log n / \left(H_m - 1\right)^3}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $H_m = \sum_{j=1}^{m} 1/j$ and $H_m^{(2)} = \sum_{j=1}^{m} 1/j^2$.

Number of Collisions in the $\beta(2,b)$-Coalescent. This is an example from mathematical biology and can be found in [7]. The authors asked for a proof of their main result (a central limit theorem) directly from the recurrence satisfied by the random variables. Indeed, our approach is applicable if one can proof the following conjecture for the recurrence

$$a_n = \sum_{j=1}^{n-1} \pi_{n,j} a_{n-j} + b_n, \qquad (n \geq 2), \tag{2}$$

where $a_1 = 0$ and

$$\pi_{n,j} = P(I_n = j) = \frac{\Gamma(n-j+b-1)\Gamma(n+1)}{(j+1)\Gamma(n-j)\Gamma(n+b)H(n,b)}$$

with $H(n,b) = b/(b+n-1) + \Psi(b+n-1) - \Psi(b) - 1$ (here Ψ is the digamma function).

Conjecture 1. *Consider (2). Let $b_n = \mathcal{O}(1/n^{\epsilon})$ with $\epsilon > 0$ suitable small. Then,*

$$a_n = c + \mathcal{O}(1/n^{\epsilon}),$$

where c is a suitable constant.

3 Summary

The original purpose of this project was to apply the moment-transfer approach to the out-degree of nodes in various classes of random trees. This goal, however, turned out to be too complicated (nevertheless, we hope to come back to this problem in the future).

Consequently, we shifted our focus to a different (but related) problem, namely, the applicability of the moment-transfer approach to certain one-sided distributional recurrence. We proposed a new variant of the moment-transfer approach which can be applied to such situations. Moreover, we demonstrated the power of our approach by applying it to several parameters from the analysis of algorithms.

One parameter which we have not considered, but which is likely to be treatable with our approach as well is the distance between two random nodes in binary search trees and its variants. This parameter is important in finger search. A future project might be dedicated to the analysis of this parameter.

References

[1] H. Chang and M. Fuchs (2009). Limit theorems for patterns in phylogenetic trees, *J. Math. Biol.*, in press.

[2] C.-H. Chern and M. Fuchs (2009). On the moment-transfer approach for random variables satisfying a one-sided distributional recurrence, preprint.

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

[4] H.-H. Chern, H.-K. Hwang, T.-H. Tsai (2002). An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms, *Journal of Algorithms*, **44**, 177-225.

[5] M. Fuchs (2008). Subtree sizes in recursive trees and binary search trees: Berry-Esseen bounds and Poisson approximations, *Combin. Probab. Comput.*, **17**, 661-680.

[6] H.-K. Hwang and R. Neininger (2002). Phase change of limit laws in the quicksort recurrences under varying toll functions, *SIAM Journal on Computing*, **31**, 1687-1722.

[7] A. Iksanov, A. Marynych, M. Möhle (2008). On the number of collisions in beta$(2, b)$-coalesents, *Bernoulli*, **15**, 829-845.

[8] M. Kuba and A. Panholzer (2007). Analysis of insertion costs in priority trees, in "Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithmics and Combinatorics", *SIAM Philadelphia*, 175-182.

[9] H. M. Mahmoud (1992). Evolution of Random Search Trees, Wiley, New York.

[10] R. Neininger and L. Rüschendorf (2004). On the contraction method with degenerate limit equation, *The Annals of Probability*, **32**, 2838-2856.

[11] A. Panholzer (2008). Analysis for some parameters for random nodes in priority trees, *Discrete Mathematics and Theoretical Computer Science*, **10**, 1-38.

[12] A. Panholzer and H. Prodinger (1998). Average case analysis priority trees: a structure for priority queue administration, *Algorithmica*, **22**, 600-630.

On the Moment-Transfer Approach for Random Variables satisfying a One-Sided Distributional Recurrence

Che-Hao CHEN and Michael FUCHS*
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, 300
Taiwan

September 24, 2009

Abstract

The moment-transfer approach is a standard tool for deriving limit laws of sequences of random variables satisfying a distributional recurrence. However, so far the approach could not be applied to certain recurrences which are "one-sided". In this paper, we propose a modified version of the moment-transfer approach which can be applied to such recurrences. Moreover, we demonstrate the usefulness of our approach by re-deriving several recent results in an almost automatic fashion.

1 Introduction

In Combinatorics and Computer Science, one often encounters sequence of random variables which satisfy a distributional recurrence. For instance, the following recurrence arises in the analysis of quicksort (see [8] for background): let X_n be a sequence of random variables satisfying

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

where $X_0 = 0$ and $I_n = \text{Uniform}\{0, \dots, n-1\}$, $X_n \stackrel{d}{=} X^*_n$ with $(I_n)_{n\geq 1}, (X_n)_{n\geq 0}, (X^*_n)_{n\geq 0}$ independent. One is then normally interested in properties such as asymptotic behavior of mean and variance as well as deeper properties such as limit laws, rates of convergence, etc.

As for limit laws, the so-called *moment-transfer approach*[1] has evolved into a major tool in recent years. Roughly speaking, the approach consists of the following steps: first, one observes that all moments (centered or non-centered) of X_n satisfy a recurrence of the same type (the so-called *underlying recurrence*). For instance, the underlying recurrence for X_n above is given by

$$a_n = \frac{2}{n}\sum_{j=0}^{n-1} a_j + b_n, \qquad (n \geq 1),$$

*Partially supported by National Science Council under the grant NSC-97-2628-M-009-008.

[1]This approach has been called *method of moments* in most previous works. However, since this name might be misleading, we decided to use *moment-transfer approach* instead.

where $a_0 = 0$ and b_n is a given sequence (called the *toll sequence*). Second, one derives general results that link the asymptotic behavior of b_n to that of a_n (called *transfer theorems*). Third, one uses the transfer theorems to obtain an asymptotic expansion for the mean. Forth, one derives the recurrences of the central moments (this step is called *shifting-the-mean*). Fifth, one uses the transfer theorems together with the expansion for the mean and induction to derive the first order asymptotic of all higher moments. Finally, the limit law is identified via the limit moment sequence. This approach has been used to treat numerous examples; see [1] and the survey article [7] for many recent references.

Overall, the main ingredients in the moment-transfer approach are the transfer theorems, the remaining steps being almost automatic. However, maybe surprisingly, the approach does not work for some sequences of random variables satisfying particular easy distributional recurrences. One such example is given by the one-sided variant of (1). More precisely, let X_n be a sequence of random variables satisfying

$$X_n \stackrel{d}{=} X_{I_n} + 1, \qquad (n \geq 1), \tag{2}$$

where $X_0 = 0$ and $I_n = \text{Uniform}\{0, \dots, n-1\}$ with $(I_n)_{n\geq 1}$ and $(X_n)_{n\geq 0}$ independent.

We provide some more details to illuminate where the approach fails. Therefore, observe that the underlying recurrence is given by

$$a_n = \frac{1}{n}\sum_{j=0}^{n-1} a_j + b_n, \qquad (n \geq 1), \tag{3}$$

where $a_0 = 0$ and b_n is a given sequence. The next step is to obtain transfer theorems. For our crude purpose the following transfer theorems are enough: for α a non-negative integer, we have

$$\begin{array}{lll} \text{(i)} & b_n \sim \log^\alpha n & \Longrightarrow \quad a_n \sim \log^{\alpha+1} n/(\alpha+1) \\ \text{(ii)} & b_n = \mathcal{O}\left(\log^\alpha n\right) & \Longrightarrow \quad a_n = \mathcal{O}\left(\log^{\alpha+1} n\right) \end{array}$$

(these and more precise results will be proved in the next section; see also [8]). Now, the mean $\mathbf{E}(X_n)$ satisfies (3) with $b_n = 1$. Hence, by transfer (i) above $\mathbf{E}(X_n) \sim \log n$. Next, we are going to shift the mean. Therefore, let $A_n^{[r]} = \mathbf{E}(X_n - \mathbf{E}(X_n))^r$. Then,

$$A_n^{[r]} = \frac{1}{n}\sum_{j=0}^{n-1} A_j^{[r]} + B_n^{[r]}, \qquad (n \geq 1),$$

where $A_0^{[r]} = 0$ and

$$B_n^{[r]} = \sum_{k=0}^{r-1} \binom{r}{k} \frac{1}{n} \sum_{j=0}^{n-1} A_j^{[k]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^{r-k}.$$

Let us first look at the variance which is obtained by setting $r = 2$. This yields

$$B_n^{[2]} = \frac{1}{n}\sum_{j=0}^{n-1} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^2 \sim \int_0^1 (1 + \log x)^2 \, \mathrm{d}x = 1,$$

where we have used the asymptotic of the mean. Hence, again by transfer (i) above $\mathrm{Var}(X_n) \sim \log n$. Finally, we want to generalize the latter argument to obtain the first order asymptotic of all central moments. Here, we should mention that it is well-known that X_n (suitable centralized and normalized) is asymptotically normal. Hence, due to the Fréchet-Shohat theorem, our goal is to show that for all $m \geq 0$

$$A_n^{[2m]} \sim \frac{(2m)!}{2^m m!} \log^m n \quad \text{and} \quad A_n^{[2m+1]} = \mathcal{O}\left(\log^m n\right). \tag{4}$$

Note that the claim trivially holds for $m = 0$. As for the induction step assume that the claim is proved for all $m' < m$. Then, in order to prove it for m, we first look at the toll sequence. In the even case, we have

$$B_n^{[2m]} = \sum_{k=0}^{2m-1} \binom{2m}{k} \frac{1}{n} \sum_{j=0}^{n-1} A_j^{[k]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^{2m-k}.$$

It is easy to see that the term with $k = 2m - 1$ is the dominant one. Hence,

$$B_n^{[2m]} \sim \frac{2m}{n} \sum_{j=0}^{n-1} A_j^{[2m-1]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right) = \mathcal{O}\left(\log^{m-1} n\right).$$

Then, by the transfer (ii) above, we obtain $A_n^{[2m]} = \mathcal{O}\left(\log^m n\right)$. This is, however, not strong enough to imply the claim. A similar problem occurs as well when considering odd central moments.

