J. Appl. Prob. **43,** 377–390 (2006)
Printed in Israel
© *Applied Probability Trust* 2006

LIMIT DISTRIBUTION OF DISTANCES IN BIASED RANDOM TRIES

RAFIK AGUECH,* *Faculté des Sciences de Monastir, Tunisia*
NABIL LASMAR,** *IPEIT, Tunisia*
HOSAM MAHMOUD,*** *The George Washington University*

Abstract

The *trie* is a sort of digital tree. Ideally, to achieve balance, the trie should grow from an unbiased source generating keys of bits with equal likelihoods. In practice, the lack of bias is not always guaranteed. We investigate the distance between randomly selected pairs of nodes among the keys in a biased trie. This research complements that of Christophi and Mahmoud (2005); however, the results and some of the methodology are strikingly different. Analytical techniques are still useful for moments calculation. Both mean and variance are of polynomial order. It is demonstrated that the standardized distance approaches a normal limiting random variable. This is proved by the contraction method, whereby the limit distribution is shown to approach the fixed-point solution of a distributional equation in the Wasserstein metric space.

Keywords: Random tree; recurrence; Mellin transform; poissonization; fixed point; contraction method

2000 Mathematics Subject Classification: Primary 05C05; 60C05
Secondary 60F05; 68P05; 68P10; 68P20

1. Introduction

We study the distances between distinct pairs of keys in random binary tries (a sort of digital tree). The mean and variance are derived by analytical methods involving the use of poissonization, as a mathematical transform, and depoissonization, as an asymptotic inverse transform. Although the chief interest lies in studying the random tree for a fixed population of n keys, the recurrence equations involved are rather unwieldy. If a Poisson-distributed number of keys is assumed instead, the functional equations involved can asymptotically be solved by the Mellin transform and its inverse. It can then be justified that the solution is a good approximation (with quantifiable small errors) for the fixed-population problem when the Poisson parameter, taken to be n, tends to infinity. The complexity of such a Mellin approach increases considerably for higher moments, which prompts us to consider a shortcut approach via the contraction method to determine the limit distribution.

The standard data model for tries is an unbiased Bernoulli probability distribution. What will become of these results if the Bernoulli model is biased? This question is of prime practical value, because our assumption about a perfect key generator may not be totally realistic, owing

Received 1 July 2004; revision received 8 March 2006.
* Postal address: Département de Mathématiques, Faculté des Sciences de Monastir, 5019 Monastir, Tunisia.
Email address: rafikaguech@ipeit.rnu.tn
** Postal address: Département de Mathématiques, IPEIT, 2 rue Jawaher Lel Nehru 1008 Montfleury, Tunis, Tunisia.
Email address: nabillasmar@yahoo.fr
*** Postal address: Department of Statistics, The George Washington University, Washington, DC 20052, USA.
Email address: hosam@gwu.edu

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

to industrial tolerance as well as ageing of equipment. Specifically, we assume a Bernoulli probability model on data (not necessarily unbiased). Keys of infinite precision are obtained from a memoryless source that emits independent bits, with $\mathrm{P}(\text{bit} = 1) = p$ and $\mathrm{P}(\text{bit} = 0) = q = 1 - p$. We say the Bernoulli model (or the trie) is unbiased if $p = q = \frac{1}{2}$; otherwise, it is biased.

Let Δ_n be the distance between two randomly chosen keys in a random trie. Two of our results concern the mean and variance, and the periodic fluctuations therein:

$$\mathrm{E}[\Delta_n] = \frac{2}{h_p} \ln n + o(\ln n),$$
$$\mathrm{var}[\Delta_n] = 2\sigma_p^2 \ln n + o(\ln n),$$

where h_p and $\sigma_p^2 = (pq/h_p^3)(\ln p - \ln q)^2$ are constants ($h_p := -p \ln p - q \ln q$ is the data entropy) and the o-terms contain ignorable oscillating functions. We write the asymptotic variance in doubled form to make explicit the fact that asymptotically Δ_n is a convolution of two random variables, each of which is the depth of a randomly chosen key in a random trie. Furthermore, if the trie is biased then

$$\frac{\Delta_n - (2/h_p) \ln n}{\sqrt{\ln n}} \xrightarrow{\mathrm{D}} \mathcal{N}(0, 2\sigma_p^2),$$

where '$\xrightarrow{\mathrm{D}}$' denotes convergence in distribution.

Curiously, in a random unbiased trie, when $p = q$ the factor σ_p^2 becomes 0; the lower-order terms dominate, realizing a significantly different behavior. In this case, we recover the results of Christophi and Mahmoud (2005), who went into more details of the o-term.

In random unbiased tries, no scaling factors exist such that the scaled distance has a nontrivial limit distribution. In the biased case there is a Gaussian limit distribution, which we prove by showing that the distance between the distribution functions of $\Delta_n^* := (\Delta_n - \mathrm{E}[\Delta_n]) \ln^{-1/2} n$ and some limit, Δ^*, diminishes to 0 in the Wasserstein metric space, and the distribution of Δ^* is the fixed-point solution of a distributional equation that can be explicitly solved.

The main results have been sketched. The rest of the paper is organized in sections, as follows. In Section 2 the definition of a trie is given. At the end of Section 2, the notation used throughout is explained. In Section 3 we show how the moments can be derived by a Mellin-poissonization–inverse Mellin-depoissonization program. In Subsection 3.1 a functional equation for the moment generating function is set up, and its Mellin transform is computed. From this functional equation, the mean is derived in Subsection 3.2, and the variance is derived in Subsection 3.3. A Gaussian limit distribution is derived by the contraction method in Section 4, where we also add a few words on the origin of the method, its use, and references to it.

2. Tries

The trie is a data structure for digital data, or data represented by their decomposition into digits. Digital data are prevalent in science and technology. They have numerous applications in computer files, telecommunication signals, DNA, etc. The trie was proposed independently by De La Briandais (1959) and Fredkin (1960) for information re*trie*val.

