Probability Metrics and Recursive Algorithms

S. T. Rachev *
University of California
at Santa Barbara

L. Rüschendorf †
University of Freiburg

Abstract

In this paper it is shown by several examples that probability metrics are a useful tool to study the asymptotic behaviour of (stochastic) recursive algorithms. The basic idea of this approach is to find a 'suitable' probability metric which yields contraction properties of the transformations describing the limits of the algorithm. In order to demonstrate the wide range of applicability of this contraction method we investigate examples from various fields, some of them have been analyzed already in the literature.

AMS Classifications: 60 K 35, 60 F 05

Key words: search algorithm, trie, contraction properties, probability metrics, image encoding

Research supported by a DFG Grant and NATO Grant CRG 900798

*postal address: University of California, Statistics Department, Santa Barbara, CA 93106-3110, USA

†postal address: Universität Freiburg, Institut Mathematische Stochastik, Hebelstr. 27, 79104 Freiburg, Germany

1 Introduction

The main methods which have been used in the literature for the asymptotic analysis of algorithms are the transformation methods, in particular those based on moment generating functions (cf. [8], [11], [12], [13], [14], [18]), the martingale method (cf. [15]), the method of branching processes (approximation by Brownian excursions) (cf. [1], [10]), and, for a more restricted class of stochastic algorithms, the method based on stochastic approximation (cf. [21] and references therin).
In this paper we study the application of probability metrics for analyzing the asymptotic behaviour of algorithms. Our method for analysing recursive algorithms consists in the following sequence of steps which together may be called:

<u>Contraction Method:</u>

1. *Find the correct normalization of the algorithms.*
   This is done usually by studying the first moments.

2. *Determine the recursion for the normalized algorithms.*

3. *Determine the limiting form of the normalized algorithms.*
   The limiting equation typically is defined via a transformation T on the set of all probability measures.

4. *Choose an 'ideal' metric μ such that T has good contraction properties w.r.t. μ.*
   This ideal metric has to reflect the structure of the algorithm. It also has to have good bounds in terms of interpretable other metrics and has to allow to estimate bounds (in terms of moments usually). As consequence of 4) we get

5. *The conjectured limiting distribution is the unique fixed point of T.*

   Finally the hope is that the recursion is stable enough so that the contraction in the limit can be pulled over to establish contraction properties of the recursion itself for $n \to \infty$. This is the technically involved step in the analysis.

6. *Establish convergence of the algorithm to the fixed point.*

We will demonstrate this idea in a series of examples. Some of the examples have already been investigated by other methods in the literature; some of the examples are modifications of known examples. The main aim of our choice of different examples is to demonstrate that the contraction method, which is developed in this paper for the first time in the field of stochastic algorithms in its generality, is an important method, supplementing the other methods in the analysis of algorithms.

The following metrics will be used throughout this paper. Define for r.v.'s on a nonatomic probability space and with values in a separable Banach-space U, and

$0 < p < \infty$

(1.1) $$L_p(X,Y) = (E\|X-Y\|^p)^{min(1,1/p)}$$

and let $\hat{L}_p$ denote the corresponding minimal metric

(1.2) $$\hat{L}_p(P_1,P_2) = \inf\{L_p(X,Y); X \overset{d}{=} P_1, Y \overset{d}{=} P_2\}.$$

$\hat{L}_p$ is a simple (r,+)-ideal metric, r=min(1,p), i.e. $\hat{L}_P(X,Y)$ depends only on the marginal distributions of X and Y and

$$\begin{aligned} \hat{L}_p(X+Z,Y+Z) &\leq \hat{L}_p(X,Y) \quad \text{for Z independent of X, Y and} \\ \hat{L}_p(cX,cY) &= |c|^r \hat{L}_p(X,Y). \end{aligned} \tag{1.3}$$

For a simple metric μ we use both notations $\mu(X,Y), \mu(P^X,P^Y)$. For Banach spaces of type $p, 1 \leq p \leq 2$ one can improve the degree of ideality (cf. [25]).

For $r > 0$, $r = m + \alpha, m \in \mathbb{N}$, $0 < \alpha \leq 1$, Zolotarev [36] found the following simple (r,+)-ideal metric

(1.4) $$\zeta_r(X,Y) = \sup\{|E(f(X)-f(Y))|; |f^{(m)}(x) - f^{(m)}(y)| \leq |x-y|^\alpha\},$$

where $f^{(m)}(x)$ denotes the Frechét-derivative of order m. For $s \in \mathbb{N}, 1 \leq p < \infty$, the following metrics $\theta_{s,p}$ were introduced in [16] for r.v.'s with values in $\mathbb{R}^k$.

(1.5) $$\theta_{s,p}(X,Y) = \sup\{|E(f(X)-f(Y))|; \|f^{(s+1)}\|_q \leq 1\},$$

where $\frac{1}{p} + \frac{1}{q} = 1$. For $k = 1$ there is the explicit representation

(1.6) $$\theta_{s,p}(X,Y) = (\int |D_{s,X,Y}(t)|^p dt)^{1/p},$$

where $D_{s,X,Y}(t) = \frac{1}{s!}(E(X-t)_+^s - E(Y-t)_+^s)$. Here $\theta_{s,p}$ is (r,+)-ideal with $r = s + \frac{1}{p}$ (cf. [16]). As limiting case for $p \to \infty$ one obtains for $s \geq 1$,

(1.7) $$\theta_s(X,Y) = \sup_t |D_{s,X,Y}(t)|,$$

which is (s,+)-ideal (cf. [25]). For r.v.'s X, Y, θ on $\mathbb{R}^k, r > 0$, we also consider the convolution type metrics

(1.8) $$\mu_{\theta,r}(X,Y) = \sup_{h \in \mathbb{R}^1} |h|^r \operatorname{ess\,sup}_t |f_{X+h\theta}(t) - f_{Y+h\theta}(t)|,$$

where θ is independent of X and Y, $f_{X+h\theta}$ the density of $X + h\theta$ and

(1.9) $$v_{\theta,r}(X,Y) = \sup_{h \in \mathbb{R}^1} |h|^r V(P_{X+h\theta}, P_{Y+h\theta}),$$

V denoting the total variation distance. The metrics $\mu_{\theta,r}, v_{\theta,r}$ are ideal of order $r-1$ resp. r (cf. [24]).

W.r.t. maxima we shall use the weighted Kolmogorov distance (for k-dimensional random vectors, $k = 1, 2, \ldots, \infty, X, Y$)

(1.10) $$\rho_r(X,Y) = \sup_{x \in \mathbb{R}^k} M(x)^r |F_X(x) - F_Y(x)|, \quad M(x) := \min_{1 \leq i \leq k} |x_i|,$$

which is ideal of order r w.r.t. maxima (cf.(3.31)). In more general spaces one can also use the minimal L_p-metrics w.r.t. maxima (cf. [25]).

2 Contraction of Transformations on Distributions

Stochastic algorithms are typically either directly described by the iterates of a transformation T on the set of all probability distributions on the basic space U or they are asymptotically closely related to the iterations (and thereby to the fixed points) of a transformation T which describes the limiting equation (see the introduction). They can e.g. differ from iterations by a stochastic sequence converging to zero. Examples for recursive algorithms which are asymptotically related to iterations of transformations T are studied in Section 3.
In this section we consider contraction properties of transformations $T : M^1(U) \to M^1(U)$, $M^1(U)$ being the set of probability measures on U supplied with probability metrics as in Section 1. From the fixed point theorems for complete, separable metric spaces we then can infer the convergence of the iterates $(T^n F)$, $F \in M^1(U)$, to a fixed point of T. The examples of transformations dealt with in this section are partially related to the examples in Section 3 on the convergence of certain algorithms. On the other hand some of the examples serve to describe the influence of the choice of the metrics, while others indicate the range of applicability to different fields.

2.1 $TF \stackrel{d}{=} \sum_{i=1}^N a_i(\tau) Y_i + C(\tau)$

For $F \in M^1(U)$ consider an iid sequence $(Y_i)_{1 \le i \le N}$ with distributions F, real random variables $C(\tau), a_i(\tau), \tau$ independent of (Y_i) and define the transformation TF as the law of $\sum_{i=1}^N a_i(\tau) Y_i + C(\tau)$,

$$TF \stackrel{d}{=} \sum_{i=1}^{N} a_i(\tau) Y_i + C(\tau). \tag{2.1}$$

Let us consider Zolotarev's ideal metric ζ_r of order $r > 0$. We obtain for $F, G \in M^1(U)$:

Proposition 1

$$\zeta_r(TF, TG) \le (\sum_{i=1}^{N} E|a_i(\tau)|^r) \zeta_r(F, G). \tag{2.2}$$

Proof: The proof of (2.2) uses the ideality properties of ζ_r namely,

$$\zeta_r(X + Z, Y + Z) \le \zeta_r(X, Y) \quad \text{for Z independent of} \quad X, Y, \tag{2.3}$$

and

$$\zeta_r(cX, cY) = |c|^r \zeta_r(X, Y). \tag{2.4}$$

Let $Y_i \stackrel{d}{=} F, Z_i \stackrel{d}{=} G$ be independent r.v.'s, then with $r = m + \alpha$ (cf. (1.4)).

$$
\begin{aligned}
\zeta_r(TF,TG) &= \sup\{|Ef(\sum a_i(\tau)Y_i + C(\tau)) - Ef(\sum a_i(\tau)Z_i + C(\tau))|; \\
&\quad |f^{(m)}(x) - f^{(m)}(y)| \leq |x-y|^\alpha\} \\
&\leq \int \zeta_r(\sum a_i(t)Y_i + C(t), \sum a_i(t)Z_i + C(t))dP^\tau(t) \\
&\leq \int \zeta_r(\sum a_i(t)Y_i, \sum a_i(t)Z_i)dP^\tau(t) \\
&\leq \sum \int |a_i(t)|^r dP^\tau(t)\zeta_r(Y_i, Z_i) = \sum E|a_i(\tau)|^r \zeta_r(F,G).
\end{aligned}
$$

□

In order to have contraction we need to have

$$
\sum E|a_i(\tau)|^r < 1 \quad \text{and} \quad \zeta_r(F,G) < \infty. \tag{2.5}
$$

The last condition can be established in some cases by the inequality

$$
\zeta_r \leq \frac{\Gamma(1+\alpha)}{\Gamma(1+r)} v_r \ \text{ for } \ v_r(X,Y) = \int \|x\|^r d|P_X - P_Y|(x) \tag{2.6}
$$

the absolute pseudomoment. On Euclidean space and $m \leq r < m+1$, (2.6) requires identical moments for X and Y of order m.