As already mentioned above, there are other approaches to show that X_n (suitable centralized and normalized) is asymptotic normal (more precisely, it is not complicated to compute the characteristic function of X_n. Then, asymptotic normality can be proved by classical tools; see [8] and [13] for another approach based on the contraction method). However, it is still an interesting question whether or not the moment-transfer approach can be modified such that it applies to X_n. To provide such a modification is the purpose of this work. Moreover, we will see that our modified version of the moment-transfer approach can be applied rather automatically to various examples from the analysis of algorithms all of them having the common feature that the recurrence satisfied by the sequence of random variables is one-sided in a similar sense as (2) above.

We conclude the introduction by giving a short sketch of the paper. In the next section, we are going to introduce our approach and apply it to X_n above. Then, in the third section, we will re-derive recent results on priority trees. This will put these results in a larger context. Moreover, our approach will yield proofs which are simpler than the previous ones. In a final section, we will discuss further examples which can be handled by our approach as well.

Notations. We will use ϵ to denote a sufficient small constant which might change from one occurrence to the next. Similarly, $\mathfrak{Pol}(x)$ will denote an unspecified polynomial which again might change from one occurrence to the next. Moreover, if needed, we will indicate its degree as a subindex.

2 Asymptotic Normality of the Stirling Cycle Distribution

In this section, we will show how to modify the moment-transfer approach such that it can be applied to (2).

Before starting, we will give some motivation as for why we are interested in (2). The easiest example of a sequence X_n leading to (2) is the number of cycles in a random permutation of size n. Indeed, let $\sigma_1 \cdots \sigma_k$ denote the the canonical cycle decomposition of a permutation of size n. Then, it is easy to see that the probability that σ_k has length j equals $1/n$. Consequently,

$$X_n \stackrel{d}{=} X_{n-I_n} + 1 \stackrel{d}{=} X_{I_n} + 1.$$

Hence, X_n satisfies our recurrence. Of course, the probability distribution of X_n is well-known

$$P(X_n = k) = \frac{c(n,k)}{n!},$$

where $c(n,k)$ denote the Stirling cycle numbers (or Stirling numbers of first kind). Apart from this interpretation of X_n, there are many others; see [8] and references therein.

Now, we are going to explain our modified moment-transfer approach which can be applied to prove asymptotic normality of X_n (suitable centralized and normalized). Again, the main ingredient will be a transfer theorem.

Proposition 1. *Consider (3).*

(i) Let $b_n = \mathcal{O}(1/n^{\epsilon})$ with $\epsilon > 0$ suitable small. Then,

$$a_n = c + \mathcal{O}\left(1/n^{1-\epsilon}\right),$$

where c is a suitable constant.

(ii) Let $b_n = \log^{\alpha} n$ with $\alpha \in \{0, 1, \ldots\}$. Then,

$$a_n = \frac{\log^{\alpha+1} n}{\alpha+1} + \mathfrak{Pol}_{\alpha}(\log n) + \mathcal{O}(1/n^{\epsilon}),$$

where $\epsilon > 0$ is suitable small.

(iii) Let $b_n = \mathcal{O}(\log^{\alpha} n)$ with $\alpha \in \{0, 1, \ldots\}$. Then, $a_n = \mathcal{O}\left(\log^{\alpha+1} n\right)$.

(iv) Item (iii) holds with $\mathcal{O}$ replaced by o as well.

Proof. It is easy to check that (3) has the general solution

$$a_n = b_n + \sum_{j=1}^{n-1} \frac{b_j}{j+1}, \qquad (n \geq 1). \tag{5}$$

Now, in order to prove (i) observe that

$$a_n = \mathcal{O}\left(\frac{1}{n^{\epsilon}}\right) + \sum_{j=1}^{n-1} \frac{b_j}{j+1} = \sum_{j=1}^{\infty} \frac{b_j}{j+1} + \mathcal{O}\left(\frac{1}{n^{\epsilon}}\right),$$

where the series is absolute convergent due to the assumption.

Also part (ii) immediately follows from (5) by a standard application of Euler-Maclaurin summation formula. Note that alternatively (ii) can also by deduced from (i) by induction.

Finally, part (iii) and (iv) are simple consequences of part (ii). ∎

The next step is to look at the mean value. Therefore, let us more generally consider r-th moments. Set $\bar{A}_n^{[r]} = \mathbf{E}(X_n^r)$. Then,

$$\bar{A}_n^{[r]} = \frac{1}{n} \sum_{j=0}^{n-1} \bar{A}_j^{[r]} + \bar{B}_n^{[r]}, \qquad (n \geq 1),$$

where $\bar{A}_0^{[r]} = 0$ and

$$\bar{B}_n^{[r]} = \sum_{k=0}^{r-1} \binom{r}{k} \frac{1}{n} \sum_{j=0}^{n-1} \bar{A}_j^{[k]}.$$

Setting $r = 1$ gives the toll sequence $\bar{B}_n^{[1]} = 1$ (as in Section 1). Hence, by our transfer theorem

$$\mathbf{E}(X_n) = \bar{A}_n^{[1]} = \log n + \mathfrak{Pol}_0(\log n) + \mathcal{O}(1/n^{\epsilon}),$$

where $\epsilon > 0$ is suitable small.

The main new step in our modified version of the moment-transfer approach is another induction to show that the above form of the expansion for the mean continues to hold for all higher moments.

Proposition 2. *For all $r \geq 1$, we have*

$$\bar{A}_n^{[r]} = \mathfrak{Pol}(\log n) + \mathcal{O}\left(1/n^{\epsilon}\right),$$

where $\epsilon > 0$ is suitable small.

Proof. Note that the claim was already proved for $r = 1$. Assume that it holds for all $r' < r$. In order to prove it for r note that the induction hypothesis and Euler-Maclaurin summation formula imply that $\bar{B}_n^{[r]} = \mathfrak{Pol}(\log n) + \mathcal{O}\left(1/n^{\epsilon}\right)$. Hence, the claim follows by the transfer theorem. ∎

Next, we turn to central moments whose recurrence was already mentioned in Section 1. Note that since

$$A_n^{[r]} = \sum_{k=0}^{r} \binom{r}{k} \bar{A}_n^{[k]} \left(\mathbf{E}(X_n)\right)^{r-k}$$

the form of the expansion from Proposition 2 also holds for all the central moments. Hence, (4) is in fact only a claim concerning the leading term of this expansion.

Proposition 3. *For all $m \geq 0$, we have*

$$A_n^{[2m]} \sim \frac{(2m)!}{2^m m!} \log^m n \quad \textit{and} \quad A_n^{[2m+1]} = \mathcal{O}\left(\log^m n\right).$$

Proof. Note that the claim trivially holds for $m = 0$. Assume now that the claim holds for all $m' < m$. We are going to prove it for m.

First, consider the even case. Then, the toll sequence is given by

$$B_n^{[2m]} = \sum_{k=0}^{2m-1} \binom{2m}{k} \frac{1}{n} \sum_{j=0}^{n-1} A_j^{[k]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^{2m-k}.$$

We start by looking at the contribution of $k = 2m - 1$ which is

$$\frac{2m}{n} \sum_{j=0}^{n-1} A_j^{[2m-1]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right) \sim c \log^{m-1} n \int_0^1 (1 + \log x) \mathrm{d}x,$$

where c is a suitable constant. Since the above integral vanishes this part contributes $o(\log^{m-1} n)$. Next, consider $k = 2m - 2$ which gives

$$\begin{aligned}\frac{2m(2m-1)}{2n} \sum_{j=0}^{n-1} A_j^{[2m-2]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^2 &\sim \frac{(2m)!}{2^m (m-1)!} \log^{m-1} n \int_0^1 (1 + \log x)^2 \,\mathrm{d}x \\ &= \frac{(2m)!}{2^m (m-1)!} \log^{m-1} n.\end{aligned}$$

As for all other parts, using a similar reasoning shows that they contribute $o(\log^{m-1} n)$. Hence,

$$B_n^{[2m]} \sim \frac{(2m)!}{2^m (m-1)!} \log^{m-1} n.$$

Using the transfer theorem proves the claim in the even case.