A *binary trie* is a digital tree consisting of *internal nodes* that each have one or two children and leaf nodes that hold data (keys). The binary trie evolves according to the following algorithm. The trie is 'fed' with n keys. If $n = 0$ then nothing needs to be done; the insertion

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

algorithm terminates. If $n = 1$ then a leaf is allocated for the single key given. If $n \geq 2$ then an internal node is specified as a *root* of the tree; keys with 0 as their most significant bit go to the left-hand subtree, and keys with 1 as their most significant bit go to the right-hand subtree. The insertion algorithm is applied recursively in the subtrees, but at level ℓ the $(\ell + 1)$th most significant bit of the key is used for branching. When the algorithm halts, each key resides in a leaf by itself. The root-to-leaf paths in the trie correspond to minimal prefixes sufficient to distinguish the keys.

Let Δ_n be the distance (i.e. the number of tree edges) between two randomly selected keys in a random trie of size n, where all $\binom{n}{2}$ pairs of keys are equally likely to appear. The recurrence equations for Δ_n will involve δ_n, the depth of a randomly selected key in a random trie of size n, with *random* meaning that all keys are equally likely to be chosen.

The introduction of some additional notation will facilitate our exposition. The notation '$\stackrel{\mathrm{D}}{=}$' will mean equality in distribution, whereas '$\stackrel{\mathrm{D}}{\to}$' will denote convergence in distribution. Likewise, '$\stackrel{\mathrm{P}}{\to}$' and '$\stackrel{\mathrm{a.s.}}{\longrightarrow}$' will respectively denote convergence in probability and almost-sure convergence. A tilded variable will refer to an independent probabilistic copy of an untilded random variable having the same distribution. For example, $\tilde{Y}$ will mean a random variable independent of Y, with $\tilde{Y} \stackrel{\mathrm{D}}{=} Y$.

The Bernoulli random variable with success probability p will be denoted by $\mathrm{Ber}(p)$. Similarly, the binomial random variable arising on n independent trials, with success rate p per trial, will be denoted by $\mathrm{Bin}(n, p)$, and the normal distribution with mean μ and variance σ^2 will be denoted by $\mathcal{N}(\mu, \sigma^2)$. The notation $\|X\|_r$ will be used for the L_r-norm of any generic random variable X, and the notation $o_{\mathcal{L}_r}(g(n))$ used to symbolize a function with L_r-norm that is negligible in comparison with $g(n)$.

The Mellin transform of a function $f(x)$ is

$$\int_0^\infty f(x)x^{s-1}\,\mathrm{d}s,$$

and will be denoted by $f^*(s)$. The Mellin transform usually exists in vertical strips in the complex s-plane of the form

$$a < \mathrm{Re}\, s < b,$$

for real numbers a and b, $a < b$. We shall denote this strip by $\langle a, b\rangle$. The function $f(x)$ can be recovered from its transform using the line integral

$$f(x) = \frac{1}{2\pi \mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} f^*(s)x^{-s}\,\mathrm{d}s,$$

for any $c \in (a, b)$. The Mellin transform was surveyed in the context of the analysis of algorithms in Flajolet *et al.* (1995), Flajolet and Sedgewick (1995), and Szpankowski (2001, pp. 398–405).

3. Moments of the random distance

Let L_n and R_n respectively denote the number of keys residing in the left- and right-hand subtrees (with $L_n + R_n = n$). In view of the Bernoulli model, $L_n \stackrel{\mathrm{D}}{=} \mathrm{Bin}(n, q)$. Owing to the independence of the keys and the bits within, the recursion of the insertion algorithm preserves the probabilistic structure in the subtrees of the trie.

Given L_n, Δ_n can be Δ_{L_n} with probability $\binom{L_n}{2}/\binom{n}{2}$ when both keys come from the left-hand subtree, $\tilde{\Delta}_{R_n}$ with probability $\binom{R_n}{2}/\binom{n}{2}$ when both keys come from the right-hand subtree,

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

or $(\delta_{L_n}+1)+(\tilde{\delta}_{R_n}+1)$ with probability $L_nR_n/\binom{n}{2}$ when the two keys come from different subtrees. Hence, we have the conditional distribution

$$\Delta_n \mid L_n = \begin{cases} \Delta_{L_n} & \text{with probability } \binom{L_n}{2}/\binom{n}{2}, \\ \tilde{\Delta}_{R_n} & \text{with probability } \binom{R_n}{2}/\binom{n}{2}, \\ (\delta_{L_n}+1)+(\tilde{\delta}_{R_n}+1) & \text{with probability } L_nR_n/\binom{n}{2}, \end{cases} \tag{1}$$

with boundary condition $\Delta_0=\Delta_1=\delta_0=\delta_1=0$.

We note that Δ_{L_n} and $\tilde{\Delta}_{R_n}$ are *dependent* through the dependency of L_n and R_n, but that, given the value of L_n (and, hence, R_n), they are *conditionally independent*; the same applies to δ_{L_n} and $\tilde{\delta}_{R_n}$. Hence, $\tilde{\Delta}_i$ is independent of Δ_j for all i and j; likewise, $\tilde{\delta}_i$ is independent of δ_j for all i and j.

3.1. Functional equations

To establish a functional equation for the distribution of Δ_n, we need a functional equation for δ_n. The asymptotic distribution of δ_n was found independently in Pittel (1986) and Jacquet and Régnier (1987) for memoryless sources; Jacquet and Szpankowski (1991) extended the results to Markovian sources. We specifically need the formulation via functional equations as elegantly surveyed in Szpankowski (2001, p. 448).

We begin by deriving a functional equation for the moment generating function Δ_n from the basic conditional recurrence (1),

$$\mathrm{E}[\mathrm{e}^{\Delta_n t} \mid L_n] = \mathrm{e}^{\Delta_{L_n} t}\frac{\binom{L_n}{2}}{\binom{n}{2}} + \mathrm{e}^{\tilde{\Delta}_{R_n} t}\frac{\binom{R_n}{2}}{\binom{n}{2}} + \exp\{[(\delta_{L_n}+1)+(\tilde{\delta}_{R_n}+1)]t\}\frac{L_nR_n}{\binom{n}{2}},$$

with an unconditional expectation

$$\begin{aligned}\binom{n}{2}\phi_{\Delta_n}(t) &:= \binom{n}{2}\mathrm{E}[\mathrm{e}^{\Delta_n t}] \\ &= \mathrm{E}\left[\binom{L_n}{2}\mathrm{e}^{\Delta_{L_n} t}\right] + \mathrm{E}\left[\binom{R_n}{2}\mathrm{e}^{\Delta_{R_n} t}\right] + \mathrm{e}^{2t}\,\mathrm{E}[L_nR_n\mathrm{e}^{(\delta_{L_n}+\tilde{\delta}_{R_n})t}]. \end{aligned} \tag{2}$$

The recurrence is not easy to solve. However, a poissonized version of the problem is amenable to solution via the Mellin transform. Consequently, suppose that instead of there being a fixed n, the number of keys to be stored in the tree is first determined by a random draw from a Poisson distribution with parameter z. Let N_z be such a random number.