The $\hat{L}_p-$metrics are ideal of order r =min(1,p). We obtain for $F, G \in M^1(U)$, $\ell_p = \hat{L}_p$,

Proposition 2

$$
\ell_p(TF,TG) \leq \sum_{i=1}^N (E|a_i(\tau)|^r)\ell_p(F,G). \tag{2.7}
$$

Proof: Let $Y_i \stackrel{d}{=} F, Z_i \stackrel{d}{=} G, 1 \leq i \leq N$, be independent pairs of random variables with laws F, G and $L_p(Y_i, Z_i) = \ell_p(F,G)$, then

$$
\begin{aligned}
\ell_p(TF,TG) &= \ell_p(\sum a_i(\tau)Y_i + C(\tau), \sum a_i(\tau)Z_i + C(\tau)) \\
&\leq \|\sum a_i(\tau)Y_i - \sum a_i(\tau)Z_i\|_p \\
&\leq \sum_{i=1}^N E|a_i(\tau)|^r \|Y_i - Z_i\|_p = (\sum_{i=1}^N E|a_i(\tau)|^r)\ell_p(F,G).
\end{aligned}
$$

□

So in order to obtain a contraction we need

$$
\sum_{i=1}^N E|a_i(\tau)|^r < 1 \quad \text{and} \quad \ell_p(F,G) < \infty. \tag{2.8}
$$

Under more specific assumptions one can improve (2.8): Let U be a Hilbert space and let F, G have identical first moments, then

Proposition 3

$$\ell_2(TF,TG) \le (\sum_{i=1}^{N} E|a_i(\tau)|^2)^{1/2}\ell_2(F,G). \tag{2.9}$$

Proof: With Y_i, Z_i as in the proof of (2.7),

$$\begin{aligned}\ell_2^2(TF,TG) &\le \|\sum a_i(\tau)(Y_i - Z_i)\|_2^2 \\ &= \sum_{i=1}^{N} E|a_i(\tau)|^2\|Y_i - Z_i\|_2^2 = (\sum E|a_i(\tau)|^2)\ell_2^2(F,G).\end{aligned}$$

□

If the Banach space U is of type p, $1 \le p \le 2$ and F, G have identical first moments (more precisely $E(Y - Z) = 0$ for $Y \stackrel{d}{=} F, Z \stackrel{d}{=} G$), then for $1 \le p \le 2$

$$\ell_p(TF,TG) \le B_p^{1/p}(\sum_{i=1}^{N} E|a_i(\tau)|^p)^{1/p}\ell_p(F,G). \tag{2.10}$$

Here B_p is the constant arising in the Woyczinski inequality (cf. [25]). For $U = L^p$ one can choose the constants $B_1 = 1, B_p = 18p^{3/2}/(p-1)^{1/2}$ for $1 < p \le 2$. The proof of (2.10) is similar to that of (2.9) but we use the Woyczinski inequality (cf. [25]) instead of the Hilbert space structure. If the underlying space is Euclidean we obtain similar contraction properties w.r.t. the metrics $\theta_{s,p}, \theta_s, \mu_{\theta,r}$ and $v_{\theta,r}$.

Example 2.1 *Let $N = 2$, let τ be uniformly distributed on (0,1) and $a_1(\tau) = \tau, a_2(\tau) = 1 - \tau$. Then the contraction factor α is w.r.t.*

$$\begin{aligned}\ell_1 - metric: \alpha &= E\tau + E(1-\tau) = 1, \text{ i.e. no contraction,} \\ \ell_2 - metric: \alpha &= (E\tau^2 + E(1-\tau)^2)^{1/2} = \sqrt{2/3}, \\ \zeta_2 - metric: \alpha &= E\tau^2 + E(1-\tau)^2 = 2/3, \\ \zeta_3 - metric: \alpha &= 1/2. \end{aligned} \tag{2.11}$$

Clearly, if for a probability metric μ on $M^1(U)$ the contraction factor is $\alpha < 1, \mu(TF,TG) \le \alpha\mu(F,G)$, then

$$\mu(T^{n+1}F, T^nF) \le \alpha^n\mu(TF,F), \tag{2.12}$$

i.e. the exponential convergence rate to a fixed point is α^n. In the example above we consider the recursion: $X_{n+1} \stackrel{d}{=} \tau_n X_n + (1-\tau_n)\bar{X}_n + C(\tau_n), \tau_n \stackrel{d}{=} \tau$, where $\bar{X}_n \stackrel{d}{=} X_n$ and $\tau_n, X_n, \bar{X}_n$ independent. The corresponding fixed point equation is: $X \stackrel{d}{=} \tau X + (1-\tau)\bar{X} + C(\tau)$. So under the comparable condition of equal first moments, for the 'ideal' metric ζ_2 the convergence rate is $(2/3)^n$ in comparison to $(2/3)^{n/2}$ for the ℓ_2-metric.

If $a_1(\tau) = \sqrt{\tau}, a_2(\tau) = \sqrt{1-\tau}$, then w.r.t. the ℓ_2-metric the contraction factor is $\alpha = E\tau + E(1-\tau) = 1$, i.e. we have no contraction, and neither do we with ζ_2. But w.r.t. ζ_3 the contraction factor is $\alpha = E\tau^{3/2} + E(1-\tau)^{3/2} = \frac{4}{5} < 1$; so we have a contraction if $\zeta_3(F,G) < \infty$.

2.2 $TF \stackrel{d}{=} \max_{1\le i\le N}\{a_i(\tau)Y_i\}$

For $U = I\!R^k, k = 1,2,\dots,\infty, F \in M^1(U)$ let $(Y_i)_{1\le i\le N}, \tau$ be as in 2.1 and consider

$$TF \stackrel{d}{=} \max_{1\le i\le N}\{a_i(\tau)Y_i\}, a_i(\tau) \ge 0. \tag{2.13}$$

To obtain a contraction of T we shall make use of the metric ρ_r (cf. (1.11)), which is ideal of order r.

Proposition 4

$$\rho_r(TF,TG) \le (\sum E(a_i(\tau))^r)\rho_r(F,G). \tag{2.14}$$

Proof: Let (Z_i) be iid, $Z_1 \stackrel{d}{=} G$, then

$$\begin{aligned}
\rho_r(TF,TG) &= \rho_r(\max_{i\le N}\{a_i(\tau)Y_i\}, \max_{i\le N}\{a_i(\tau)Z_i\}) \\
&= \sup_x |x|^r \left| \int (F_{\max\{a_i(t)Y_i\}}(x) - F_{\max\{a_i(t)Z_i\}}(x)) dP^\tau(t)\right| \\
&\le \int \rho_r(\max\{a_i(t)Y_i\}, \max\{a_i(t)Z_i\}) dP^\tau(t) \\
&\le \sum_i \int a_i(t)^r \, dP^\tau(t)\rho_r(Y_i,Z_i) = (\sum_i Ea_i(\tau)^r)\rho_r(Y_1,Z_1).
\end{aligned}$$

□

For more general maxima we can again use the ℓ_p−metrics. Let $U = L^\lambda(\mu)$, $1 \le \lambda < \infty$, and for $F \in M^1(U)$ consider

$$TF \stackrel{d}{=} \bigvee_{i=1}^N a_i(\tau)Y_i + C(\tau), a_i(\tau) \ge 0, \tag{2.15}$$

where (Y_i) are iid, $Y_1 \stackrel{d}{=} F, \tau$ independent of (Y_i) and $C(\tau)$ has values in U. For any p, λ we have

$$\ell_p(TF,TG) \le (\sum_{i=1}^N Ea_i(\tau)^r)\ell_p(F,G), r = \min(1,p). \tag{2.16}$$

For $1 \le p \le \lambda < \infty$ we have the following improvement:

Proposition 5 *If* $1 \le p \le \lambda < \infty$, *then*

$$\ell_p(TF,TG) \le \sum_{i=1}^N (Ea_i(\tau)^p)^{1/p}\ell_p(F,G). \tag{2.17}$$

Proof: Let $Y_i \stackrel{d}{=} F, Z_i \stackrel{d}{=} G$ satisfy $\|Y_i - Z_i\|^p_{\lambda,\mu} = \ell_p(F,G)^p$, where $\|X\|_{\lambda,\mu} = (\int |X(t)|^\lambda \, d\mu(t))^{1/\lambda}$. Then

$$
\begin{aligned}
\ell_p(TF,TG) &\le (E\| \vee a_i(\tau)Y_i - \vee a_i(\tau)Z_i\|^p_{\lambda,\mu})^{1/p} \\
&= (\int E(\int | \vee a_i(t)Y_i(s) - \vee a_i(t)Z_i(s)|^\lambda \, d\mu(s))^{p/\lambda} \, dP^\tau(t))^{1/p} \\
&\le [\int E(\sum_i \int a_i(t)^\lambda |Y_i(s) - Z_i(s)|^\lambda \, d\mu(s))^{p/\lambda} \, dP^\tau(t))^{1/p} \\
&\le (\int \sum_i E[\int a_i(t)^\lambda |Y_i(s) - Z_i(s)|^\lambda \, d\mu(s)]^{p/\lambda} \, dP^\tau(t))^{1/p} \quad (\text{since } p/\lambda \le 1) \\
&= (\sum_i Ea_i(\tau)E\|Y_i - Z_i\|^p_{\lambda,\mu})^{1/p} = (\sum_i Ea_i(\tau)^p)^{1/p} \, \ell_p(F,G).
\end{aligned}
$$

□

2.3 Bootstrap Estimators

For a separable Banach space U and $F \in M^1(U)$, let $\mu(F) = \int x \, dF(x), \mu_n(F) = \frac{1}{n}\sum_{i=1}^n X_i$, where (X_i) are iid, $X_i \stackrel{d}{=} F$ and F_n is the empirical measure of $X_1, \ldots, X_n$. Then from the strong law of large numbers for any $p > 0$ and $F \in \Gamma_p$, the class of distributions with finite p-th moment, we obtain

$$\ell_p(F_n, F) \to 0 \quad \text{a.s. and} \quad E\ell_p(F_n, F) \to 0 \tag{2.18}$$

(cf. [23], p. 666). Let $X_1^*, \ldots, X_m^*$ be a bootstrap sample, i.e. (X_i^*) are iid, $X_1^* \stackrel{d}{=} F_n$ (conditionally on $X_1, \ldots, X_n$), and $F^*_{n,m}$ is the empirical distribution of $X_1^*, \ldots, X_m^*, m = m(n)$. The condition $\ell_p(F^*_{n,m}, F) \to 0$ a.s. (conditionally on X) is equivalent to

$$\frac{1}{m}\sum_{i=1}^m f(X_i^*) \to \int f dF \quad \text{a.s.}, f \in C_b(U), \tag{2.19}$$

$$\frac{1}{m}\sum_{i=1}^m d^p(X_i^*, a) \to \int d^p(x,y) dF(x), \quad \text{a.s.} \tag{2.20}$$

(cf. [23], p. 651). Therefore, we need a version of the SLLN for real r.v.'s to establish this convergence. For the case $(U,d) = (\mathbb{R}^r, \| \ \|), p > 1$, we can obtain a rate of convergence for the bootstrap approximation. Let $\gamma = kr/[(k-r)(k-2)], k > r, k > 2$ and $\int \|x\|^\gamma F(dx) < \infty$, then

$$E\ell_p^p(F_n, F) \le C(r,k,p) n^{-(1-\frac{1}{p})/k} \tag{2.21}$$

(cf. [23], pp. 667, 668), and thus

$$
\begin{aligned}
E\ell_p^p(F^*_{n,m}, F) &\le 2^p E E_{X_1,\ldots,X_n} \ell_p^p(F^*_{n,m}, F_n) + 2^p C(r,k,p) n^{-(1-\frac{1}{p})/k} \\
&\le C^*(r,k,p)\left(m^{-(1-\frac{1}{p})/k} + n^{-(1-\frac{1}{p})/k}\right).
\end{aligned} \tag{2.22}
$$

If (X_i) are in the domain of an $\alpha-$stable distribution, then it is more natural to choose the bootstrap estimator from a distribution $\tilde{F}_n$, which has a similar tail behaviour to F. Let $\tilde{X}_1, \ldots, \tilde{X}_n$ be a bootstrap sample with

$$\ell_p(\tilde{F}_n, F) \to 0. \tag{2.23}$$

Consider $\tilde{V}_n = \frac{1}{n^{1/\alpha}} \sum_{i=1}^n (\tilde{X}_i - \tilde{E}_{\tilde{F}_n} \tilde{X}_i) \stackrel{d}{=} T_n(\tilde{F}_n)(1 \le \alpha \le 2)$, as a bootstrap estimator of $V_n = \frac{1}{n^{1/\alpha}} \sum_{i=1}^n (X_i - E_F X_i) \stackrel{d}{=} T_n(F)$. Then for a Banach space of type $p, 1 \le p \le 2$, and, $1 \le \alpha \le p$, it follows from (2.10) that

$$\begin{aligned} \ell_p(\tilde{V}_n, V_n) &\le B_p\, n^{\frac{1}{p}-\frac{1}{\alpha}} \ell_p(\tilde{X}_1 - E\tilde{X}_1, X_1 - E_F X_1) \\ &\le B_p\, n^{\frac{1}{p}-\frac{1}{\alpha}} (\ell_p(\tilde{F}_n, F) + |E\tilde{X}_1 - E_F X_1|) \to 0. \end{aligned} \tag{2.24}$$

Since $\ell_p(V_n, Y_{(\alpha)}) \to 0$, and $Y_{(\alpha)}$ a symmetric $\alpha-$stable r.v. in case $\ell_p(X_1, Y_{(\alpha)}) < \infty$, it follows from (2.24) that

$$\ell_p(\tilde{V}_n, Y_{(\alpha)}) \to 0, \tag{2.25}$$

the rate of convergence being of order o $(n^{1/p-1/\alpha})$. In the case of $p = 2$ and $F \in \Gamma_2$ condition (2.23) is satisfied by (2.19) for $\tilde{F}_n = F_n^*$. This case has been considered for Euclidean spaces by Bickel and Freedman (1981). Their investigation of more general functions of the empirical measures can also be extended to our framework.

2.4 Transformation by Markov Kernels, Image Encoding

Let (U, d) be a separable metric space and let $w_i : U \to U, 1 \le i \le N$ be mappings satisfying

$$d(w_i x, w_i y) \le s_i\, d(x, y). \tag{2.26}$$

For a probability distribution $(p_i)_{1 \le i \le N}$ define the Markov kernel

$$K(x, \cdot) = \sum_{i=1}^N p_i \varepsilon_{w_i(x)}, \quad N \le \infty, \tag{2.27}$$

and the corresponding transformation on $M^1(U)$

$$TF = KF, KF(A) = \int K(x, A) F(dx). \tag{2.28}$$

If $f \in LipU$ is a Lipschitz function, $|f(x) - f(y)| \le d(x, y)$, then for $Kf(x) = \sum p_i f \circ w_i(x)$, we have

$$\begin{aligned} |Kf(x) - Kf(y)| &\le \sum p_i |f \circ w_i(x) - f \circ w_i(y)| \\ \le \sum p_i d(w_i(x), w_i(y)) &\le (\sum p_i s_i) d(x, y). \end{aligned} \tag{2.29}$$

Therefore, for the Kantorovich metric we have

$$\mu_L(F, G) = \sup\{|E(f(X) - f(Y))|; X \stackrel{d}{=} F, Y \stackrel{d}{=} G, f \in Lip(U)\}, \tag{2.30}$$

$$\begin{aligned}
\mu_L(TF,TG) &= \sup\{|E(f(K(X,\cdot)-f(K(Y,\cdot)))|; f\in Lip(U)\}\\
&= \sup\{|E(Kf(X)-Kf(Y))|; f\in Lip(U)\}\\
&\leq (\sum p_i s_i)\sup\{|E(g(X)-g(Y))|; g\in Lip(U)\}\\
&= (\sum p_i s_i)\mu_L(F,G). \qquad (2.31)
\end{aligned}$$

If $\sum p_i s_i < 1$, then T is a contraction. By the Kantorovich- Rubinstein theorem μ_L is identical with the minimal L_1-metric and, therefore,

$$\sum p_i s_i < 1 \quad \text{implies} \quad \ell_1(TF,TG) \leq (\sum p_i s_i)\ell_1(F,G). \qquad (2.32)$$

For any $p > 0$ we can extend this result in the following way.

Proposition 6

$$\ell_p(TF,TG) \leq (\sum p_i s_i^p)^{1/p\wedge 1}\ell_p(F,G). \qquad (2.33)$$

Proof: Let $X \stackrel{d}{=} F, Y \stackrel{d}{=} G$ satisfy

$$\ell_p(F,G) = \begin{cases} (E\, d(X,Y)^p)^{1/p}, & \text{if } p\geq 1\\ E\, d(X,Y)^p, & \text{if } p<1.\end{cases}$$

Let $\tilde{I}$ be a random variable with values in $\{1,2,\ldots,N\}$ and distribution (p_i) independent of X,Y. Then for $1 \leq p$,

$$\begin{aligned}
\ell_p^p(TF,TG) &\leq E\, d(w_{\tilde{I}}(X),\ w_{\tilde{I}}(Y))^p\\
&= \sum_{i=1}^N E(d(w_i(X),w_i(Y))^p\ I(\tilde{I}=i))\\
&= \sum_{i=1}^N p_i\ E\, d(w_i(X),w_i(Y))^p\\
&\leq (\sum_{i=1}^N p_i s_i^p) E\, d(X,Y)\\
&= (\sum p_i s_i^p)\ell_p^p(F,G). \qquad (2.34)
\end{aligned}$$

The proof for the case $0 < p < 1$ is similar. □

Remark:

a) Another proof of (2.31) can be given via the dual representation of ℓ_p (cf. [19]). For $0 < p < \infty, p' = 1 \vee p$,

$$\begin{aligned}
\ell_p^{p'}(F,G) = \sup\{\int f dF + \int g dG;\ \ f,g \text{ bounded}\\
\text{continuous}, f(x)+g(y) \leq d^p(x,y), \forall x,y\in U\}. \qquad (2.35)
\end{aligned}$$