As for the odd case, here the toll sequence becomes

$$B_n^{[2m+1]} = \sum_{k=0}^{2m} \binom{2m+1}{k} \frac{1}{n} \sum_{j=0}^{n-1} A_j^{[k]} \left(1 + \mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^{2m+1-k}.$$

Using similar reasoning as above, the term with $k = 2m$ contributes $o(\log^m n)$. All other terms give a smaller contribution. Hence, $B_n^{[2m+1]} = o(\log^m n)$. By the transfer theorem $A_n^{[2m+1]} = o(\log^{m+1} n)$. Due to the remark preceding the proposition this implies our claim in the odd case. ∎

Overall, we have proved the following result.

Theorem 1. *As $n \to \infty$, we have*

$$\frac{X_n - \log n}{\sqrt{\log n}} \xrightarrow{d} \mathcal{N}(0,1).$$

To summarize, the only difference of our approach to the previous version of the moment-transfer approach are two induction steps instead of only one. The first induction step establishes a certain shape of all higher moments (centered and non-centered). Then, the second induction is used to derive more details concerning the leading term. Again, the main tool is the transfer theorem. Once such a result is established, the remaining proof is rather automatic.

We will apply our new approach to a couple of other examples in the subsequent sections.

3 Analysis of Priority Trees

Priority trees have been analyzed in several recent papers; see [9], [14], [15]. Since we are here just interested in the applicability of our modified moment-transfer approach, we will just give the probabilistic problem and direct the interested reader to the latter papers for background.

Length of the Left Path. We only briefly discuss this example due to its similarity to the example from the previous section. Let X_n be the length of the left path in a random priority tree build from n records. Then, we have

$$\begin{aligned} X_n &\stackrel{d}{=} Y_{I_n} + Z_{n-1-I_n}, \qquad (n \geq 1), \\ Y_n &\stackrel{d}{=} Y_{I_n} + 1, \qquad (n \geq 1), \\ Z_n &\stackrel{d}{=} Z_{I_n} + 1, \qquad (n \geq 1), \end{aligned}$$

where $X_0 = Z_0 = 0, Y_0 = 1$ and $I_n = \text{Uniform}\{0, \ldots, n-1\}$ with $(I_n)_{n\geq 1}, (Y_n)_{n\geq 0}, (Z_n)_{n\geq 0}$ independent.

So, the central moments of Y_n and Z_n can be treated as in the previous section. Moreover, due to the first recurrence, the (centered and non-centered) moments of X_n are connected to those of Y_n and Z_n. Using this connection it is straightforward to prove that $\mathbf{E}(X_n) \sim 2\log n$ and that the r-th central moment of X_n (denoted as in the previous section) satisfies

$$A_n^{[2m]} \sim \frac{(2m)!}{m!} \log^m n \quad \text{and} \quad A_n^{[2m+1]} = \mathcal{O}\left(\log^m n\right).$$

Consequently, we have the following result.

Theorem 2. *As $n \to \infty$, we have*

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

Number of Key Comparisons for Insertion. This is a more sophisticated example whose proof of the central limit theorem was briefly sketched in [9]. We will see that our approach applies quite straightforwardly. So, let X_n denote the number of key comparisons when inserting a random node in a random priority tree build from n records. Then, for $n \geq 1$,

$$X_n|\,(I_n=j) \stackrel{d}{=} \begin{cases} Y_j + U_{n-1-j}, & \text{with probability } (j+1)/(n+1), \\ Z_{n-1-j} & \text{with probability } (n-j)/(n(n+1)), \end{cases}$$

$$Y_n|\,(I_n=j) \stackrel{d}{=} \begin{cases} Y_j + 1, & \text{with probability } (j+1)/(n+1), \\ X_{n-1-j}+2 & \text{with probability } (n-j)/(n(n+1)), \end{cases}$$

$$Z_n|\,(I_n=j) \stackrel{d}{=} \begin{cases} X_j + U_{n-1-j}+2, & \text{with probability } (j+1)/(n+1), \\ Z_{n-1-j} & \text{with probability } (n-j)/(n(n+1)), \end{cases}$$

where $P(I_n = j) = 1/n, 0 \leq j < n, X_0 = 0, Y_0 = Z_0 = 1$, the probability generating function of U_n is given by

$$\mathbf{E}\left(w^{U_n}\right) = \binom{w+n-1}{n},$$

and $(U_n)_{n\geq 0}, (X_n)_{n\geq 0}, (Y_n)_{n\geq 0}$ are independent.

The first step is to find the underlying recurrence which needs some tedious (but straightforward) computations. Therefore, let

$$X(s,t) = \sum_{n\geq 0}(n+1)\mathbf{E}\left(e^{tX_n}\right)s^n;$$
$$Y(s,t) = \sum_{n\geq 0}(n+1)\mathbf{E}\left(e^{tY_n}\right)s^n;$$
$$Z(s,t) = \sum_{n\geq 0}(n+1)\mathbf{E}\left(e^{tZ_n}\right)s^n.$$

Then, from the above distributional recurrences, we get

$$\frac{\partial}{\partial s}X(s,t) = \frac{1}{(1-s)^{e^t}}Y(s,t) + \frac{1}{1-s}Z(s,t);$$
$$\frac{\partial}{\partial s}Y(s,t) = \frac{e^t}{1-s}Y(s,t) + \frac{e^{2t}}{1-s}X(s,t);$$
$$\frac{\partial}{\partial s}Z(s,t) = \frac{e^{2t}}{(1-s)^{e^t}}X(s,t) + \frac{1}{1-s}Z(s,t)$$

with initial conditions $X(0,t) = 1$ and $Y(0,t) = Z(0,t) = e^t$. Eliminating $Y(s,t)$ and $Z(s,t)$ gives

$$\frac{\partial^3}{\partial s^3}X(s,t) - \frac{3+2e^t}{1-s}\frac{\partial^2}{\partial s^2}X(s,t) + 2e^t\left(\frac{2}{(1-s)^2} - \frac{e^t}{(1-s)^{e^t+1}}\right)\frac{\partial}{\partial s}X(s,t) + \frac{2e^{2t}}{(1-s)^{e^t+2}}X(s,t) = 0$$

with initial conditions $X(0,t) = 1, \frac{\partial}{\partial s}X(0,t) = 2e^t, \frac{\partial^2}{\partial s^2}X(0,t) = 2e^t + 4e^{2t}$. Now, let $\bar{P}_n(t) = (n+1)\mathbf{E}(e^{tX_n})$. Reading off coefficients from the above differential equation yields

$$\bar{P}_n(t) = \sum_{j=0}^{n-1}\left(\sum_{l=1}^{2} c_l(t)\frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} + \frac{2e^{2t}}{3!}\frac{\binom{e^t+n-j-2}{n-j-2}j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}}\right)\bar{P}_j(t), \qquad (n\geq 3), \qquad (6)$$

where $c_1(t) = -4e^t$ and $c_2(t) = 3+2e^t$ and initial conditions $\bar{P}_0(t) = 1$, $\bar{P}_1(t) = 2e^t$, and $\bar{P}_2(t) = e^t+2e^{2t}$. Finally, set $\bar{A}_n^{[r]} = \mathbf{E}(X_n^r)$. Then, by differentiating r times and setting $t = 0$, we obtain

$$(n+1)\bar{A}_n^{[r]} = \sum_{j=0}^{n-1}\left(\sum_{l=0}^{2} c_l \frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}}\right)(j+1)\bar{A}_j^{[r]} + \bar{B}_n^{[r]}, \qquad (n \geq 3), \tag{7}$$

where $c_0 = c_1 = -2$ and $c_2 = 5$, initial conditions $\bar{A}_0^{[r]} = 0$, $\bar{A}_1^{[r]} = 2$, $\bar{A}_2^{[r]} = 1 + 2^{r+1}$, and toll sequence

$$\bar{B}_n^{[r]} = \sum_{k=1}^{r}\binom{r}{k}\sum_{j=0}^{n-1}\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left(\sum_{l=1}^{2} c_l(t)\frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} + \frac{2e^{2t}}{3!}\frac{\binom{e^t+n-j-2}{n-j-2}j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}}\right)\Bigg|_{t=0}(j+1)\bar{A}_j^{[r-k]}.$$

Hence, the underlying recurrence is given by

$$a_n = \sum_{j=0}^{n-1}\left(\sum_{l=0}^{2} c_l \frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}}\right)a_j + b_n, \qquad (n \geq 3) \tag{8}$$

with initial conditions as above (note that in slight difference to the previous sections, this is the recurrence satisfied by the r-th moment of X_n multiplied with $n + 1$). So, we need a transfer theorem for this recurrence. Fortunately, this and more general recurrences were already studied in [3].