In a step toward poissonization, we multiply both sides of (2) by z^n and sum over all possible n. We do the calculations on the right-hand side by conditioning on L_n, and obtain

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{z^n\phi_{\Delta_n}(t)\binom{n}{2}}{n!} &= \sum_{n=0}^{\infty}\sum_{\ell=0}^{n}\frac{(pz)^\ell\phi_{\Delta_\ell}(t)\binom{\ell}{2}}{\ell!}\frac{(qz)^{n-\ell}}{(n-\ell)!} \\ &\quad + \sum_{n=0}^{\infty}\sum_{\ell=0}^{n}\frac{(qz)^{n-\ell}\phi_{\Delta_{n-\ell}}(t)}{(n-\ell)!}\frac{(pz)^\ell\binom{\ell}{2}}{\ell!} \\ &\quad + \mathrm{e}^{2t}\sum_{n=0}^{\infty}\sum_{\ell=0}^{n}\frac{(pz)^\ell\ell}{\ell!}\phi_{\delta_\ell}(t)\frac{(qz)^{n-\ell}(n-\ell)}{(n-\ell)!}\phi_{\delta_{n-\ell}}(t),\end{aligned}$$

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

where $\phi_{\delta_j}(t) := \mathrm{E}[\mathrm{e}^{\delta_j t}]$. To complete the poissonization we introduce the bivariate generating function $\Phi(t, z) = \mathrm{e}^{-z} \sum_{n=0}^{\infty} \binom{n}{2} \phi_{\Delta_n}(t) z^n / n!$, such that

$$\Phi(t, z) = \mathrm{E}\left[\binom{N_z}{2} \phi_{\Delta_{N_z}}(t)\right],$$

as can be seen by conditioning on N_z.

Direct work with $\Phi(t, z)$ gives rise to a technical difficulty in using the Mellin transform. To ensure the existence of the transform, we shift $\Phi(t, z)$ down by $\mathrm{e}^{2t} z^2/2$, and define

$$P(t, z) = \Phi(t, z) - \mathrm{e}^{2t} \frac{z^2}{2}.$$

We can now express the recurrence in the form

$$P(t, z) = P(t, pz) + P(t, qz) + \mathrm{e}^{2t}[(Q(t, pz) + pz)(Q(t, qz) + qz)] - \mathrm{e}^{2t} pqz^2, \qquad (3)$$

where $Q(t, z)$ is the shifted, poissonized function $\mathrm{E}[N_z \mathrm{e}^{\delta_{N_z} t}] - z$ for the random depth.

3.2. The mean

The moments can be derived in a systematic manner by techniques, such as poissonization and depoissonization (see Jacquet and Szpankowski (1998) or Szpankowski (2001, p. 465)), and singularity analysis and Mellin transformation (see Flajolet *et al.* (1995)), belonging to the tool-kit of analysis of algorithms.

The kth derivative of (3) yields a functional equation for the (shifted, poissonized) kth moment of Δ_n. The first derivative gives

$$A(z) := \frac{\partial}{\partial t} P(t, z)\bigg|_{t=0} = \mathrm{E}\left[\binom{N_z}{2} \Delta_{N_z}\right] - z^2.$$

Thus,

$$A(z) = A(pz) + A(qz) + pz a(qz) + qz a(pz),$$

where $a(z) := (\partial/\partial t) Q(t, z)|_{t=0} = \mathrm{E}[N_z \delta_{N_z}]$. The Mellin transform of $A(z)$ is

$$A^*(s) = \frac{(qp^{-1-s} + pq^{-1-s}) a^*(s+1)}{1 - p^{-s} - q^{-s}},$$

where $a^*(s)$ is the Mellin transform of $a(z)$, which can be found in a number of reference sources. (It was developed in Szpankowski (2001, p. 448) via a shortcut argument that avoids recurrence and uses poissonization as a paradigm.) These references give

$$a^*(s) = -\frac{\Gamma(s+1)}{1 - p^{-s} - q^{-s}}.$$

Upon substituting this into $A^*(s)$, we obtain the Mellin transform

$$A^*(s) = -\frac{(pq^{-(s+1)} + qp^{-(s+1)})\Gamma(s+2)}{(1 - p^{-s} - q^{s})(1 - p^{-(s+1)} - q^{-(s+1)})},$$

existing in $\langle -3, -2 \rangle$.

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

The poissonized average is retrieved by the inversion

$$A(z) = \frac{1}{2\pi \mathrm{i}} \int_{-5/2-\mathrm{i}\infty}^{-5/2+\mathrm{i}\infty} A^*(s) z^{-s}\, \mathrm{d}s.$$

We evaluate this integral using 'the method of closing the box' (see Szpankowski (2001, p. 408)). It starts from a rectangular contour integral connecting the points $-\frac{5}{2} \pm \mathrm{i}M$ and $\theta \pm \mathrm{i}M$. The sides are chosen so that no poles are crossed, with M large and θ an arbitrary positive real number. Then we argue that as $M \to \infty$ this contour integral reduces to the required integral for the inverse Mellin transform, plus an error arising from integration on the side joining the points $\theta \pm \mathrm{i}\infty$ (the other two integrals on the top and bottom sides vanish, in view of the presence of the gamma function). The contour integral itself is evaluated via the residues of the poles in the infinite rectangle (strip). The conclusion is

$$A(z) = -\sum_{\text{poles}} \text{residues of poles in } \left\langle -\frac{5}{2}, \theta \right\rangle + o(z^{-\theta}). \tag{4}$$

This leaves us with residue calculations at the poles of the gamma function and the roots of the equation

$$(1 - p^{-s} - q^{-s})(1 - p^{-(s+1)} - q^{-(s+1)}) = 0.$$

Inside the strip, the gamma function has a singularity at -2, and the equation

$$1 - p^{-(s+1)} - q^{-(s+1)} = 0$$

has roots both at $s_0 = -2$ and, for $\ln p / \ln q = r/m$, where $\gcd(r, m) = 1$ for integers r and m, at $s_k = -2 + 2\pi \mathrm{i} k m / \ln q$ for $k = \pm 1, \pm 2, \ldots$. Thus, the transform $A^*(s)$ has a double pole at s_0 and, for $\ln p / \ln q$ rational, has simple poles at s_k for each nonzero k. The roots of the equation $1 - p^{-s} - q^{-s} = 0$ are at -1 and, for $\ln p / \ln q = r/m$, where $\gcd(r, m) = 1$ for integers r and m, at $-1 + 2\pi \mathrm{i} m j / \ln q$ for $j = 0, \pm 1, \pm 2, \ldots$. These contribute $O(z)$ terms that can be subsumed in the error because ultimately we have to divide by z^2.