Therefore,

$$\begin{aligned} \ell_p^{p'}(TF,TG) &= \sup\{E\,f(K(X,\cdot)) + E\,g(K(Y,\cdot)); \\ &\qquad f(x)+g(y) \le d^p(x,y)\}, \end{aligned} \tag{2.36}$$

where $X \stackrel{d}{=} F, Y \stackrel{d}{=} G$. Since

$$Kf(x) + Kg(y) \le \sum p_i d^p(w_i(x), w_i(y)) \le (\sum p_i s_i^p) d(x,y),$$

we obtain

$$\begin{aligned} \ell_p^{p'}(TF,TG) &= \sup\{E\,Kf(X) + E\,Kg(Y)), f(x)+g(y) \le d^p(x,y)\} \\ &\le (\sum p_i s_i^p)\ell_p^{p'}(F,G). \end{aligned} \tag{2.37}$$

b) Barnsley and Elton (1988) use this Markov chain for the construction of images by so-called iterated function systems (IFS). They establish the existence of a unique attractive invariant measure μ under the assumption

$$\prod_{i=1}^N d(w_i(x), w_i(y))^{p_i} \le r\,d(x,y), \quad r<1, \tag{2.38}$$

which is implied by condition (2.26) with

$$\prod_{i=1}^N s_i^{p_i} < 1. \tag{2.39}$$

In addition they also prove an a.s. convergence result. We note that the condition established in (2.33) for contraction, namely $(\Sigma p_i s_i^p)^{1/p} < 1$ is not directly comparable to (2.39). Note that a.s. convergence is also an immediate consequence of our results since, as is well-known, exponential rates of stochastic convergence imply a.s. convergence.

$\square$

3 Convergence of Recursive Algorithms

In this section we will apply probability metrics to study the asymptotic behaviour of recursive stochastic algorithms with respect to the average case behaviour. There is a large and growing literature on the asymptotic analysis of algorithms. For the general literature we refer to Knuth (1969), Kemp (1984), Mehlhorn (1986), Hofri (1987) and Mahmoud (1992). Typically, in the literature the asymptotic behaviour of the first moments is investigated. There are some exceptions where more information is given on the asymptotic distribution of the algorithms (cf. [9], [10], [12], [15], [22], [29]). The contraction property w.r.t. ℓ_p-metrics is used by

Rösler (1988) to obtain partial information on the asymptotic distribution of the Quicksort algorithm. This is the only example in the literature, as far as we know, where probability metrics are used for the analysis of algorithms.

The basic idea of the application of probability metrics is to transform the recursive equations in such a way that w.r.t. a 'suitable' probability metric one obtains contraction properties in the limit, i.e. we consider sequences $X_n = \tilde{Y}_n + \tilde{W}_n$, such that $(\tilde{Y}_n)$ has suitable contraction properties, while $\tilde{W}_n \to 0$ (see also the intoduction for a more complete description of the contraction method). We shall demonstrate this idea in a series of examples. We start with a simple example.

3.1 Learning Algorithm

Let $Y_1, Y_2, \dots$ be an iid sequence of r.v.'s with values in a separable Banach space and first moment μ and define the recursive sequence: X_1 arbitrary with finite first moment,

$$X_{n+1} = \frac{n}{n+1} X_n + \frac{1}{n+1} Y_{n+1}. \tag{3.1}$$

Proposition 7 *$\zeta_r(X_n, \mu) \to 0$ if $\zeta_r(X_1, \mu) < \infty$.*

Proof: Let $\ell_n = EX_n$, then:

Claim 1: $\ell_n \to \mu$.

Proof. From (3.1), we obtain

$$\begin{aligned} \ell_{n+1} &= \frac{n}{n+1}\ell_n + \frac{1}{n+1}\mu \\ &= \frac{n-1}{n+1}\ell_{n-1} + \frac{2}{n+1}\mu \\ &= \frac{1}{n+1}\ell_1 + \frac{n}{n+1}\mu, \end{aligned} \tag{3.2}$$

the last step following from induction. This implies Claim 1.

Define

$$Z_n = X_n - \ell_n, \quad W_n = Y_n - \mu. \tag{3.3}$$

Then:

$$\begin{aligned} Z_{n+1} + \ell_{n+1} &= \frac{n}{n+1}(Z_n + \ell_n) + \frac{1}{n+1}(W_{n+1} + \mu), \quad \text{(by (3.1))} \\ Z_{n+1} &= Z_n \frac{n}{n+1} + \frac{1}{n+1} W_{n+1} - (\ell_{n+1} - \ell_n \frac{n}{n+1} - \mu \frac{1}{n+1}) \\ &= Z_n \frac{n}{n+1} + W_{n+1} \frac{1}{n+1} \quad \text{(by (3.2))}. \end{aligned} \tag{3.4}$$

Let μ_r be an ideal metric of order r, $1 < r < 2$ and $b_n = \mu_r\ (Z_n, 0)$ (e.g. we can take $\mu_r = \zeta_r$ from (1.4)). Then we have the following:

Claim 2. $\mu_r(Z_n, 0) \to 0$ if $a = \mu_r(W_1, 0) < \infty$.

Proof. We have

$$
\begin{aligned}
b_{n+1} &= \mu_r(Z_{n+1},0) = \mu_r(Z_n \frac{n}{n+1} + W_{n+1}\frac{1}{n+1},0) \\
&\leq \mu_r(\frac{n}{n+1}Z_n,0) + \mu_r(\frac{1}{n+1}W_{n+1},0) \quad \text{(since } Z_n \text{ is independent of } W_{n+1}) \\
&= (\frac{n}{n+1})^r\ \mu_r(Z_n,0) + (\frac{1}{n+1})^r \mu_r(W_{n+1},0) \\
&= (\frac{n}{n+1})^r\ b_n + (\frac{1}{n+1})^r\ a.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
b_{n+1} &\leq (\frac{n}{n+1})^r[(\frac{n-1}{n})^r\ b_{n-1} + (\frac{1}{n})^r\ a] + (\frac{1}{n+1})^r_a \\
&= (\frac{n-1}{n+1})^r\ b_{n-1} + 2(\frac{1}{n+1})^r\ a \\
&\leq (\frac{1}{n+1})^r\ b_1 + n(\frac{1}{n+1})^r\ a, \qquad (3.5)
\end{aligned}
$$

the last inequality following from obvious induction arguments. Since $1 < r, b_n \to 0$.

In particular for $\mu_r = \zeta_r$ we obtain from Claim 1:

$$\zeta_r(X_n,\mu) \to 0 \quad \text{if} \quad \zeta_r(X_1,\mu) < \infty. \qquad (3.6)$$

□

For the case of Euclidean spaces the condition $\zeta_r(Y_1,\mu) < \infty$ is satisfied if Y_1 has a finite absolute r-th moment, $1 < r$. So under the assumption of a finite r-th moment we obtain convergence of X_n to μ. (X_n) is a simple example of a learning algorithm (on μ). Its convergence to μ in the real case can be obtained also by an application of the Robbins-Siegmund lemma (cf. e.g. [30] and [21], p. 206) under the stronger assumption of a finite second moment. In this simple example we can of course directly prove the convergence of X_n to μ under the assumption of a finite first moment. The discussion above shows how to typically use the properties of ideal metrics and shows that in this simple case the method of probability metrics works with weaker assumptions than the method from stochastic approximation based on the Robbins-Siegmund lemma. Some further simple examples of the Robbins-Monroe-type recursion $X_{n+1} = f_n(X_n, Y_{n+1})$ can be treated similarly. Note that our method only needs a metric ideal of order $r > 1$, such that $\mu_r(X_n - \ell_n, 0) \to 0$ implies that $X_n - \ell_n \to 0$ stochastically. The ℓ_p metrics will not work in this example since their degree of ideality is only $r = \min(1,p)$.

3.2 Size of Tries

The trie structure is a basic data structure which is of importance for many tree and stack communications protocols. The size L_n of a trie corresponds to the time under this algorithm to dispose a group of n colliding users. It satisfies the following recursive equations: $L_0 = L_1 = 1$,

$$L_n \stackrel{d}{=} 1 + L_{I_n} + \bar{L}_{n-I_n}, \tag{3.7}$$

where $I_n \stackrel{d}{=} \mathcal{B}(n,p), L_k \stackrel{d}{=} \bar{L}_k$ and $(\bar{L}_k)$ are independent of (L_k). A slot (string) of length n is divided into two parts; each element of the slot (string) is independently distributed to subslot 1 with probability p and to subslot 2 with probability $1-p$. Set

$$\ell_n = EL_n, \quad v_n = \mathrm{Var}(L_n). \tag{3.8}$$

It was proved in [8], [9], [12] based on a Mellin transformation technique that

$$\begin{aligned} \ell_n &= \frac{n}{b(p\log\frac{1}{p} + q\log\frac{1}{q})} G(\log n) + o(n) \\ v_n &= n\, r(p,b) G_{p,b}(\log n) + o(n), \end{aligned} \tag{3.9}$$

where G, $G_{p,b}$ are bounded periodic functions with mean 1 if $\log p/\log q$ is rational and where $G(n) = 1 + O(n^{-\varepsilon}), \varepsilon > 0, G_{p,b}(n) = 1 + O(n^{-\varepsilon}), \varepsilon > 0$, otherwise, and where $r(p,b)$ are constants depending on p, b. Furthermore, based on some involved bounds for the moment generating functions it was proved by Jacquet and Régnier (1988) that the centered and normalized r.v.'s $\frac{L_n - \ell_n}{\sqrt{v_n}}$ converge to the standard normal distribution

$$\frac{L_n - \ell_n}{\sqrt{v_n}} \xrightarrow{\mathcal{D}} N(0,1), \tag{3.10}$$

and, moreover, that all moments converge to the corresponding moments of the normal distribution.