Proposition 4. *Consider (8).*

(i) Let $b_n = \mathcal{O}(n^{1-\epsilon})$ with $\epsilon > 0$ suitable small. Then,

$$a_n = cn + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where c is a suitable constant.

(ii) Let $b_n = n\log^{\alpha} n$ with $\alpha \in \{0, 1, \ldots\}$. Then,

$$a_n = \frac{8n\log^{\alpha+1} n}{\alpha+1} + n\mathfrak{Pol}_{\alpha}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small.

(iii) Let $b_n = \mathcal{O}(n\log^{\alpha} n)$ with $\alpha \in \{0, 1, \ldots\}$. Then, $a_n = \mathcal{O}\left(n\log^{\alpha+1} n\right)$.

(iv) Item (iii) holds with $\mathcal{O}$ replaced by o as well.

Proof. See Section 2 in [3]. ∎

Before we use this result to treat (centered and non-centered) moments, we need a technical lemma.

Lemma 1. *We have*

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\binom{e^t+n}{n-1}\Bigg|_{t=0} = \frac{n^2}{2}\log^k n + n^2\mathfrak{Pol}_{k-1}(\log n) + \mathcal{O}\left(n^{2-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small.

Proof. The proof uses induction on k. First for $k = 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\binom{e^t+n}{n-1}\bigg|_{t=0} = \binom{e^t+n}{n-1}\sum_{j=2}^{n}\frac{e^t}{e^t+j}\bigg|_{t=0} = \frac{(n+1)n}{2}\left(H_{n+1}-\frac{3}{2}\right),$$

where $H_n = \sum_{j=1}^{n} 1/j$ is the n-th harmonic number. Hence, the claim follows from the well-known asymptotic expansion $H_n = \log n + \gamma + \mathcal{O}(1/n)$, where γ denotes Euler's constant.

Assume now that the claim holds for all $k' < k$. In order to prove it for k, observe that

$$\begin{aligned}\frac{\mathrm{d}^k}{\mathrm{d}t^k}\binom{e^t+n}{n-1}\bigg|_{t=0} &= \frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}}\left(\binom{e^t+n}{n-1}\sum_{j=2}^{n}\frac{e^t}{e^t+j}\right)\bigg|_{t=0}\\ &= \sum_{i=0}^{k-1}\binom{k-1}{i}\frac{\mathrm{d}^i}{\mathrm{d}t^i}\binom{e^t+n}{n-1}\bigg|_{t=0}\frac{\mathrm{d}^{k-1-i}}{\mathrm{d}t^{k-1-i}}\sum_{j=2}^{n}\frac{e^t}{e^t+j}\bigg|_{t=0}.\end{aligned}$$

For the first derivative inside the sum, we can use the induction hypothesis. For the second derivative, one shows by another induction (left as an exercise) that

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\sum_{j=2}^{n}\frac{e^t}{e^t+j}\bigg|_{t=0} = \log n + c + \mathcal{O}(1/n)$$

for all $k \geq 0$ with a suitable constant c. Plugging this in and doing some straightforward simplification yields the claim. ∎

Corollary 1. *We have,*

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\binom{e^t+n}{n}\bigg|_{t=0} = n\log^k n + n\mathfrak{Pol}_{k-1}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small. Moreover, we have

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}e^{2t}\binom{e^t+n}{n-1}\bigg|_{t=0} = \frac{n^2}{2}\log^k n + n^2\mathfrak{Pol}_{k-1}(\log n) + \mathcal{O}\left(n^{2-\epsilon}\right)$$

and

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}e^{2t}\binom{e^t+n}{n}\bigg|_{t=0} = n\log^k n + n\mathfrak{Pol}_{k-1}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small.

Proof. All of the claims follow similarly. Hence, we just prove the first one. Therefore, note that

$$\begin{aligned}\frac{\mathrm{d}^k}{\mathrm{d}t^k}\binom{e^t+n}{n}\bigg|_{t=0} &= \frac{1}{n}\frac{\mathrm{d}^k}{\mathrm{d}t^k}(e^t+1)\binom{e^t+n}{n-1}\bigg|_{t=0}\\ &= \frac{2}{n}\frac{\mathrm{d}^k}{\mathrm{d}t^k}\binom{e^t+n}{n-1}\bigg|_{t=0} + \frac{1}{n}\sum_{i=1}^{k}\frac{\mathrm{d}^{k-i}}{\mathrm{d}t^{k-i}}\binom{e^t+n}{n-1}\bigg|_{t=0}.\end{aligned}$$

Plugging in the result of the above lemma immediately yields the claim. ∎

Now, we can turn to the mean which satisfies (7) with $r = 1$. Hence, the toll sequence is given by

$$\begin{aligned}\bar{B}_n^{[1]} &= \sum_{j=0}^{n-1} \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{l=1}^{2} c_l(t)\frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} + \frac{2e^{2t}}{3!}\frac{\binom{e^t+n-j-2}{n-j-2}j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}}\right)\Bigg|_{t=0}(j+1)\\ &:= \alpha_n + \beta_n\end{aligned}$$

which we break into the two parts

$$\alpha_n = \sum_{j=0}^{n-1}\frac{\mathrm{d}}{\mathrm{d}t}\sum_{l=1}^{2} c_l(t)\frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}}\Bigg|_{t=0}(j+1)$$

and

$$\beta_n = \sum_{j=0}^{n-1}\frac{\mathrm{d}}{\mathrm{d}t}\frac{2e^{2t}}{3!}\frac{\binom{e^t+n-j-2}{n-j-2}j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}}\Bigg|_{t=0}(j+1)$$

First, for α_n observe that

$$\begin{aligned}\alpha_n &= -4\sum_{j=0}^{n-1}\frac{j(j+1)(n-1-j)}{n(n-1)(n-2)} + 2\sum_{j=0}^{n-1}\frac{(j-1)j(j+1)}{n(n-1)(n-2)}\\ &= n\left(-4\int_0^1 x^2(1-x)\mathrm{d}x + 2\int_0^1 x^3\mathrm{d}x + \mathcal{O}(1/n)\right) = n/6 + \mathcal{O}(1),\end{aligned}$$

where we have used Euler-Maclaurin summation formula. Next, we are going to treat β_n. Here, we apply Corollary 1 and obtain

$$\begin{aligned}\beta_n =& \sum_{j=0}^{n-1}\frac{2j(j+1)}{n(n-1)(n-2)}\left((n-j-2)\log(n-j-2) + c_1(n-j-2) + \mathcal{O}\left((n-j-2)^{1-\epsilon}\right)\right)\\ &- \sum_{j=0}^{n-j}\frac{2(j+1)}{n(n-1)(n-2)}\left(\frac{(n-j-2)^2}{2}\log(n-j-2) + c_2(n-j-2) + \mathcal{O}\left((n-j-2)^{2-\epsilon}\right)\right),\end{aligned}$$

where c_1 and c_2 are suitable constants. By another application of the Euler-Maclaurin summation formula and a trivial estimate for the remainder,

$$\begin{aligned}\beta_n &= \left(2\int_0^1 x^2(1-x)\mathrm{d}x - \int_0^1 x(1-x)^2\mathrm{d}x\right)n\log n + cn + \mathcal{O}\left(n^{1-\epsilon}\right)\\ &= \frac{1}{12}n\log n + n\mathfrak{Pol}_0(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),\end{aligned}$$

where c is a suitable constant and $\epsilon > 0$ is suitable small. Overall,

$$\bar{B}_n^{[1]} = \frac{1}{12}n\log n + n\mathfrak{Pol}_0(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right).$$

Hence, by the transfer theorem

$$\mathbf{E}(X_n) = \bar{A}_n^{[1]} = \frac{1}{3}\log^2 n + \mathfrak{Pol}_1(\log n) + \mathcal{O}\left(1/n^{\epsilon}\right).$$

As before, the next step is to show that a similar expansion more generally holds for all higher moments.

Proposition 5. *For all $r \geq 1$, we have*

$$\bar{A}_n^{[r]} = \mathfrak{Pol}(\log n) + \mathcal{O}\left(1/n^{\epsilon}\right),$$

where $\epsilon > 0$ is suitable small.