Although θ can be taken to be arbitrarily large, we shall fix its value just past all the poles at -0.99, because eventually we depoissonize and this operation gives an $o(n^2)$ error term anyway. By collecting contributions of the residues of the poles in the strip and substituting them into (4), we arrive at

$$\begin{aligned} A(z) &= \mathrm{E}\left[\binom{N(z)}{2} \Delta_{N(z)}\right] - z^2 \\ &= \frac{1}{2pqh_p^2}[2pqh_p z^2 \ln z \\ &\qquad + (2pqh_p\gamma + (\ln^2 q - \ln^2 p)p^3 + (2\ln p \ln q - 3\ln^2 q)p^2 \\ &\qquad + (3\ln^2 q - 2\ln p \ln q)p - \ln^2 q + 2pqh_p^2)z^2] \\ &\quad + 2z^2\beta(z) + o(z^2), \end{aligned}$$

where γ is Euler's constant and $\beta(\cdot)$ is the function

$$\beta(z) = \begin{cases} -\dfrac{1}{2h_p} \displaystyle\sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \Gamma\left(\frac{2\pi \mathrm{i} k m}{\ln q}\right) z^{-2\pi \mathrm{i} k m / \ln q} & \text{if } \ln p / \ln q \text{ is rational,} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

In all cases, $|\beta(z)|$ is a small function for the typical values of p staying away from 0 and 1. For instance, when $r/m = \frac{1}{2}$, we have $p^2 = q$ or $p = \frac{1}{2}(\sqrt{5}-1) \approx 0.618$, and $|\beta(z)| \leq 0.18 \times 10^{-8}$, uniformly in z.

The $O(z^2 \ln z)$-term of $A(z)$ satisfies the conditions for depoissonization (see Theorem 10 of Jacquet and Szpankowski (1998)). The main result of this section follows, substituting n for z. This introduces a negligible $o(1)$ error in the average. The final result can be considerably simplified by straightforward algebraic manipulation.

Proposition 1. *In a trie of n random keys following the Bernoulli model, the average distance between two randomly selected keys is*

$$\mathrm{E}[\Delta_n] = \frac{2}{h_p}\ln n + \frac{2\gamma}{h_p} + 2 - \frac{1}{pqh_p^2}(p^3\ln^2 p + 2pq\ln p\ln q + q^3\ln^2 q) + 4\beta(n) + o(1),$$

where $\beta(n)$ is the oscillating function given in (5).

3.3. The variance

The second derivative of (3) gives rise to a functional equation for the second moment of Δ_n. There are an enormous number of details, and we shall only sketch the calculation. By taking the second derivative we obtain

$$B(z) := \frac{\partial^2}{\partial^2 t}P(t,z)|_{t=0} = \mathrm{E}\left[\binom{N_z}{2}\Delta^2_{N_z}\right] - 2z^2.$$

This gives

$$B(z) = B(pz) + B(qz) + 4pzA(qz) + 4qzA(pz) + pzb(qz) + qzb(pz) + 2a(pz)a(qz),$$

with $b(z) = (\partial^2/\partial^2 t)Q(t,z)|_{t=0} = \mathrm{E}[N_z\delta^2_{N_z}]$. We write

$$\begin{aligned} B(z) = {} & B(pz) + B(qz) + 4pza(qz) + 4qza(pz) + pzb(qz) + qzb(pz) \\ & + 2\left(a(pz) - \frac{pz\ln(pz)}{h_p}\right)\left(a(qz) - \frac{qz\ln(qz)}{h_p}\right) \\ & + 2a(pz)\frac{qz\ln(qz)}{h_p} + 2a(qz)\frac{pz\ln(pz)}{h_p} - \frac{2pqz^2\ln(pz)\ln(qz)}{h_p^2}. \end{aligned}$$

To aid in symmetrizing the equation with respect to the roles of p and q, we write the term $-2pqz^2\ln(pz)\ln(qz)/h_p^2$ as $-f(z) + f(pz) + f(qz)$, where

$$\begin{aligned} f(z) = {} & \frac{z^2\ln^2 z}{h_p^2} + \left(\frac{p^2\ln p + q^2\ln q}{pqh_p^2} + \frac{\ln(pq)}{h_p^2}\right)z^2\ln z \\ & + \left(\frac{p^2\ln^2 p + q^2\ln^2 q}{2pqh_p^2} + \left(\frac{p^2\ln p + q^2\ln q}{pqh_p^2} + \frac{\ln(pq)}{h_p^2}\right)\frac{p^2\ln p + q^2\ln q}{2pq} \right. \\ & \left. + \frac{\ln p\ln q}{h_p^2}\right)z^2. \end{aligned}$$

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

The function $f(z)$ looks like an asymptotic expansion of some function, $F(z)$, that has a Mellin transform in the strip $\langle -3, -2\rangle$ and satisfies $F^*(s) = \lambda(s)\Gamma(s)$ for a function $\lambda(\cdot)$ of simple form. Indeed, $w(z) = f(z) - F(z)$ is a regular function of order $O(z)$, and the term

$$2\left(a(pz) - \frac{pz\ln(pz)}{h_p}\right)\left(a(qz) - \frac{qz\ln(qz)}{h_p}\right) + w(pz) + w(qz) - w(z)$$

is $O(z^2)$.