We next show the asymptotic normality of L_n based on some information on the first moments and on the structure of the algorithm by a suitable probability metric. This clarifies a basic feature of our approach to the asymptotic analysis of algorithms. By a suitable choice of the probability metrics we can combine information on some moments with information on the structure of the algorithms to obtain the complete limiting distribution. Our method also allows us to modify the binomial protocol. Let I_n in (3.7) be distributed on $\{0, \dots, n\}$, define

$$Y_n = \frac{L_n - \ell_n}{\sqrt{n}}, \quad \sigma_n^2 = EY_n^2 \tag{3.11}$$

and assume that

$$\begin{aligned} b_r : &= \overline{lim} E((\frac{I_n}{n})^{r/2} + (\frac{n - I_n}{n})^{r/2}) < 1 \\ & \sigma_n^2 \to \sigma^2 \\ & \sup_n E|Y_n|^r < \infty \quad \text{for some} \quad r \in (2,3]. \end{aligned} \tag{3.12}$$

This is fulfilled for L_n in the binomial case by the result of Jacquet and Régnier (1988) if $\log p/\log q$ is not rational. Note that our conditions allow considerable departures from the binomial protocol and establish robustness of the algorithm.

Theorem 3.1 *If (3.12) holds, then*

$$Y_n \xrightarrow{\mathcal{D}} N(0,\sigma^2). \tag{3.13}$$

Proof: From (3.8) one obtains the recursion

$$Y_n \stackrel{d}{=} (\frac{I_n}{n})^{1/2} Y_{I_n} + (\frac{n-I_n}{n})^{1/2} \bar{Y}_{n-I_n} + c_n(I_n), \tag{3.14}$$

where $c_n(I_n) = \frac{1}{\sqrt{n}} + \frac{\ell_{I_n}}{\sqrt{n}} + \frac{\ell_{n-I_n}}{\sqrt{n}} - \frac{\ell_n}{\sqrt{n}}$ and $\bar{Y}_k = \frac{\bar{L}_k - \ell_k}{\sqrt{k}}$. Let $r = 2 + 1/\alpha, 1 \le \alpha \le \infty, \frac{1}{\alpha} + \frac{1}{\beta} = 1 (\beta = \frac{1}{3-r})$ and

$$\mu_r(X,Y) = \vartheta_{1,\alpha}(X,Y) = \sup\{|E(f(X) - f(Y))|; \|f''\|_\beta \le 1\},$$

the ideal metric of order r (cf. (1.5)).

Step I. Let $a_n = \mu_r(Y_n, Z_n)$, where $Z_n \stackrel{d}{=} N(0,\sigma_n^2)$, then

$$a = \sup a_n < \infty. \tag{3.15}$$

Proof. Let

$$\kappa_r(X,Y) = r \int |x|^{r-1} |F_X(x) - F_Y(x)| dx \tag{3.16}$$

denote the difference pseudomoment of order r. Then by a well-known inequality (cf. [16])

$$\mu_r(X,Y) \le \frac{\Gamma(1+\frac{1}{\alpha})}{\Gamma(1+r)} \kappa_r(X,Y) \tag{3.17}$$

and, therefore,

$$\sup a_n \le C \sup_n (E|Y_n|^r + E|Z|^r) < \infty.$$

Define

$$Z_n^* = (\frac{I_n}{n})^{1/2} Z_{I_n} + (\frac{n-I_n}{n})^{1/2} \bar{Z}_{n-I_n} + c_n(I_n), \tag{3.18}$$

then:

Step II. $\limsup_n \mu_r(Z_n^*, Y_n) \le a b_r < \infty, q = 1-p$.

Proof.

$$\begin{aligned}
\mu_r(Z_n^*, Y_n) &\le \sum_{k=0}^n P(I_n = k) \mu_r((\frac{k}{n})^{1/2} Z_k + (\frac{n-k}{n})^{1/2} \bar{Z}_{n-k} + c_n(k), \\
&\quad (\frac{k}{n})^{1/2} Y_k + (\frac{n-k}{n})^{1/2} \bar{Y}_{n-k} + c_n(k)) \\
&\le \sum_{k=0}^n P(I_n = k) [\mu_r((\frac{k}{n})^{1/2} Z_k, (\frac{k}{n})^{1/2} Y_k) \\
&\quad + \mu_r((\frac{n-k}{n})^{1/2} \bar{Z}_{n-k}, (\frac{n-k}{n})^{1/2} \bar{Y}_{n-k})] \\
&\le \sum_{k=0}^n P(I_n = k) \mu_r(Z_k, Y_k) [(\frac{k}{n})^{r/2} + (\frac{n-k}{n})^{r/2}] \\
&\le a[E(\frac{I_n}{n})^{r/2} + E(\frac{n-I_n}{n})^{r/2}]
\end{aligned}$$

and thus $\limsup_n \mu_r(Z_n^*, Y_n) \le ab_r < \infty$, using the ideality of order r of μ_r at various steps.

As a consequence of Step II one obtains (cf. [16])

$$EZ_n^* = EY_n = 0, E(Z_n^*)^2 = EY_n^2 = \sigma_n^2. \tag{3.19}$$

Step III. $b_n = \mu_r(Z_n, Z_n^*) \to 0$.

Proof. Since the mean and the variance of Z_n coincide with those of Z_n^* ($\mu_r(Z_n^*, Y_n) < \infty$ implies $E((Z_n^*)^j - Y_n^j) = 0,\ j = 1, 2$) it follows that $b_n \le c\,\kappa_r(Z_n, Z_n^*)$. Let N_o be a standard normal r.v. independent of (I_n); then

$$\begin{aligned} Z_n^* &= (\frac{I_n}{n})^{1/2} Z_{I_n} + (\frac{n-I_n}{n})^{1/2} \tilde{Z}_{n-I_n} + c_n(I_n), \\ &\stackrel{d}{=} (\frac{I_n}{n}\sigma_{I_n}^2 + \frac{n-I_n}{n}\sigma_{n-I_n}^2)^{1/2} N_o + c_n(I_n) \\ &=: \eta_n N_o + c_n(I_n). \end{aligned} \tag{3.20}$$

From assumption (3.12) we conclude that

$$\eta_n \to (p\sigma^2 + (1-p)\sigma^2)^{1/2} = \sigma. \tag{3.21}$$

Since $Z_n^* \stackrel{d}{=} \eta_n N_o + c_n(I_n)$ has the same mean and variance as $Z_n \stackrel{d}{=} \sigma_n N_o$,

$$\begin{aligned} \sigma_n &= E(Z_n^*)^2 = E(\eta_n N_o + c_n(I_n))^2 \\ &= E\eta_n^2 + E(c_n(I_n))^2. \end{aligned}$$

From $\eta_n \xrightarrow{L^2} \sigma$ we conclude that

$$c_n(I_n) \xrightarrow{p} 0 \tag{3.22}$$

and, therefore, $b_n = \mu_r(Z_n, Z_n^*) \to 0$.

With

$$a_n = \mu_r(Y_n, Z_n) \le \mu_r(Y_n, Z_n^*) + b_n \tag{3.23}$$

and $\bar{a} = \overline{\lim}\, a_n$ we finally obtain from Steps I - III.

Step IV. $\bar{a} = 0$.

Proof. For $\varepsilon > 0$ let $a_k \le \bar{a} + \varepsilon$ for $k \ge n_o$. Then for $n \ge n_o$ by (3.23),

$$\begin{aligned} a_n \ &\le \ \sum_{k=0}^{n_o-1} P(I_n = k)[(\frac{k}{n})^{r/2} + (\frac{n-k}{n})^{r/2}] \sup_{k \le n_o - 1} a_k \\ &+ \ \sum_{k=n_o}^{n} P(I_n = k)[(\frac{k}{n})^{r/2} + (\frac{n-k}{n})^{r/2}](\bar{a} + \varepsilon) + b_n. \end{aligned}$$

Letting $n \to \infty$, we obtain $\bar{a} \le 0 \sup_{0 \le k \le n_o - 1} a_k + b_r(\bar{a} + \varepsilon) + 0 = b_r\bar{a} < \infty$ by Step II.