Proof. We use induction on r. Note that the claim for $r = 1$ was already proved above. Next, we assume that the claim holds for all $r' < r$. In order to show it for r, we again break the toll sequence of (7) into two parts α_n and β_n, where

$$\alpha_n = \sum_{k=1}^{r} \binom{r}{k} \sum_{j=0}^{n-1} \frac{\mathrm{d}^k}{\mathrm{d}t^k} \sum_{l=1}^{2} c_l(t) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} \Bigg|_{t=0} (j+1)\bar{A}_j^{[r-k]}$$

and

$$\beta_n = \sum_{k=1}^{r} \binom{r}{k} \sum_{j=0}^{n-1} \frac{\mathrm{d}^k}{\mathrm{d}t^k} \frac{2e^{2t}}{3!} \frac{\binom{e^t+n-j-2}{n-j-2} j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}} \Bigg|_{t=0} (j+1)\bar{A}_j^{[r-k]}$$

Now, α_n and β_n are treated with exactly the same ideas as for the mean value above. For instance, when plugging the induction hypothesis into α_n one obtains sums such as

$$\sum_{j=1}^{n-1} \frac{j(j+1)(n-1-j)}{n(n-1)(n-2)} \left(\mathfrak{Pol}(\log j) + \mathcal{O}\left(1/j^{\epsilon}\right)\right)$$

which due to Euler-Maclaurin summation formula yield $n\mathfrak{Pol}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right)$. Hence, $\alpha_n = n\mathfrak{Pol}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right)$. Similarly, by plugging the induction hypothesis into β_n and using Euler-Maclaurin summation formula and Corollary 1 one obtains that $\beta_n = n\mathfrak{Pol}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right)$. Overall,

$$\bar{B}_n^{[r]} = n\mathfrak{Pol}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right).$$

Applying the transform theorem concludes the induction. ∎

Next, we turn to central moments. As in the previous section, the above proposition implies that the central moments are of the same general shape. So, what is left is again to derive a more detailed information of the leading term.

First, we need the recurrence of the central moments. Therefore, set $P_n(t) = (n+1)\mathbf{E}(e^{t(X_n - \mathbf{E}(X_n))})$. Then, from (6), we obtain for $n \geq 3$,

$$P_n(t) = \sum_{j=0}^{n-1} \left(\sum_{l=1}^{2} c_l(t) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} + \frac{2e^{2t}}{3!} \frac{\binom{e^t+n-j-2}{n-j-2} j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}} \right) e^{t(\mathbf{E}(X_j) - \mathbf{E}(X_n))} P_j(t)$$

with initial conditions $P_0(t) = 1, P_1(t) = 2e^{-t}$, and $P_2(t) = e^{-4t} + 2e^{-3t}$. Next, set $A_n^{[r]} = \mathbf{E}(X_n - \mathbf{E}(X_n))^r$. Taking derivatives r times and setting $t = 0$ yields

$$(n+1)A_n^{[r]} = \sum_{j=0}^{n-1} \left(\sum_{l=0}^{2} c_l \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} \right) (j+1)\bar{A}_j^{[r]} + B_n^{[r]}, \qquad (n \geq 3) \tag{9}$$

with initial conditions $A_0^{[r]} = 0, A_1^{[r]} = 2(-1)^r, A_2^{[r]} = (-1)^r(4^r + 2 \cdot 3^r)$ and toll sequence

$$B_n^{[r]} = \sum_{\substack{i_1+i_2+i_3=r \\ i_3 \neq r}} \binom{r}{i_1, i_2, i_3} \sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \left(\sum_{l=1}^{2} c_l(t) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} + \frac{2e^{2t}}{3!} \frac{\binom{e^t+n-j-2}{n-j-2} j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}} \right) \Bigg|_{t=0} \left(\mathbf{E}(X_j) - \mathbf{E}(X_n)\right)^{i_2} (j+1)A_j^{[i_3]}.$$

Let us again first look at the variance. Hence, we have to choose $r = 2$ in (9). As for the mean we break the toll sequence into two parts α_n and β_n, where

$$\alpha_n = \sum_{\substack{i_1+i_2+i_3=2\\ i_3\neq 2}} \binom{2}{i_1, i_2, i_3} \underbrace{\sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \sum_{l=1}^{2} c_l(t) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} \bigg|_{t=0} (\mathbf{E}(X_j) - \mathbf{E}(X_n))^{i_2} (k+1) A_j^{[i_3]}}_{:=P^{\alpha_n}_{i_1,i_2,i_3}}$$

and

$$\beta_n = \sum_{\substack{i_1+i_2+i_3=2\\ i_3\neq 2}} \binom{2}{i_1, i_2, i_3} \underbrace{\sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \frac{2e^{2t}}{3!} \frac{\binom{e^t+n-j-2}{n-j-2} j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}} \bigg|_{t=0} (\mathbf{E}(X_j) - \mathbf{E}(X_n))^{i_2} (j+1) A_j^{[i_3]}}_{:=P^{\beta_n}_{i_1,i_2,i_3}}.$$

For α_n, we first consider $i_2 = 2$ and $i_3 = 0$,

$$\begin{aligned} P^{\alpha_n}_{0,2,0} &\sim \sum_{j=0}^{n-1} \sum_{l=1}^{2} c_l(0) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} \left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n \right)^2 (j+1) \\ &\sim \frac{4}{9} \left(-4 \int_0^1 x^2(1-x)\log^2 x \mathrm{d}x + 5 \int_0^1 x^3 \log^2 x \mathrm{d}x \right) n \log^2 n = -\frac{13}{1944} n \log^2 n. \end{aligned}$$

For the other terms in α_n, we can use similar ideas to obtain the bound $\mathcal{O}(n \log n)$. Hence,

$$\alpha_n \sim -\frac{13}{1944} n \log^2 n.$$

Next, we are going to treat β_n. Here, we first consider $i_3 = 0$. Then, after a similar computation as for α_n,

$$\begin{aligned} &\sum_{i_1+i_2=2} \binom{2}{i_1, i_2} P^{\beta_n}_{i_1,i_2,0} \\ &\quad\sim \sum_{i_1+i_2=2} \binom{2}{i_1, i_2} \sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \frac{2e^{2t}}{3!} \frac{\binom{e^t+n-j-2}{n-j-2} j - \binom{e^t+n-j-2}{n-j-3}}{\binom{n}{3}} \bigg|_{t=0} \left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n \right)^{i_2} (j+1) \\ &\quad\sim -\frac{41}{1944} n \log^2 n - \frac{1}{108} n \log^2 n + \frac{1}{12} n \log^2 n = \frac{103}{1944} n \log^2 n. \end{aligned}$$

The other terms in β_n are easily shown to contribute only $\mathcal{O}(n \log n)$. Hence,

$$\beta_n \sim \frac{103}{1944} n \log^2 n.$$

Overall, we obtain for the toll sequence

$$B_n^{[2]} \sim -\frac{13}{1944} n \log^2 n + \frac{103}{1944} n \log^2 n = \frac{5}{108} n \log^2 n.$$

Using our transfer theorem yields

$$\mathrm{Var}(X_n) \sim \frac{10}{81} \log^3 n.$$

Now, the final step is to generalize these arguments to all central moments.

Proposition 6. *For all $m \geq 0$, we have*

$$A_n^{[2m]} \sim \frac{(2m)!}{2^m m!}\left(\frac{10}{81}\right)^m \log^{3m} n \quad \textit{and} \quad A_n^{[2m+1]} = \mathcal{O}\left(\log^{3m+1} n\right).$$

Proof. We are going to use induction on m. Note that the claim holds for $m = 0$. Next assume that the claim holds for all $m' < m$. We will show that it holds for m as well.

First, let us consider the even case. Then, as for the variance, we are going the break the toll sequence of (9) into two parts α_n and β_n, where

$$\alpha_n = \sum_{\substack{i_1+i_2+i_3=2m\\ i_3\neq 2m}} \binom{2m}{i_1, i_2, i_3} \underbrace{\sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \sum_{l=1}^{2} c_l(t) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}}\Bigg|_{t=0} (\mathbf{E}(X_j) - \mathbf{E}(X_n))^{i_2} (j+1) A_j^{[i_3]}}_{:=T^{\alpha_n}_{i_1,i_2,i_3}}$$

and

$$\beta_n = \sum_{\substack{i_1+i_2+i_3=2m\\ i_3\neq 2m}} \binom{2m}{i_1, i_2, i_3} \underbrace{\sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \frac{2e^t}{3!} \frac{\left(\binom{e^t+n-j-2}{n-j-2} j - \binom{e^t+n-j-2}{n-j-3}\right)}{\binom{n}{3}}\Bigg|_{t=0} (\mathbf{E}(X_j) - \mathbf{E}(X_n))^{i_2} (j+1) A_j^{[i_3]}}_{:=T^{\beta_n}_{i_1,i_2,i_3}}.$$

We first treat α_n which we again break into two parts $x_n^{[\alpha]}$ and $y_n^{[\alpha]}$ according to whether i_3 is even or not, i.e.,

$$x_n^{[\alpha]} = \sum_{\substack{i_1+i_2+2i_3=2m\\ i_3\neq m}} \binom{2m}{i_1, i_2, 2i_3} T^{\alpha_n}_{i_1,i_2,2i_3}$$

and

$$y_n^{[\alpha]} = \sum_{i_1+i_2+2i_3+1=2m} \binom{2m}{i_1, i_2, 2i_3+1} T^{\alpha_n}_{i_1,i_2,2i_3+1}.$$