Let $B(z) + F(z) = B_1(z) + B_2(z)$, where

$$\begin{aligned} B_1(z) &= B_1(pz) + B_1(qz) + 4pza(qz) + 4qza(pz) + pzb(qz) + qzb(pz) \\ &\quad + 2\frac{pz\ln(pz)}{h_p}a(qz) + 2\frac{qz\ln(qz)}{h_p}a(pz), \\ B_2(z) &= B_2(pz) + B_2(qz) + 2\left(a(pz) - \frac{pz\ln(pz)}{h_p}\right)\left(a(qz) - \frac{qz\ln(qz)}{h_p}\right) \\ &\quad - w(z) + w(pz) + w(qz). \end{aligned}$$

Lemma 1. *In the cone* $S_\theta = \{z \in \mathbb{C}\colon |\arg(z)| < \theta\}$, $0 \le \theta < \pi/2$, *we have*

$$B_2(z) = O(z^2) \quad \text{as } z \to \infty.$$

Proof. We prove this lemma by induction in vertical sections of the cone. Write $B_2(z)$ as

$$B_2(z) = B_2(pz) + B_2(qz) + \rho(z),$$

where $\rho(z) = O(z^2)$. Thus, there is a value, $z_0 > 1$, such that $|\rho(z)| \le K|z|^2$ for all $|z| \ge z_0$. Let $\nu_1 = \min(p, q)$, $\nu_2 = \max(p, q)$, and $z_0' = z_0/\nu_1$. For positive integer m, let $D_m = \{z\colon \mathrm{Re}(z) \in [z_0', z_0'\nu_2^{-m}]\}$. Let us add $D_0 = \{z\colon \mathrm{Re}(z) \in [z_0, z_0']\}$. We chose the starting section so that the bound for $|\rho(z)|$ is already in effect in D_0 and beyond.

Whenever $z \in D_{m+1} - D_m$, both pz and qz fall in $D_0 \cup D_m$. Let $c = \sup_{z\in D_0\cup D_1} |B_2(z)|$. We start the induction at $m = 1$, where we already have

$$B_2(z) \le c < c|z|^2.$$

Let us take $\alpha = \max(c/\nu_1^2, 2c + K, K/(1 - p^2 - q^2))$. In D_1, we have $|B_2(z)| \le \alpha|z|^2$. Assume that $|B_2(z)| \le \alpha|z|^2$ in D_m, for $m \ge 1$. Suppose now that $z \in D_{m+1} - D_m$. The maximal value of pq is $\frac{1}{4}$, and $K \le K/(2pq)$. The functional equation for $B_2(z)$ gives us

$$\begin{aligned} |B_2(z)| &\le \max(c, \alpha p^2|z|^2) + \max(c, \alpha q^2|z|^2) + K|z|^2 \\ &\le \alpha p^2|z|^2 + \alpha q^2|z|^2 + \alpha(1 - p^2 - q^2)|z|^2 \\ &= \alpha|z|^2, \end{aligned}$$

completing the induction in all the sections.

The Mellin transform of $B_1(z)$ is

$$\begin{aligned} B_1^*(s) = 2\Bigg[&\left(2 + \frac{\ln p - \ln q}{h_p}\right)pq^{-(s+1)} + \left(2 + \frac{\ln q - \ln p}{h_p}\right)qp^{-(s+1)}\Bigg] \\ &\times \frac{a^*(s+1)}{1 - p^{-s} - q^{-s}} \\ &+ \left(\frac{qp^{-(s+1)} + pq^{-(s+1)}}{1 - p^{-s} - q^{-s}}\right)\left(b^*(s+1) + \frac{2}{h_p}\frac{\mathrm{d}}{\mathrm{d}s}a^*(s+1)\right). \end{aligned}$$

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

By inverting this Mellin transform and residue computation, we obtain

$$\begin{aligned}\mathrm{E}\left[\binom{N_z}{2}\Delta_{N_z}^2\right] &= \frac{2}{h_p^2}z^2\ln^2 z \\ &\quad - \left[\frac{3}{pqh_p^3}(q^4\ln^2 q + 2pq\ln p\ln q(1-pq) + p^4\ln^2 p)\right. \\ &\quad \left. - \frac{4\gamma}{h_p^2} + \frac{p^2\ln p + q^2\ln q}{pqh_p^2} + \frac{\ln(pq)}{h_p^2} - 8\frac{\beta(z)}{h_p}\right]z^2\ln z \\ &\quad + O(z^2).\end{aligned}$$

Again, the conditions for depoissonization are met. The complicated expressions above can be simplified using straightforward algebra to the expression in the following statement.

Theorem 1. *In a trie of n random keys following the Bernoulli model, the variance of the distance between two randomly selected keys is*

$$\mathrm{var}[\Delta_n] = 2\left(\frac{pq}{h_p^3}(\ln(p) - \ln(q))^2\right)\ln n + O(1).$$

Note that the case $p = q$ presents a degeneracy, which was handled in Christophi and Mahmoud (2005).

Corollary 1. *As $n \to \infty$, $\Delta_n/\ln n \xrightarrow{\mathrm{P}} 2/h_p$.*

Proof. By Chebyshev's inequality, we have

$$\mathrm{P}(|\Delta_n - \mathrm{E}[\Delta_n]| > \varepsilon\,\mathrm{E}[\Delta_n]) \le \frac{\mathrm{var}[\Delta_n]}{\varepsilon^2\,\mathrm{E}^2[\Delta_n]}.$$

The orders of magnitude found in Proposition 1 and Theorem 1 yield

$$\mathrm{P}\left(\left|\frac{\Delta_n}{\mathrm{E}[\Delta_n]} - 1\right| > \varepsilon\right) = O\left(\frac{1}{\ln n}\right).$$

We thus have

$$\frac{\Delta_n}{\mathrm{E}[\Delta_n]} \xrightarrow{\mathrm{P}} 1,$$

which can be combined with the convergence $\mathrm{E}[\Delta_n]/\ln n \to 2/h_p$ to give the statement of the corollary.

4. Limit distributions

In principle, one can continue pumping higher moments by the methods utilized for the mean and variance, and aspire to determine limit distributions by a method of recursive moments (see Chern *et al.* (2002), for example). However, as was already mentioned, the explosive complexity is forbidding. The contraction method offers a shortcut. A solution is guessed based on some heuristics in the structure of the problem, and then the guess is verified by showing convergence of the distribution function of $\Delta_n^* := (\Delta_n - \mathrm{E}[\Delta_n])\ln^{-1/2} n$ to that of the guessed limit in some metric space. Recently, the Wasserstein and Zolotarev metrics have been popularized in the context of the contraction method.