For $\varepsilon > 0$ let $a_k \leq \bar{a} + \varepsilon$ for $k > n_o = n_o(\varepsilon)$, then as in the proof of the first step

$$\begin{aligned} a_n &\leq \mu_r(Z_n^*, Y_n) + b_n \\ &\leq \Big(\sum_{k=0}^{n_o-1} + \sum_{k=n-n_o}^{n}\Big) P(I_n = k) \sup_{\substack{0\leq k\leq n_o-1 \\ n-n_o\leq k\leq n}} (a_k + a_{n-k})\Big(\big(\frac{k}{n}\big)^{r/2} + \big(\frac{n-k}{n}\big)^{r/2}\Big) \\ &+ \sum_{k=n_o}^{n-n_o-1} P(I_n = k)\Big[\big(\frac{k}{n}\big)^{r/2}(\bar{a}+\varepsilon) + \big(\frac{n-k}{n}\big)^{r/2}(\bar{a}+\varepsilon)\Big] + b_n. \end{aligned}$$

Letting $n \to \infty$, we obtain

$$\begin{aligned} \bar{a} &\leq \limsup\Big(\sum_{k=0}^{n_o-1} + \sum_{k=n-n_o}^{n}\Big) P(I_n = k) 4a \\ &\quad +(\bar{a}+\varepsilon)(p^{r/2} + (1-p)^{r/2}) \\ &= (\bar{a}+\varepsilon)(p^{r/2} + (1-p)^{r/2}). \end{aligned}$$

Since $b_r < 1$, we obtain $\bar{a} = 0$ and thus the proof of the Theorem is complete since μ_r convergence implies weak convergence. $\square$

Remark:

a) As a consequence of this approach one can modify the binomial protocol. If one is able to obtain information on the first moments, then one obtains the full information about the limiting distribution. In the binomial case $I_n \stackrel{d}{=} \mathcal{B}(n,p)$ we have

(3.24) $$b_r = p^{r/2} + q^{r/2} < 1$$

since $r > 2$. We remark that we also can modify the recursive equation (3.7) to

(3.25) $$L_n \stackrel{d}{=} k_n + L_{I_n} + \bar{L}_{n-I_n}, \quad \frac{k_n}{\sqrt{n}} = o(1).$$

If for this scheme one can establish (3.13), then asymptotic normality follows since still $\tilde{c}_n(I_n)$ – the modified $c_n(I_n)$ – converges to zero. From a practical point of view our result suggests normality if we establish empirically linearity of the mean and variance.

b) Asymptotic normality can be expected in the trie example from the 'limiting fixed point equation' of (3.14)

(3.26) $$Y \stackrel{d}{=} \sqrt{p}Y + \sqrt{q}\bar{Y}.$$

c) Numerical properties of the approximations in (3.12), (3.13) will be investigated in a separate paper.

$\square$

3.3 Quicksort

The analysis of quicksort in Rösler (1991) allows a nice comparison to the analysis of Section 3.2. The defining equation is

$$L_n \stackrel{d}{=} n - 1 + L_{I_n} + \bar{L}_{n-I_n}, \tag{3.27}$$

where I_n is uniformly distributed on $\{1, \ldots, n\}$. We repeat briefly the main steps in Rösler (1991). Define $\ell_n = EL_n$; then

$$\ell_n = 2n \log n + n(2\gamma - 4) + 2\log n + 2\gamma + 1 + o(1), \tag{3.28}$$

where γ denotes Euler's constant. Define the normalization

$$Y_n = \frac{L_n - \ell_n}{n} \stackrel{d}{=} \frac{I_n - 1}{n} Y_{I_n - 1} + \frac{n - I_n}{n} \bar{Y}_{n-I_n} + c_n(I_n) \tag{3.29}$$

with $c_n(j) = (n-1)/n + 1/n(\ell_{j-1} + \ell_{n-j} - \ell_n)$. Then with
$c(x) = 2x \log x + 2(1-x)\log(1-x) + 1$,

$$\sup_{x \in (0,1]} |c_n([nx]) - c(x)| \leq \frac{4}{n} \log n + 0(\frac{1}{n}). \tag{3.30}$$

Since $(I_n)/n \to \tau$, the limiting fixed point equation is

$$Y \stackrel{d}{=} \tau Y + (1-\tau)\bar{Y} + c(\tau), \tag{3.31}$$

which gives recursive formulas for the moments of Y. Finally, for any $p \geq 2$ it is proved by Rösler (1991) that

$$\ell_p(Y_n, Y) \to 0. \tag{3.32}$$

In this example the fixed point equation has contraction properties w.r.t. the ℓ_p-metrics, $p \geq 2$ (cf. Example 2.1), which allows one to apply the ℓ_p-metrics for proving convergence of the approximative equation (3.29). An interesting point of this proof is the fact that for the proof one needs only the asymptotic behavior of the first moment of L_n, which is related to the fact that we can use a probability metric which is ideal of order one only. So it is not possible to prove convergence of the size of tries example in Section 3.2 by means of the ℓ_2 metric. To make this difference clear was the reason for the short repeat of Rösler's analysis.

Let φ_Y denote the characteristic function of Y (the solution of (3.31)) and let

$$\psi(z) = \varphi_Y(z) z^{2iz} = \varphi_Y(z) e^{2iz \log z}.$$

Then from (3.31),

$$\varphi_Y(t) = \int_0^1 \varphi_Y(tu)\varphi_Y(t(1-u)) e^{itc(u)} du.$$

Since $e^{itc(u)} = u^{2itu}(1-u)^{2(1-u)it}e^{it}$,

$$\begin{aligned} \psi(t) &= e^{it}\int_0^1 \psi(tu)\psi(t(1-u))du \\ &= e^{it}\int_0^t \psi(z)\psi(t-z)dz. \end{aligned} \tag{3.33}$$

Remark: In equation (3.33) consider the Fourier-transformation:

$\ell(s) := \int_0^\infty e^{isu}\psi(u)du, s \in \mathbb{C}$. From (3.33): $\psi(u)e^{-iu} = \int_0^u \psi(z)\psi(u-z)dz$ and $\int_0^\infty e^{isu}\psi(u)e^{-iu}du = \int_0^\infty e^{isu}(\int_0^u \psi(z)\psi(u-z)dz)du$, i.e. $\ell(s-1) = \ell(s)\ell(s)$.
Let $L(s)$ be the main branch of $\log \ell(s)$; i.e. $L(s) = \text{Log } \ell(s)$; then

$$L(s) = 2L(s+1), s \in \mathbb{C}. \tag{3.34}$$

If we put $L(s) = 2^{\lambda s + \mu}$, then $\lambda = -1$ and we obtain for the Laplace-transform

$$\int_0^\infty e^{-su}\psi(u)du = \exp\{-cz^{-is}\}, c = 2^\mu > 0, s \in \mathbb{C}, Re\ s > 0.$$

If we put more generally $LL(s) = \log L(s) = \log 2 + LL(s+1)$ and $LL(s) = \sum_{n=0}^N a_n s^n, (N = 1, 2, \ldots, \infty)$, then this leads to the system of equations

$$\begin{aligned} a_o &= \log 2 + \sum_{k=0}^N a_k, \\ a_k &= \sum_{n=k}^N a_n \binom{n}{k}, k = 1, 2, \ldots, \end{aligned}$$

which can be investigated numerically. □

3.4 Behaviour of Maxima

A sample of size n is divided into two parts of random size I_n and $n - I_n$, where I_n is a random variable. We consider the recursion of maxima

$$L_n \stackrel{d}{=} c_n + L_{I_n} \vee \bar{L}_{n-I_n}, \tag{3.35}$$

where $L_n, \bar{L}_n$ are independent and identically distributed, (I_n) are independent and (c_n) is a sequence of real numbers.
Given $\alpha > 0$, define the normalizations

$$Y_n = n^{-1/\alpha}L_n, \quad \bar{Y}_n = n^{-1/\alpha}\bar{L}_n. \tag{3.36}$$

By (3.35),

$$Y_n \stackrel{d}{=} c_n n^{-1/\alpha} + (\frac{I_n}{n})^{1/\alpha}Y_{I_n} \vee (\frac{n-I_n}{n})^{1/\alpha}\bar{Y}_{n-I_n}. \tag{3.37}$$

If $n^{-1/\alpha}$ is the right normalization to obtain $Y_n \xrightarrow{\mathcal{D}} Z, \bar{Y}_n \xrightarrow{\mathcal{D}} \bar{Z}$ and $\frac{I_n}{n} \to \tau$, τ a random variable independent of Z, $\bar{Z}$, then in the limit one obtains the fixedpoint equation

$$Z \stackrel{d}{=} \tau^{1/\alpha} Z \vee (1-\tau)^{1/\alpha} \bar{Z}. \tag{3.38}$$

Obviously, the extreme value distribution

$$F_Z(x) = \begin{cases} e^{-x^{-\alpha}}, & x > 0 \\ 0, & x \le 0 \end{cases} \tag{3.39}$$

satisfies (3.38).

For motivation of equation (3.35) consider $L_n = \max\{X_1, \dots, X_n\}, c_n = 0$, where$(X_i)$ are iid of Paretian type ($F(x) \sim 1 - x^{-\alpha}$ for $x \to \infty$). Then by Gnedenko's extreme-value theorem $Y_n = n^{-1/\alpha} L_n \xrightarrow{\mathcal{D}} Z$, with F_Z as in (3.39). Formula (3.35) concerns some modifications of this recursion, where the maxima are produced by a (random) scheme determined by I_n (e.g. $I_n \stackrel{d}{=} \mathcal{B}(n,p)$) and c_n corresponding to some weighting of the number of steps in this reduction (e.g. $c_n = 1$). Note that (3.35) with $c_n = 1$ also describes the maximum search length of a search algorithm dividing a slot of size n succesively into two parts of size I_n and $n - I_n$.

Define $a_k := \ell_r(Y_k, Z)$ for $0 < \alpha < r \le 1$ or $a_k := \rho_r(Y_k, Z), \rho_r$ the weighted Kolmogorov metric (cf. 1.10) for $1 \le \alpha < r < \alpha + 1$, and consider the following assumptions:

$$\begin{gathered} \overline{\lim}\, a_k < \infty, \\ c_n n^{-1/\alpha} \to 0, \\ \frac{I_n}{n} \to \tau \quad \text{a.s. with} \quad E\tau^{r/\alpha} + (1-\tau)^{r/\alpha} < 1. \end{gathered} \tag{3.40}$$

The first condition corresponds to the condition that $n^{-1/\alpha}$ is the right normalization for L_n (as e.g. in the case $L_n = \max(X_1, \dots, X_n)$). If $I_n = \mathcal{B}(n,p)$, then $\frac{I_n}{n} \to p$ and for $\alpha < r, p^{r/\alpha} + (1-p)^{r/\alpha} < 1$.