As for $x_n^{[\alpha]}$, plugging in the expansion for the mean and the induction hypothesis gives

$$T^{\alpha_n}_{i_1,i_2,2i_3} \sim \sum_{j=0}^{n-1} \frac{\mathrm{d}^{i_1}}{\mathrm{d}t^{i_1}} \sum_{l=1}^{2} c_l(t) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}}\Bigg|_{t=0} \left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n\right)^{i_2} (j+1) \frac{(2i_3)!}{2^{i_3}(i_3)!}\left(\frac{10}{81}\log^3 j\right)^{i_3}.$$

Using ideas as in the proof of Proposition 5, we obtain $T^{\alpha_n}_{i_1,i_2,2i_3} \sim c_{i_1,i_2,i_3} n \log^{i_2+3i_3} n$, where c_{i_1,i_2,i_3} are suitable constants. Hence, choosing $i_1 = 0, i_2 = 2$, and $i_3 = m-1$ gives the main contribution. So,

$$\begin{aligned} x_n^{[\alpha]} &\sim \binom{2m}{2, 2m-2} T^{\alpha_n}_{0,2,2m-2} \\ &\sim \binom{2m}{2, 2m-2} \sum_{j=0}^{n-1}\sum_{l=1}^{2} c_l(0) \frac{l!}{3!} \frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}} \left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n\right)^2 (j+1) \frac{(2i_3)!}{2^{i_3}(i_3)!}\left(\frac{10}{81}\log^3 j\right)^{m-1} \\ &\sim -\frac{13}{1944}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n. \end{aligned}$$

Next, for $y_n^{[\alpha]}$, we first consider the term with $i_1 = 0, i_2 = 1$ and $i_3 = m-1$. Note that by Proposition 5 we know that $A_j^{[2m-1]}$ is a polynomial in $\log n$. Therefore, by induction hypothesis, we know that $A_j^{[2m-1]} \sim c\log^{3m-2} j$ with a suitable constant c. Consequently,

$$\begin{aligned} T_{0,1,2m-1}^{\alpha_n} &\sim \sum_{j=0}^{n-1}\sum_{l=1}^{2} c_l(0)\frac{l!}{3!}\frac{\binom{j}{l}\binom{n-1-j}{2-l}}{\binom{n}{3}}\left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n\right)(j+1)c\log^{3m-2} j \\ &\sim -\frac{17c}{216} n\log^{3m-1} n. \end{aligned}$$

For $i_3 < m-1$, we can use similar ideas to show that $T_{i_1,i_2,i_3}^{\alpha_n} = o\left(n\log^{3m-\frac{3}{2}} n\right)$. Therefore,

$$y_n^{[\alpha]} = 2mT_{0,1,2m-1}^{\alpha_n} + o\left(n\log^{3m-\frac{3}{2}} n\right) \sim -\frac{34cm}{216} n\log^{3m-1} n.$$

Overall, we obtain for the contribution of α_n,

$$\alpha_n = x_n^{[\alpha]} + y_n^{[\alpha]} \sim -\frac{13}{1944}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n - \frac{34cm}{216} n\log^{3m-1} n.$$

Next, we turn to β_n which is handled in exactly the same manner. So, we again break it into two parts $x_n^{[\beta]}$ and $y_n^{[\beta]}$ according to whether i_3 is even or not. Consequently,

$$x_n^{[\beta]} = \sum_{\substack{i_1+i_2+2i_3=2m \\ i_3\neq m}} \binom{2m}{i_1, i_2, 2i_3} T_{i_1,i_2,2i_3}^{\beta_n}$$

and

$$y_n^{[\beta]} = \sum_{i_1+i_2+2i_3+1=2m} \binom{2m}{i_1, i_2, 2i_3} T_{i_1,i_2,2i_3+1}^{\beta_n}.$$

As for $x_n^{[\beta]}$ the same reasoning as above shows that $T_{i_1,i_2,2i_3}^{\beta_n} \sim c_{i_1,i_2,i_3} n\log^{i_1+i_2+3i_3} n$, where c_{i_1,i_2,i_3} are suitable constants. Hence, $i_3 = m-1$ gives the main contribution. Consequently, by Corollary 1 and the induction hypothesis,

$$\begin{aligned} x_n^{[\beta]} &\sim \sum_{i_1+i_2=2}\binom{2m}{i_1, i_2} T_{i_1,i_2,2m-2}^{\beta_n} \\ &\sim 2\sum_{i_1+i_2=2}\binom{2m}{i_1, i_2}\sum_{j=0}^{n-1}\left((n-j-2)j - \frac{(n-j-2)^2}{2}\right)\log^{i_1}(n-j-2) \\ &\qquad \left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n\right)^{i_2}(j+1)\frac{(2m-2)!}{2^{m-1}(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-3} n \\ &\sim \left(-\frac{41}{1944}\binom{2m}{0,2} - \frac{1}{216}\binom{2m}{1,1} + \frac{1}{12}\binom{2m}{2,0}\right)\frac{(2m-2)!}{2^{m-1}(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n \\ &= \frac{103}{1944}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n. \end{aligned}$$

For $y_n^{[\beta]}$, we again start by looking at $i_3 = m-1$ for which due to Proposition 5 we have $A_j^{[2m-1]} \sim c\log^{3m-2} j$ with a suitable constant c. Hence, as above

$$\sum_{i_1+i_2=1} T^{\beta_n}_{i_1,i_2,2m-2} \sim \sum_{i_1+i_2=1} \binom{2m}{i_1,i_2} \sum_{j=0}^{n-1} \left((n-j-2)j - \frac{(n-j-2)^2}{2} \right)$$
$$\log^{i_1}(n-j-2) \left(\frac{1}{3}\log^2 j - \frac{1}{3}\log^2 n \right)^{i_2} (j+1)c\log^{3m-3} n$$
$$\sim \frac{34cm}{216} n\log^{3m-1} n.$$

Similar arguments for $i < m-1$ show that $T^{\beta_n}_{i_1,i_2,i_3} = o\left(n\log^{3m-\frac{5}{2}} n\right)$. Therefore,

$$y_n^{[\beta]} \sim \frac{34cm}{216} n\log^{3m-1} n.$$

Overall, we have

$$\beta_n \sim \frac{103}{1944}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n + \frac{34cm}{216} n\log^{3m-1} n.$$

Now, collecting everything gives

$$\begin{aligned} B_n^{[2m]} &= \alpha_n + \beta_n \\ &\sim -\frac{13}{1944}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n - \frac{34cm}{216} n\log^{3m-1} n \\ &\qquad + \frac{103}{1944}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n + \frac{34cm}{216} n\log^{3m-1} n \\ &= \frac{5}{108}\frac{(2m)!}{2^m(m-1)!}\left(\frac{10}{81}\right)^{m-1} n\log^{3m-1} n. \end{aligned}$$

and using the transfer theorem concludes the proof in the even case.

Next, we briefly sketch the odd case which can be treated with the same ideas as the even case. Again, we break the toll sequence into two parts α_n and β_n which are defined as above (with the only difference that $2m$ is replaced by $2m+1$). Then, as above, one shows that

$$\alpha_n \sim -(2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!} n\log^{3m+1} n$$

and

$$\beta_n \sim (2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!} n\log^{3m+1} n.$$

Hence,

$$\begin{aligned} B_n^{[2m+1]} &= \alpha_n + \beta_n \\ &\sim -(2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!} n\log^{3m+1} n + (2m+1)\frac{17}{216}\frac{(2m)!}{2^m m!} n\log^{3m+1} n \\ &= o\left(n(\log n)^{3m+1}\right). \end{aligned}$$

Using the transfer theorem shows that $A_n^{[2m+1]} = o\left(n(\log n)^{3m+1}\right)$ and due to the remark in the paragraph succeeding Proposition 5 the claim is established. ∎

Finally, by the Fréchet-Shohat theorem, the last proposition implies the following theorem.

Theorem 3. *As $n \to \infty$, we have*

$$\frac{X_n - \log^2 n/3}{\sqrt{10\log^3 n/81}} \xrightarrow{d} \mathcal{N}(0,1).$$

Another example which is very similar to the one above is the depth of a random node in a random priority tree of size n; see [14]. Here, the underlying recurrence is as above. Hence, one can again use the transfer theorem to derive the central limit theorem. Since the details are straightforward, we leave them as an exercise for the reader.

4 Further Examples

In this final section, we will briefly sketch some further examples. It should by now be clear that our approach essentially rests on the transfer theorem. Once such a result is established, the remaining proof is rather automatic.

Number of Key Comparisons for Insertion and Depth in Binary Search Trees. These examples are similar but more easier than the examples discussed in the previous section. For instance, let X_n denote the number of key comparisons when inserting a random node in a random binary search tree build from n records (this quantity is also called "unsuccessful search"; see Chapter 2 in [11] for background). Then, for $n \geq 1$,

$$X_n|\,(I_n = j) \stackrel{d}{=} \begin{cases} X_j + 1, & \text{with probability } (j+1)/(n+1), \\ X_{n-1-j} + 1, & \text{with probability } (n-1-j)/(n+1) \end{cases}$$

with $P(I_n = j) = 1/n, 0 \leq j < n$ and $X_0 = 0$. From this, a straightforward computation reveals that the underlying recurrence (with a scaling factor $n+1$ as in the previous section) is given by

$$a_n = \frac{2}{n}\sum_{j=0}^{n-1} a_j + b_n, \qquad (n \geq 1) \tag{10}$$

with $a_0 = 0$. A transfer theorem for this recurrence of similar type as in the previous section is easily derived and can be found in [8].