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

The contraction method was introduced by Rösler (1991). Rachev and Rüschendorf (1995) added several useful extensions. Recently general contraction theorems and multivariate extensions were added by Rösler (2001), Neininger (2001), and Neininger and Rüschendorf (2004). Rösler and Rüschendorf (2001) provides a valuable survey.

We start from the recursive representation (1), written in the form

$$\Delta_n = \Delta_{L_n} I_n + \tilde{\Delta}_{R_n} J_n + (\delta_{L_n} + \tilde{\delta}_{R_n} + 2) K_n,$$

where I_n is the indicator of the event that both keys are chosen from the left-hand subtree, J_n is the indicator of the event that both keys are chosen from the right-hand subtree, and K_n is the indicator of the event that the keys are chosen from different subtrees. The indicators are inserted to truncate an irrelevant choice; they are of course mutually exclusive ($I_n + J_n + K_n \equiv 1$). For $n \geq 2$, we can reorganize the above relation as

$$\begin{aligned}\frac{\Delta_n - \mathrm{E}[\Delta_n]}{\sqrt{\ln n}} \stackrel{\mathrm{D}}{=} \frac{\Delta_{L_n} - \mathrm{E}[\Delta_{L_n}]}{\sqrt{\ln L_n}} I_n \sqrt{\frac{\ln L_n}{\ln n}} + \frac{\tilde{\Delta}_{R_n} - \mathrm{E}[\tilde{\Delta}_{R_n}]}{\sqrt{\ln R_n}} J_n \sqrt{\frac{\ln R_n}{\ln n}} \\ + Y_n^* K_n + \frac{1}{\sqrt{\ln n}} (\mathrm{E}[\Delta_{L_n}] I_n + \mathrm{E}[\tilde{\Delta}_{R_n}] J_n \\ + (\mathrm{E}[\delta_{L_n}] + \mathrm{E}[\tilde{\delta}_{R_n}] + 2) K_n - \mathrm{E}[\Delta_n]),\end{aligned}$$

where

$$Y_n^* := \frac{\delta_{L_n} - \mathrm{E}[\delta_{L_n}]}{\sqrt{\ln L_n}} \sqrt{\frac{\ln L_n}{\ln n}} + \frac{\tilde{\delta}_{R_n} - \mathrm{E}[\tilde{\delta}_{R_n}]}{\sqrt{\ln R_n}} \sqrt{\frac{\ln R_n}{\ln n}}.$$

This equation can be written in terms of the normed variables as

$$\Delta_n^* = \Delta_{L_n}^* I_n \sqrt{\frac{\ln L_n}{\ln n}} + \tilde{\Delta}_{R_n}^* J_n \sqrt{\frac{\ln R_n}{\ln n}} + Y_n^* K_n + G_n, \tag{6}$$

where

$$G_n := \frac{1}{\ln n} (\mathrm{E}[\Delta_{L_n}] I_n + \mathrm{E}[\tilde{\Delta}_{R_n}] J_n + (\mathrm{E}[\delta_{L_n}] + \mathrm{E}[\tilde{\delta}_{R_n}] + 2) K_n - \mathrm{E}[\Delta_n]).$$

We first find the limit of this equation heuristically and then confirm it by an inductive proof in the Wasserstein metric space. The additive terms in the representation (6) are dependent. For instance, Δ_{L_n} and $\tilde{\Delta}_{R_n}$ are dependent through L_n and R_n (though these copies are conditionally independent when L_n and R_n are given). Also I_n, J_n, and K_n are dependent, etc. However, the fact that the binomial distribution of L_n is sharply concentrated around its average, namely

$$\frac{L_n}{n} \xrightarrow{\text{a.s.}} q, \qquad \frac{R_n}{n} \xrightarrow{\text{a.s.}} p, \tag{7}$$

loosens the dependence. As an immediate consequence of (7) we have the convergence

$$\sqrt{\frac{\ln L_n}{\ln n}} \xrightarrow{\text{a.s.}} 1, \qquad \sqrt{\frac{\ln R_n}{\ln n}} \xrightarrow{\text{a.s.}} 1. \tag{8}$$

If Δ_n^* converges to a limit then so do $\Delta_{L_n}^*$ and $\tilde{\Delta}_{R_n}^*$, because both L_n and R_n grow to infinity almost surely, and these limits would be eventually independent. The limit variable, δ^*, of

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

$(\delta_n - \mathrm{E}[\delta_n])\ln^{-1/2} n$ is known to be $\mathcal{N}(0, \sigma_p^2)$ for biased tries (it does not exist in unbiased tries); see Pittel (1986). So, by the same token as before, $(\delta_{L_n} - \mathrm{E}[\Delta_{L_n}])\ln^{-1/2} L_n$ and $(\tilde{\delta}_{R_n} - \mathrm{E}[\tilde{\Delta}_{R_n}])\ln^{-1/2} R_n$, although dependent, would eventually be independent copies of $\mathcal{N}(0, \sigma_p^2)$.

Each of the indicators I_n, J_n, and K_n is a *conditional* Bernoulli random variable. For instance, for any $n \geq 2$,

$$I_n = \mathrm{Ber}\left(\frac{L_n(L_n-1)}{n(n-1)}\right),$$

which is to be interpreted as $\mathrm{Ber}(\ell(\ell-1)/(n(n-1)))$, whenever $L_n = \ell$. Also, the indicators (I_n, J_n, K_n) tend to a vector, (I, J, K), of three jointly distributed Bernoulli random variables on the nonzero vertices of the unit simplex in three dimensions, with marginals

$$I_n \xrightarrow{\text{a.s.}} I = \mathrm{Ber}(q^2), \tag{9}$$

$$J_n \xrightarrow{\text{a.s.}} J = \mathrm{Ber}(p^2), \tag{10}$$

$$K_n \xrightarrow{\text{a.s.}} K = \mathrm{Ber}(2pq). \tag{11}$$

Lemma 2. *As $n \to \infty$, $G_n \to 0$.*

Proof. By utilizing Proposition 1, we can bound the term $\mathrm{E}[\Delta_{L_n}]$ by conditioning as follows:

$$\begin{aligned}
\mathrm{E}[\Delta_{L_n}] &= \sum_{\ell=0}^{n} \mathrm{E}[\Delta_\ell]\binom{n}{\ell} p^\ell q^{n-\ell} \\
&= \sum_{\ell=2}^{n} \binom{n}{\ell} p^\ell q^{n-\ell}\left(\frac{2}{h_p}\ln \ell + O(1)\right) \\
&\leq \left(\frac{2}{h_p}\ln n + O(1)\right)\sum_{\ell=2}^{n}\binom{n}{\ell} p^\ell q^{n-\ell} \\
&< \frac{2}{h_p}\ln n + O(1).
\end{aligned}$$