Theorem 3.2 *Let (L_n) satisfy the recursion (3.35), and define $a_k := \ell_r(Y_k, Z)$ if $0 < \alpha < r \le 1$, or $a_k := \rho_r(Y_k, Z)$ if $0 \le \alpha < r < \alpha + 1$, F_Z as in (3.39).*
Then Assumption (3.40) implies

$$\lim_k a_k = 0.$$

Proof: We consider first the case $0 < \alpha < r \le 1$ and $a_k = \ell_r(Y_k, Z)$.
Let (Z_i) be an iid sequence with common extreme value distribution (3.39) and define

$$Z_n^* = n^{-1/\alpha} c_n + (\frac{I_n}{n})^{1/\alpha} Z_{I_n} \vee (\frac{n - I_n}{n})^{1/\alpha} \bar{Z}_{n - I_n}. \tag{3.41}$$

Then

$$
\begin{aligned}
\ell_r(Y_n, Z_n^*) &= \ell_r((\frac{I_n}{n})^{1/\alpha} Y_{I_n} \vee (\frac{n-I_n}{n})^{1/\alpha} \bar{Y}_{n-I_n}, \\
&\quad (\frac{I_n}{n})^{1/\alpha} Z_{I_n} \vee (\frac{n-I_n}{n})^{1/\alpha} \bar{Z}_{n-I_n}) \\
&\leq \sum_{k=1}^{n} P(I_n = k)[\ell_r((\frac{k}{n})^{1/\alpha} Y_k \vee (\frac{n-k}{n})^{1/\alpha} \bar{Y}_{n-k}, \\
&\quad (\frac{k}{n})^{1/\alpha} Z_k \vee (\frac{n-k}{n})^{1/\alpha} \bar{Z}_{n-k})] \\
&\leq \sum_{k=1}^{n} P(I_n = k)[(\frac{k}{n})^{r/\alpha} a_k + (\frac{n-k}{n})^{r/\alpha} a_{n-k}] \\
&= E(\frac{I_n}{n})^{r/\alpha} a_{I_n} + E(\frac{n-I_n}{n})^{r/\alpha} a_{n-I_n} \qquad (3.42)
\end{aligned}
$$

by the ideality of ℓ_r w.r.t. maxima.
Define $b_n := \ell_r(Z_n^*, Z_n)$ and use the relation

$$
\ell_r(X, Y) \leq (\ell_1(X, Y))^r \qquad (3.43)
$$

for any $r \leq 1$ which is a consequence of the Monge-Kantorovich-theorem (cf. [23]). Note that $\ell_1(X, Y) = \int |F_X(x) - F_Y(x)| dx$.

Claim 1. $b_n \to 0$.

Proof. By (3.41) we obtain

$$
\begin{aligned}
b_n^{1/r} &\leq \ell_1(n^{-1/\alpha} c_n + (\frac{I_n}{n})^{1/\alpha} Z_{I_n} \vee (\frac{n-I_n}{n})^{1/\alpha} \bar{Z}_{n-I_n}, Z_n) \\
&\leq \sum_{k=1}^{n} P(I_n = k) \ell_1(n^{-1/\alpha} c_n + (\frac{k}{n})^{1/\alpha} Z_k \vee (\frac{n-k}{n})^{1/\alpha} \bar{Z}_{n-k}, Z_n) \\
&= \sum_{k=1}^{n} P(I_n = k) \ell_1(c_n n^{-1/\alpha} + (\frac{k}{n} + \frac{n-k}{n})^{1/\alpha} Z, Z),
\end{aligned}
$$

using that for the extreme value distributions Z_n,

$$
(\frac{k}{n})^{1/\alpha} Z_k \vee (\frac{n-k}{n})^{1/\alpha} \bar{Z}_{n-k} \stackrel{d}{=} (\frac{k}{n} + \frac{n-k}{n})^{1/\alpha} Z = Z.
$$

Therefore, $b_n^{1/r} \leq c_n n^{-1/\alpha} \to 0$.

From the triangle inequality and (3.42), (3.43),

$$
a_n \leq \ell_r(Y_n, Z_n^*) + \ell_r(Z_n^*, Z_n), \qquad (3.44)
$$

and, therefore, with $\bar{a} = \overline{\lim}\, a_n$,

$$
\bar{a} \leq \bar{a} E(\tau^{r/\alpha} + (1-\tau)^{r/\alpha}) + 0, \qquad (3.45)
$$

implying that $\bar{a} = 0$.

For the general case $\alpha > 0$ we shall make use of the weighted Kolmogorov metric ρ_r (cf. 1.10). The following inequality for ρ_r follows easily from the definition. For $\varepsilon \geq 1$ and $X, Y \geq 0$,

$$\rho_r(X + a, Y + \varepsilon) \leq (\varepsilon a)^r + \varepsilon^r \rho_r(X, Y). \tag{3.46}$$

Define $a_k = \rho_r(Y_k, Z), b_k = \rho_r(Z_k^*, Z_k)$; then as in (3.41) by (3.46) with $\varepsilon = 1$,

$$\begin{aligned} \rho_r(Y_n, Z_n^*) &\leq c_n^r n^{-r/\alpha} + \rho_r((\frac{I_n}{n})^{1/\alpha} Z_{I_n} \vee (\frac{n - I_n}{n})^{1/\alpha} \bar{Z}_{n-I_n}, \\ &\quad (\frac{I_n}{n})^{1/\alpha} Y_{I_n} \vee (\frac{n - I_n}{n})^{1/\alpha} \bar{Y}_{n-I_n}) \\ &\leq c_n^r n^{-r/\alpha} + E(\frac{I_n}{n})^{r/\alpha} a_{I_n} + E(\frac{n - I_n}{n})^{r/\alpha} a_{n-I_n} \end{aligned}$$

(cf. also (2.14) for a similar derivation). This implies that

$$\overline{\lim}\, \rho_r(Y_n, Z_n^*) \leq E(\tau^{r/\alpha} + (1 - \tau)^{r/\alpha})\bar{a} \quad \text{for} \quad r > \alpha. \tag{3.47}$$

Let now $\alpha < r \leq \alpha + 1$, then as above

$$\rho_r(Z_n, Z_n^*) \leq \sum_{k=1}^{n} P(I_n = k)\rho_r(c_n n^{-1/\alpha} + Z_n, Z_n) = \rho_r(c_n n^{-1/\alpha} + Z, Z).$$

We next prove that $\rho_r(a + Z, Z) \to 0$ for $a \to 0$. Let

$$X(x, a) = x^r |F_Z(x) - F_{Z+a}(x)| = x^r |e^{-x^{-\alpha}} - e^{-(x-a)^{-\alpha}}|.$$

Then $\sup_{0 \leq x \leq a} X(x, a) \leq a^r$, $\sup_{a \leq x \leq 2a} X(x, a) \leq (2a)^r e^{-a^{-\alpha}}$. Furthermore,

$$\sup_{2a \leq x \leq 1} x^r |e^{-x^{-\alpha}} - e^{-(x-a)^{\alpha}}| = \sup_{2a \leq x \leq 1} x^r \alpha \int_{x-a}^{x} y^{-\alpha-1} e^{-y^{-\alpha}}\, dy$$

$$\leq \sup_{2a \leq x \leq 1} x^r \alpha a \frac{1}{(x - a)^{\alpha+1}} e^{-x^{-\alpha}} \leq \sup_{2a \leq x \leq 1} (\frac{x}{x - a})^r \frac{\alpha a}{(x - a)^{\alpha+1-r}} e^{-1} \leq 2^r \alpha a^{r-\alpha} e^{-1}$$

and

$$\sup_{1 \leq x < \infty} x^r \alpha \int_{x-a}^{x} y^{-\alpha-1} e^{-y^{-\alpha}}\, dy \leq \sup_{1 \leq x < \infty} \frac{x^{\alpha+1}}{(x - y)^{\alpha+1}} \alpha a\, e^{-x^{-\alpha}} \leq 2^{\alpha+1} \alpha a.$$

Combining all this we obtain $\rho_r(a + Z, Z) \leq C\, a^{r-\alpha}$ for $\alpha < r \leq \alpha + 1$. (Note that $\sup_x X(x, u) = \infty$. for $r > \alpha + 1$.) From the triangle inequality, therefore,

$$\bar{a} = \overline{\lim}\, \rho_r(Y_n, Z) \leq \overline{\lim}\, \rho_r(Y_n, Z_n^*) + \overline{\lim}\, b_n \leq E(\tau^{r/\alpha} + (1 - \tau)^{r/\alpha})\bar{a},$$

which implies that $\bar{a} = 0$. □

3.5 Environmental Processes

Let (Y_i, Z_i) be a sequence of iid pairs of rv's with values in $U \times I\!R, U$ being a separable Banach space and define a sequence of r.v.'s (S_n) by: $S_o \geq 0$,

$$S_{n+1} = (Y_n + S_n)Z_n. \tag{3.48}$$

This kind of process has found several applications in environmental modelling and has been studied in several papers (cf. [35], [27], [28]). If we write $\tau_n = (Y_n, Z_n)$, and $a(\tau_n) = Z_n, C(\tau_n) = Y_n Z_n$, then

$$S_{n+1} = a(\tau_n)S_n + C(\tau_n), \tag{3.49}$$

so we have a special case of (2.1). Under the condition that $E|a(\tau)|^r < 1$ we obtain contraction of the operator $TS \stackrel{d}{=} a(\tau)S + C(\tau)$ and, therefore, (S_n) converges w.r.t. the ideal metric of order r (in Section 2 we considered ζ_r and ℓ_r) to a fixed point, i.e. a solution of

$$S \stackrel{d}{=} (Y + S)Z. \tag{3.50}$$

Properties of solutions of this equation are discussed in [27], [28]. Note that the results of Section 2 allow us to consider more general recursions than those in (3.48).