Proposition 7. *Consider (10).*

(i) Let $b_n = \mathcal{O}\left(n^{1-\epsilon}\right)$ with $\epsilon > 0$ suitable small. Then,

$$a_n = cn + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where c is a suitable constant.

(ii) Let $b_n = n\log^\alpha n$ with $\alpha \in \{0, 1, \ldots\}$. Then,

$$a_n = \frac{2n\log^{\alpha+1} n}{\alpha+1} + n\mathfrak{Pol}_\alpha(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small.

(iii) Let $b_n = \mathcal{O}\left(n\log^\alpha n\right)$ with $\alpha \in \{0, 1, \ldots\}$. Then, $a_n = \mathcal{O}\left(n\log^{\alpha+1} n\right)$.

(iv) Item (iii) holds with $\mathcal{O}$ replaced by o as well.

Hence, our approach applies as in the last section (the technical details being easier) and we obtain the following theorem.

Theorem 4. *As $n \to \infty$, we have*

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

Similarly, the depth of a random node satisfies almost the same distributional recurrence (again see Chapter 2 in [11] for background). Hence, again a central limit theorem follows from the above transfer theorem by applying our approach.

Depth of Variants of Binary Search Trees. The previous example of the depth can be extended to several extensions of binary search trees. Here, we are going to discuss three of them, namely, median-of-$(2t+1)$ binary search trees (see [3]), m-ary search trees (see [2]), and quadtrees (see [1]). Subsequently, let X_n denote the depth of a randomly chosen record in the random tree build from n records. Moreover, the underlying recurrence will be satisfied by all centered and non-centered moments multiplied by n.

First, for median-of-$(2t+1)$ binary search trees, X_n satisfies the distributional recurrence for $n \geq 2t+1$

$$X_n|\,(I_n = j) \stackrel{d}{=} \begin{cases} X_j + 1, & \text{with probability } j/n, \\ X_{n-1-j} + 1, & \text{with probability } (n-1-j)/n, \\ 0, & \text{with probability } 1/n \end{cases}$$

with $P(I_n = j) = \binom{j}{t}\binom{n-1-j}{t}/\binom{n}{2t+1}, 0 \leq j < n$ and suitable initial conditions. Hence, the underlying recurrence is given by

$$a_n = \frac{2}{\binom{n}{2t+1}} \sum_{j=0}^{n-1} \binom{j}{t}\binom{n-1-j}{t} a_j + b_n, \qquad (n \geq 2t+1) \tag{11}$$

with suitable initial conditions. This recurrence was extensively studied in [3]. In particular, the following transfer theorem can be proved with the tools of the latter paper.

Proposition 8. *Consider (11).*

(i) Let $b_n = \mathcal{O}\left(n^{1-\epsilon}\right)$ with $\epsilon > 0$ suitable small. Then,

$$a_n = cn + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where c is a suitable constant.

(ii) Let $b_n = n\log^{\alpha} n$ with $\alpha \in \{0, 1, \ldots\}$. Then,

$$a_n = \frac{n\log^{\alpha+1} n}{(H_{2t+2} - H_{t+1})(\alpha+1)} + n\mathfrak{Pol}_{\alpha}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small and $H_n = \sum_{j=1}^{n} 1/j$ denotes the n-th harmonic number.

(iii) Let $b_n = \mathcal{O}\left(n\log^{\alpha} n\right)$ with $\alpha \in \{0, 1, \ldots\}$. Then, $a_n = \mathcal{O}\left(n\log^{\alpha+1} n\right)$.

(iv) Item (iii) holds with $\mathcal{O}$ replaced by o as well.

Hence, our approach applies and yields the following result (see [4] for a different approach).

Theorem 5. *As $n \to \infty$, we have*

$$\frac{X_n - \log n/(H_{2t+2} - H_{t+1})}{\sqrt{(H_{2t+2}^{(2)} - H_{t+1}^{(2)}) \log n/(H_{2t+2} - H_{t+1})^3}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $H_n^{(2)} = \sum_{j=1}^{n} 1/j^2$.

Next, for the m-ary search tree, we have for $n \geq m-1$

$$X_n | \left(I_n^{[1]} = j_1, \dots, I_n^{[m]} = j_m\right) \stackrel{d}{=} \begin{cases} X_{j_1} + 1, & \text{with probability } j_1/n, \\ & \vdots \\ X_{j_m} + 1, & \text{with probability } j_m/n, \\ 0, & \text{with probability } (m-1)/n \end{cases}$$

with $P(I_n^{[1]} = j_1, \dots, I_n^{[m]} = j_m) = 1/\binom{n}{m-1}, j_1, \dots, j_m \geq 0, j_1 + \ldots + j_m = n - m + 1$ and $X_0 = \cdots = X_{m-2} = 0$. The underlying recurrence is given by

$$a_n = \frac{m}{\binom{n}{m-1}} \sum_{j=0}^{n-m+1} \binom{n-1-j}{m-2} a_j + b_n, \qquad (n \geq m-1) \tag{12}$$

with $a_0 = \cdots = a_{m-2} = 0$. Also, this recurrence was already investigated before and transfer theorems can be found in [2] and [5].

Proposition 9. *Consider (12).*

(i) Let $b_n = \mathcal{O}\left(n^{1-\epsilon}\right)$ with $\epsilon > 0$ suitable small. Then,

$$a_n = cn + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where c is a suitable constant.

(ii) Let $b_n = n \log^{\alpha} n$ with $\alpha \in \{0, 1, \dots\}$. Then,

$$a_n = \frac{n \log^{\alpha+1} n}{(H_m - 1)(\alpha + 1)} + n\mathfrak{Pol}_{\alpha}(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small and $H_n = \sum_{j=1}^{n} 1/j$ denotes the n-th harmonic number.

(iii) Let $b_n = \mathcal{O}\left(n \log^{\alpha} n\right)$ with $\alpha \in \{0, 1, \dots\}$. Then, $a_n = \mathcal{O}\left(n \log^{\alpha+1} n\right)$.

(iv) Item (iii) holds with $\mathcal{O}$ replaced by o as well.

Then, again by our approach, the following result can be proved (see also [4] and [12] for different approaches).

Theorem 6. *As $n \to \infty$, we have*

$$\frac{X_n - \log n/(H_m - 1)}{\sqrt{(H_m^{(2)} - 1)\log n/(H_m - 1)^3}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $H_n^{(2)} = \sum_{j=1}^{n} 1/j^2$.

Finally, for d-dimensional quadtrees, we have for $n \geq 1$

$$X_n \big| \left(I_n^{[1]} = j_1, \ldots, I_n^{[2^d]} = j_{2^d}\right) \stackrel{d}{=} \begin{cases} X_{j_1} + 1, & \text{with probability } j_1/n, \\ & \vdots \\ X_{j_{2^d}} + 1, & \text{with probability } j_{2^d}/n, \\ 0, & \text{with probability } 1/n \end{cases}$$

with $X_0 = 0$ and

$$P(I_n^{[1]} = j_1, \ldots, I_n^{[2d]} = j_{2^d}) = \binom{n-1}{j_1, \ldots, j_{2^d}} \int_{[0,1]^d} q_1(\mathbf{x})^{j_1} \cdots q_{2^d}(\mathbf{x})^{j_2^d} \mathrm{d}\mathbf{x},$$

where $j_1, \ldots, j_{2^d} \geq 0, j_1 + \cdots + j_{2^d} = n-1, \mathbf{x} = (x_1, \ldots, x_d)$ and

$$q_h(\mathbf{x}) = \prod_{i=1}^{d} \left((1-b_i)x_i + b_i x_i\right), \qquad (1 \leq h \leq 2^d)$$

with $(b_1, \ldots, b_d)_2$ the binary representation of $h-1$. From this, we obtain for the underlying recurrence

$$a_n = 2^d \sum_{j=0}^{n-1} \pi_{n,j} a_j + b_n, \qquad (n \geq 1) \tag{13}$$

with $a_0 = 0$ and

$$\pi_{n,j} = \binom{n-1}{j} \int_{[0,1]^d} (x_1 \cdots x_d)^j (1 - x_1 \cdots x_d)^{n-1-j} \, \mathrm{d}\mathbf{x}.$$

This recurrence was studied in [1]. The following transfer theorem can be proved with tools from the latter paper.

Proposition 10. *Consider (13).*

(i) Let $b_n = \mathcal{O}\left(n^{1-\epsilon}\right)$ with $\epsilon > 0$ suitable small. Then,

$$a_n = cn + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where c is a suitable constant.