Likewise, by conditioning the known result about the asymptotic mean random depth from Pittel (1986) (see also Jacquet and Régnier (1987) and Jacquet and Szpankowski (1991)), we have $\mathrm{E}[\delta_{L_n}] < (1/h_p)\ln n + O(1)$. By symmetry, similar bounds hold in the right-hand subtree for the terms $\mathrm{E}[\Delta_{R_n}]$ and $\mathrm{E}[\delta_{R_n}]$.

So, now we can represent G_n as

$$\begin{aligned}
G_n = \frac{1}{\sqrt{\ln n}}\Bigg(&\left[\frac{2}{h_p}\ln n + O(1)\right] I_n + \left[\frac{2}{h_p}\ln n + O(1)\right] J_n \\
&+ \left(\left[\frac{1}{h_p}\ln n + O(1)\right] + \left[\frac{1}{h_p}\ln n + O(1)\right] + 2\right) K_n \\
&- \left(\frac{2}{h_p}\ln n + O(1)\right)\Bigg) \\
= \frac{1}{\sqrt{\ln n}}&\left(\frac{2}{h_p}(I_n + J_n + K_n)\ln n - \frac{2}{h_p}\ln n + O(1)\right) \\
= O&\left(\frac{1}{\sqrt{\ln n}}\right).
\end{aligned}$$

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

According to the form of (6) and the convergence relations (7)–(11) and Lemma 2, if Δ_n^* converges to a limit, say Δ^*, then the limit satisfies the distributional equation

$$\Delta^* \stackrel{\mathrm{D}}{=} \Delta^* I + \tilde{\Delta}^* J + Y^* K, \tag{12}$$

where $Y^* \stackrel{\mathrm{D}}{=} \delta^* + \tilde{\delta}^*$ and $(\Delta^*, \tilde{\Delta}^*, Y^*)$ is independent of (I, J, K).

Let F_n^* be the distribution function of Δ_n^*, and let F^* be the distribution function of Δ^*. To actually prove that a limit for Δ_n^* exists in distribution, and satisfies the fixed-point limit equation (12), it suffices to show that the second-order distance

$$d_2(F_n^*, F^*) = \inf \|V_n - W\|_2$$

converges to 0 as $n \to \infty$; the infimum in this definition is taken over all pairs (V_n, W) of random variables with respective distribution functions F_n^* and F^*.

The L_2-norm $\|V_n - W\|_2 = \sqrt{\mathrm{E}[(V_n - W)^2]}$, for any particular pair (V_n, W), gives an upper bound on $d_2(F_n^*, F^*)$. In particular, we have

$$d_2^2(F_n^*, F^*) \le b_n := \mathrm{E}[(\Delta_n^* - \Delta^*)^2].$$

Lemma 3. *As $n \to \infty$, $b_n \to 0$.*

Proof. In (6) we can replace several factors by their limit plus an asymptotically negligible corrective term. For example, the difference $|I_n - I|^m$ is always bounded (by 1) for any index m, and the convergence in (9) implies convergence in L^m; consequently, we have $I_n \sqrt{\ln L_n / \ln n} = I + o_{\mathcal{L}_1}(1)$. By subtracting (12) from (6) we obtain

$$\begin{aligned} b_n &:= \mathrm{E}[(\Delta_{L_n}^* - \Delta^*) I + (\tilde{\Delta}_{R_n}^* - \tilde{\Delta}^*) J + (Y_n^* - Y^*) K + G_n + o_{\mathcal{L}_1}(1)] \\ &:= \mathrm{E}[(V_1 + V_2 + V_3 + G_n + o_{\mathcal{L}_1}(1))^2]. \end{aligned} \tag{13}$$

Upon squaring the five additive terms in the latter equation, fifteen terms appear. The terms $\mathrm{E}[V_1^2]$ and $\mathrm{E}[V_2^2]$ define a recurrence for b_n, as follows. The limiting indicator I is independent of L_n, as the latter depends on only finitely many keys. We find that

$$\begin{aligned} \mathrm{E}[V_1^2] &= \mathrm{E}[(\Delta_{L_n}^* - \Delta^*)^2 I] \\ &= \mathrm{E}[I] \sum_{\ell=0}^{n} \mathrm{E}[(\Delta_{L_n}^* - \Delta^*)^2 \mid L_n = \ell] \, \mathrm{P}(L_n = \ell) \\ &= q^2 \sum_{\ell=0}^{n} \mathrm{E}[(\Delta_\ell - \Delta^*)^2] \binom{n}{\ell} p^{n-\ell} q^\ell \\ &= q^2 \sum_{\ell=0}^{n} \binom{n}{\ell} b_\ell \, p^{n-\ell} q^\ell \end{aligned}$$

and, similarly,

$$\mathrm{E}[V_2^2] = p^2 \sum_{r=0}^{n} \binom{n}{r} b_r \, p^r q^{n-r}.$$

The three terms $2\,\mathrm{E}[V_1 V_2]$, $2\,\mathrm{E}[V_1 V_3]$, and $2\,\mathrm{E}[V_2 V_3]$ are exactly 0, because they contain cross-products of mutually exclusive indicators; for example,

$$\mathrm{E}[V_1 V_2] = \mathrm{E}[(\Delta_{L_n}^* - \Delta^*) I (\tilde{\Delta}_{R_n}^* - \tilde{\Delta}^*) J] = 0$$

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

as $IJ \equiv 0$. All the other terms combined contribute only $O(\ln^{-1/2} n)$, due to the exact centering of the random variables and the smallness of the corrections, for example. We illustrate the calculations for only one of the terms,

$$2\,\mathrm{E}[V_3 G_n] = 2\,\mathrm{E}[(Y_n^* - Y^*)K G_n].$$

By the Cauchy–Schwarz inequality and the fact that $K \le 1$, we have

$$\mathrm{E}[(Y_n^* - Y^*)G_n K] \le \sqrt{\mathrm{E}[G_n^2]}\sqrt{\mathrm{E}[(Y_n^* - Y^*)^2]}.$$

The proof of Lemma 2 gives $O(1/\ln n)$ as an upper bound for $\mathrm{E}[G_n^2]$. Furthermore,

$$\sqrt{\mathrm{E}[(Y_n^* - Y^*)^2]}$$

is $o(1)$, as shown, for example, in Jacquet and Szpankowski (1991).

We rewrite (13) using our results for the fifteen terms in the expansion:

$$\begin{aligned} b_n &= q^2 \sum_{\ell=0}^{n} \binom{n}{\ell} b_\ell p^{n-\ell} q^\ell + p^2 \sum_{r=0}^{n} \binom{n}{r} b_r p^r q^{n-r} + O\left(\frac{1}{\sqrt{\ln n}}\right) \\ &= \frac{1}{1 - p^{n+2} - q^{n+2}} \left(\sum_{\ell=0}^{n-1} \binom{n}{\ell} b_\ell p^{n-\ell} q^{\ell+2} + \sum_{r=0}^{n-1} \binom{n}{r} b_r p^{r+2} q^{n-r} \right) \\ &\quad + O\left(\frac{1}{\sqrt{\ln n}}\right). \end{aligned}$$

We can show by induction that $b_n \to 0$. Indeed, $d_2(F_n^*, F^*) \le b_n \to 0$ as $n \to \infty$. The convergence $d_2(F_n^*, F^*) \to 0$ verifies the convergence $\Delta_n^* \xrightarrow{\mathrm{D}} \Delta^*$.

Theorem 2. *In a trie of n random keys following the biased Bernoulli model, the distance, Δ_n, between two randomly selected keys satisfies*

$$\frac{\Delta_n - (2/h_p)\ln n}{\sqrt{\ln n}} \xrightarrow{\mathrm{D}} \mathcal{N}(0, 2\sigma_p^2).$$

Proof. The limiting random variable Δ^* has a distribution that satisfies the distributional equation (12). Let $\phi_X(t)$ be the characteristic function of a generic random variable X. Conditioning on $\boldsymbol{M} = (I, J, K)$, we find the representation

$$\begin{aligned} \phi_{\Delta^*}(t) &= \mathrm{E}[\mathrm{e}^{t(I\Delta^* + J\tilde{\Delta}^* + Y^* K)}] \\ &= \mathrm{E}[\mathrm{e}^{t(I\Delta^* + J\tilde{\Delta}^* + Y^* K)} \mid \boldsymbol{M} = (1,0,0)]\,\mathrm{P}(\boldsymbol{M} = (1,0,0)) \\ &\quad + \mathrm{E}[\mathrm{e}^{t(I\Delta^* + J\tilde{\Delta}^* + Y^* K)} \mid \boldsymbol{M} = (0,1,0)]\,\mathrm{P}(\boldsymbol{M} = (0,1,0)) \\ &\quad + \mathrm{E}[\mathrm{e}^{t(I\Delta^* + J\tilde{\Delta}^* + Y^* K)} \mid \boldsymbol{M} = (0,0,1)]\,\mathrm{P}(\boldsymbol{M} = (0,0,1)) \\ &= q^2 \phi_{\Delta^*}(t) + p^2 \phi_{\Delta^*}(t) + 2pq\phi_{Y^*}(t). \end{aligned}$$

Thus,

$$\phi_{\Delta^*}(t) = \phi_{Y^*}(t).$$

Since $Y^* = \delta^* + \tilde{\delta}^*$ and both δ^* and $\tilde{\delta}^*$ are independent copies of the limit of the normalized random depth, which is known to be $\mathcal{N}(0, \sigma_p^2)$ (see Pittel (1986), for example), we therefore have $Y^* \stackrel{\mathrm{D}}{=} \mathcal{N}(0, 2\sigma_p^2)$.

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press

Acknowledgement

The third author wishes to thank Dr. Costas Christophi for several insightful discussions on the content of this paper.

References

Chern, H. H., Hwang, H. K. and Tsai, T. H. (2002). An asymptotic theory for Cauchy–Euler differential equations with applications to the analysis of algorithms. *J. Algorithms* **44,** 177–225.

Christophi, C. and Mahmoud, H. (2005). The oscillatory distribution of distances in random tries. *Ann. Appl. Prob.* **15,** 1536–1564.

De La Briandais, R. (1959). File searching using variable length keys. In *Proc. Western Joint Computer Conference*, AFIPS, San Francisco, CA, pp. 295–298.

Flajolet, P. and Sedgewick, R. (1995). Mellin transforms and asymptotics: finite differences and Rice's integrals. *Theoret. Comput. Sci.* **144,** 101–124.

Flajolet, P., Gourdon, X. and Dumas, P. (1995). Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.* **144,** 3–58.

Fredkin, E. (1960). Trie memory. *Commun. ACM* **3,** 490–499.

Jacquet, P. and Régnier, M. (1987). Normal limiting distribution of the size of tries. In *Performance '87*, North-Holland, Amsterdam, pp. 209–223.

Jacquet, P. and Szpankowski, W. (1991). Analysis of digital tries with Markovian dependency. *IEEE Trans. Inf. Theory* **37,** 1470–1475.

Jacquet, P. and Szpankowski, W. (1998). Analytical depoissonization and its applications. *Theoret. Comput. Sci* **201,** 1–62.

Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to quicksort. *Random Structures Algorithms* **19,** 498–524.

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

Pittel, B. (1986). Paths in a random digital tree: limiting distributions. *Adv. Appl. Prob.* **18,** 139–155.

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

Rösler, U. (1991). A limit theorem for "Quicksort". *RAIRO Informatique Théoret. Appl.* **25,** 85–100.

Rösler, U. (2001). On the analysis of stochastic divide and conquer algorithms. *Algorithmica* **29,** 238–261.

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

Szpankowski, W. (2001). *Average Case Analysis of Algorithms on Sequences*. Wiley-Interscience, New York.

https://doi.org/10.1239/jap/1152413729 Published online by Cambridge University Press