(ii) Let $b_n = n \log^\alpha n$ with $\alpha \in \{0, 1, \ldots\}$. Then,

$$a_n = \frac{2n \log^{\alpha+1} n}{d(\alpha+1)} + n\mathfrak{Pol}_\alpha(\log n) + \mathcal{O}\left(n^{1-\epsilon}\right),$$

where $\epsilon > 0$ is suitable small and $H_n = \sum_{j=1}^{n} 1/j$ denotes the n-th harmonic number.

(iii) *Let* $b_n = \mathcal{O}(n \log^\alpha n)$ *with* $\alpha \in \{0, 1, \ldots\}$*. Then,* $a_n = \mathcal{O}\left(n \log^{\alpha+1} n\right)$*.*

(iv) *Item (iii) holds with* $\mathcal{O}$ *replaced by* o *as well.*

Using our approach then gives the following result (see also [4] and [6] for different approaches).

Theorem 7. *As* $n \to \infty$*, we have*

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

Number of Collisions in the $\beta(2, b)$-Coalescent. This is an example from coalescent theory (see [10] for background). Let X_n be a sequence of random variables satisfying

$$X_n \stackrel{d}{=} X_{n-I_n} + 1. \qquad (n \geq 2)$$

with $X_1 = 0$ and $(I_n)_{n\geq1}$ independent of $(X_n)_{n\geq1}$ with distribution

$$\pi_{n,j} = P(I_n = j) = \frac{\Gamma(n-j+b-1)\Gamma(n+1)}{(j+1)\Gamma(n-j)\Gamma(n+b)H(n,b)} \qquad (1 \leq j \leq n-1),$$

where $b > 0$ and

$$H(n, b) = \frac{b}{b+n-1} + \Psi(b+n-1) - \Psi(b) - 1.$$

The authors of [10] asked for a proof of their main result (a central limit theorem for X_n suitable centralized and normalized) directly from the above recurrence. Indeed, our approach is able to solve this problem once a suitable transfer theorem for the underlying recurrence is proved. Therefore, note that the underlying recurrence (without a scaling factor) is given by

$$a_n = \sum_{j=1}^{n-1} \pi_{n,j} a_{n-j} + b_n, \qquad (n \geq 2), \tag{14}$$

where $a_1 = 0$. Unfortunately, due to the more complicated nature of $\pi_{n,j}$ this recurrence is more involved. In particular, we have not been able to prove an analogous result to part (i) of the transfer results above. However, we strongly conjecture that the following claim holds true.

Conjecture 1. *Consider (14). Let* $b_n = \mathcal{O}(1/n^\epsilon)$ *with* $\epsilon > 0$ *suitable small. Then,*

$$a_n = c + \mathcal{O}(1/n^\epsilon),$$

where c *is a suitable constant.*

As before, apart from this property, we need a couple of other transfer properties. However, once this conjecture is established, the other properties can be deduced from it.

Proposition 11. *Assume that the above conjecture holds.*

(i) *Let* $b_n = \log^\alpha n$ *with* $\alpha \in \{-1, 0, 1, \ldots\}$*. Then,*

$$a_n = \frac{\log^{\alpha+2} n}{(\alpha+2)m_1} + n\mathfrak{Pol}_{\alpha+1}(\log n) + \mathcal{O}(1/n^\epsilon),$$

where $\epsilon > 0$ *is suitable small and* $m_1 = \zeta(2, b)$ *with* $\zeta(z, b)$ *the Hurwitz zeta function.*

(ii) *Let* $b_n = \mathcal{O}(\log^\alpha n)$ *with* $\alpha \in \{-1, 0, 1, \ldots\}$. *Then,* $a_n = \mathcal{O}\left(\log^{\alpha+2} n\right)$.

(iii) *Item (ii) holds with* $\mathcal{O}$ *replaced by* o *as well.*

Proof. All these properties follow from the conjecture by using similar ideas as in [10]. ∎

Finally, by applying our approach, we obtain the following result.

Theorem 8. *As* $n \to \infty$, *we have*

$$\frac{X_n - \log^2 n/(2m_1)}{\sqrt{m_2 \log^3 n/(3m_1^3)}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $m_2 = 2\zeta(3, b)$.

References

[1] H.-H. Chern, M. Fuchs, H.-K. Hwang (2007). Phase changes in random point quadtrees, *ACM Transactions on Algorithms*, **3**, 51 pages.

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

[3] H.-H. Chern, H.-K. Hwang, T.-H. Tsai (2002). An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms, *Journal of Algorithms*, **44**, 177-225.

[4] L. Devroy. Universal limit laws for depths in random trees, *SIAM Journal on Computing*, **28**, 409-432.

[5] J. A. Fill and N. Kapur. Transfer theorems and asymptotic distributional results for m-ary search trees, *Random Structures and Algorithms*, **26**, 359-391.

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

[7] H.-K. Hwang (2004). Phase changes in random recursive structures and algorithms (a brief survey), In "Proceedings of the Workshop on Probability with Applications to Finance and Insurance", World Scientific, 82V97.

[8] H.-K. Hwang and R. Neininger (2002). Phase change of limit laws in the quicksort recurrences under varying toll functions, *SIAM Journal on Computing*, **31**, 1687-1722.

[9] M. Kuba and A. Panholzer (2007). Analysis of insertion costs in priority trees, in "Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithmics and Combinatorics", *SIAM Philadelphia*, 175-182.

[10] A. Iksanov, A. Marynych, M. Möhle (2008). On the number of collisions in beta$(2, b)$-coalesents, to appear.

[11] H. M. Mahmoud (1992). Evolution of Random Search Trees, Wiley, New York.

[12] H. M. Mahmoud and B. Pittel (1988). On the joint distribution of the insertion path length and the number of comparisons in search trees, *Discrete Applied Mathematics*, **20**, 243-251.

[13] R. Neininger and L. Rüschendorf (2004). On the contraction method with degenerate limit equation, *The Annals of Probability*, **32**, 2838-2856.

[14] A. Panholzer (2008). Analysis for some parameters for random nodes in priority trees, *Discrete Mathematics and Theoretical Computer Science*, **10**, 1-38.

[15] A. Panholzer and H. Prodinger (1998). Average case analysis priority trees: a structure for priority queue administration, *Algorithmica*, **22**, 600-630.

Participation in conferences within NSC 97-2628-M-009-008

by

Michael Fuchs

This is a short report concerning participation in international conferences within my national science counsel project NSC 97-2628-M-009-008.

I participated in the 9th International Conference on Finite Fields and Their Applications, Dublin, Ireland, July 13-19, 2009. The conference was to honor the 65th birthday of Prof. Harald Niederreiter (retired from National University of Singapore) who was the supervisor of my master thesis. Participation was only possible after submitting and acceptance of an abstract of the proposed talk. The abstract of my talk which took place on July 16th, 2009 is enclosed.

After my talk, I had a couple of interesting discussions with Alain Lasjaunias (University of Bordeaux) and we started to work on an interesting problem concerning continued fractions in the field of formal Laurent series over a finite field. We have already achieved some partial results and this might lead to a future research paper.

Metric Diophantine Approximation for Formal Laurent Series over Finite Fields

Michael Fuchs

National Chiao Tung University

Let $\mathbf{F}_q((T^{-1}))$ be the field of formal Laurent series endowed with the valuation $|\cdot|$ induced by the degree function. Consider the set

$$\mathbf{L} = \{f \in \mathbf{F}_q((T^{-1})) : \ |f| < 1\}$$

together with the Haar probability measure. Several recent studies investigated the diophantine approximation problem

$$\left| f - \frac{P}{Q} \right| < \frac{1}{q^{2n+l_n}}, \quad \deg Q = n, \quad \gcd(P,Q) = 1, \tag{1}$$

where $f \in \mathbf{L}$ and l_n is a sequence of non-negative integers.

For instance, in [1] a strong law of large numbers with error term for the number of pairs (P,Q) with (1) with $\deg Q \leq N$ was proved. Moreover, in [2] a similar result for (1) without the comprimeness assumption was established, however, under further assumptions on l_n and without an error term.

In this talk, we will discuss improvements of these results as well as generalizations to inhomogeneous Diophantine approximation, restricted Diophantine approximation, and simultaneous Diophantine approximation. A typical result which improves the main result in [2] reads as follows:

Theorem *The number of pairs (P,Q) satisfying (1) without the coprimeness condition and $\deg Q \leq N$ is almost surely given by*

$$\Psi(N) + \mathcal{O}\left((\Psi(N))^{1/2}(\log \Psi(N))^{2+\epsilon}\right),$$

where $\epsilon > 0$ and $\Psi(N) = \sum_{n \leq N} q^{-l_n}$.

References

[1] K. Inoue and H. Nakada (2003). On metric diophantine approximation in positive characteristic, *Acta Arith.*, **110**, 215-218.

[2] H. Nakada and R. Natsui (2006). Asymptotic behavior of the number of solutions for non-Archimedean Diophantine approximations, *Acta Arith.*, **125**, 203-214.