3.6 A Random Search Algorithm

Let (T_n) be an independent sequence of random natural numbers, $T_n \leq n-1$. We consider a random search in a set of n ordered states $\{1, 2, \ldots, n\}$, starting in the largest state n. After one step of the search we reach state $T_n \leq n-1$. The search is continued in the smaller set $\{1, \ldots, T_n\}$ in the same way, reaching in the next step the state $T_{T_n} \leq T_n - 1$. The search ends if state one is reached. Let S_n denote the number of steps needed in this random search to reach state 1. Then S_n satisfies the recursion

$$S_n \stackrel{d}{=} 1 + S_{T_n}, \quad S_1 = 1. \tag{3.51}$$

With rv's T_n uniformly distributed on $\{1, \ldots, n-1\}$ this model has been used in [32], p. 118, [3] to estimate the mean number of steps in the simplex method (with n extreme points). For applications to max-search problems cf. [19], [20]. In this case there are given independent rv's $X_1, \ldots, X_n$ and T_n is the largest index $k \leq n-1$ such that $X_k > X_n$ (here we add the index 0 to the state space; $T_n = 0$ meaning that no larger value than X_n occurs).

Define rv's $I_1, \ldots, I_n$, where I_k is defined as 1 or 0 if state k is visited by the search process or not. Then

$$S_n \stackrel{d}{=} \sum_{k=1}^{n} I_k. \tag{3.52}$$

Let $a_i \in [0,1], i \geq 1, a_1 = 1$, and consider the special search strategy

$$P(T_n = k) = (\prod_{m=k+1}^{n-1} b_m)a_k, 1 \leq k \leq n-1, \tag{3.53}$$

where $b_m = 1 - a_m$ and $\prod_{m=n}^{n-1} b_m = 1$.

Special cases:

a) If $a_k = 1/k, b_k = (k-1)/k$, then $\alpha_{n,k} = (\prod_{m=k+1}^{n-1} b_m)a_k = \frac{1}{k}\frac{k}{k+1}\cdots\frac{n-2}{n-1} = \frac{1}{n-1}$, i.e. we obtain uniform search on $\{1,\dots,n-1\}$.

b) If $a_k = 1 - e^{-\alpha_k},\ b_k = e^{-\alpha_k},\ (\alpha_1 = -\infty)$, then $\alpha_{n,k} = e^{-\sum_{m=k+1}^{n-1}\alpha_k}(1 - e^{-\alpha_n})$.

For our choice of the search probabilities in (3.53) it is not difficult to prove that

$$I_1,\dots,I_n \quad \text{are independent,} \quad I_i \stackrel{d}{=} \mathcal{B}(1, a_i). \tag{3.54}$$

This implies that

$$S_n \stackrel{d}{=} \sum_{i=1}^n I_i \tag{3.55}$$

is a sum of independent binomial random variables; in particular

$$ES_n = \sum_{i=1}^n a_i, \quad \mathrm{Var}(S_n) = \sum_{i=1}^n a_i b_i. \tag{3.56}$$

In the uniform search case this implies

$$\begin{aligned} ES_n &= \log n + \gamma + 0(\frac{1}{n}) \\ \mathrm{Var}(S_n) &= \log n + \gamma - \frac{\pi^2}{6} + 0(\frac{1}{n}), \end{aligned} \tag{3.57}$$

$\gamma = 0.5772$ being Euler's constant.
If $\lambda_n = \sum_{i=2}^n a_i$, and $\sum_{i=2}^n a_i^2 / \sum_{i=2}^n a_i = r_n$ is small for $n \to \infty$, then from the results of Deheuvels and Pfeifer (1988) one obtains for the Kolmogorov distance

$$\rho(X,Y) = \sup_x |F_X(x) - F_Y(x)| \tag{3.58}$$

between $S_n - 1$ and a Poisson distributed random variable Z_n with mean λ_n the asymptotic approximation

$$\rho(S_n - 1, Z_n) = \frac{1}{2\sqrt{2\pi e}} r_n + O(\max\left\{\sum_2^n p_i^2/(\sum p_i)^{3/2}, r_n^2\right\}), \tag{3.59}$$

i.e.

$$P(S_n = k+1) = e^{-\lambda_n}\lambda_n^k/k! + O(r_n). \tag{3.60}$$

Some alternative approximations of S_n by different probability metrics were derived in [26].

3.7 Bucket Algorithms, Hybrid Sort

Consider n iid rv's $X_1, \dots, X_n$ with density f on [0,1] and divide [0,1] into m intervals $A_i = [\frac{i-1}{m}, \frac{i}{m}], 1 \le i \le m$. Let $N = (N_1, \dots, N_m)$ be the vector of the number of rv's in the m-buckets $A_1, \dots, A_m$, $N_i = \sum_{j=1}^n 1_{A_i}(X_j)$. The total number of comparisons needed to sort n random numbers by the bucket algorithm is given by (cf. [6])

$$C_n = \sum_{i=1}^m \binom{N_i}{2} = \frac{1}{2}(T_n - n), T_n = \sum_{i=1}^m N_i^2. \tag{3.61}$$

Since N is multinomial $M(m; p_1, \dots, p_m)$ distributed with $p_i = \int_{A_i} f(x)dx$, we obtain

$$EC_n = \frac{n(n-1)}{2} \sum_{i=1}^m p_j^2. \tag{3.62}$$

In the case of a uniform distribution $p_j = 1/m$ and for $m/n \to \alpha \in (0,1)$, therefore, $EC_n \approx n/2\alpha$. In the general case we obtain

$$EC_n \approx \frac{n}{2\alpha} \int_0^1 f^2(x)dx \quad \text{(cf. [6])}, \tag{3.63}$$

$$\begin{aligned} \frac{1}{n}\operatorname{Var} C_n &\to \frac{4}{\alpha^2}\Big(\int f^3(x)dx - \Big(\int f^2(x)dx\Big)^2\Big) + \frac{2}{\alpha}\int f^2(x)dx \\ &\ge \frac{2}{\alpha}\int f^2(x)dx \quad \text{(cf. [6])}. \end{aligned} \tag{3.64}$$

For the case of f not square integrable cf. [6] and [17].

We demonstrate the applicability of probability metrics in order to obtain the asymptotic distribution of C_n in the special case $m = 2$ and $n \to \infty$, leading to $\alpha = 0$ in (3.63). In this case the normalization should be of order $o(n^{-1})$. We have

$$C_n = \frac{1}{n}(T_n - n), \quad T_n = \Big(\sum_{i=1}^n \zeta_i\Big)^2 + \Big(n - \sum_{i=1}^n \zeta_i\Big)^2, \tag{3.65}$$

where ζ_i's are iid Bernoulli random variables with success probability p. Define the approximating U-statistic based on a normal sample

$$D_n = \frac{1}{2}(S_n - n), \quad S_n = \Big(\sum_{i=1}^n \eta_i\Big)^2 + \Big(n - \sum_{i=1}^n \eta_i\Big)^2, \tag{3.66}$$

where $\eta_i \stackrel{d}{=} N(p, pq), q := 1 - p$. A description of the distribution of S_n can be found for example in [34].

By a superscript $\circ$ we denote the normalized quantities, $\overset{\circ}{\zeta}_i = (\zeta_i - p)/\sqrt{pq}$, $\overset{\circ}{\eta}_i = (\eta_i - p)/\sqrt{pq}$, $\overset{\circ}{T}_n = (\sum \overset{\circ}{\zeta}_i)^2 + (n - \sum \overset{\circ}{\zeta}_i)^2$ and so on.

Theorem 3.3 *For $m = 2$ and $n \to \infty$:*

$$\ell_2\left(\frac{\overset{\circ}{C}_n}{n^{3/2}}, \frac{\overset{\circ}{D}_n}{n^{3/2}}\right) = O(n^{-1/4}), \tag{3.67}$$

$$\ell_1\left(\frac{\overset{\circ}{C}_n}{n^{3/2}}, \frac{\overset{\circ}{D}_n}{n^{3/2}}\right) = O(n^{-1/2}). \tag{3.68}$$

Proof: For the proof of (3.67) we first note that the normalization $n^{3/2}$ is of the right order since

$$\begin{aligned}
\mathrm{Var}\left(\frac{\overset{\circ}{D}_n}{n^{3/2}}\right) &= \frac{1}{4}\mathrm{Var}(n^{-3/2}\overset{\circ}{S}_n) \\
&= \frac{1}{4}\mathrm{Var}\left(\frac{2\sum_{i=1}^n \sum_{j=1}^n \overset{\circ}{\eta}_i \overset{\circ}{\eta}_j - 2n\sum \overset{\circ}{\eta}_i}{n^{3/2}}\right) \\
&\approx \text{const. } \mathrm{Var}\left(\frac{2n\sum \overset{\circ}{\eta}_i}{n^{3/2}}\right) \approx \text{const.} > 0.
\end{aligned}$$

Since $\ell_2 = \hat{L}_2$, $\ell_2(\overset{\circ}{C}_n, \overset{\circ}{D}_n) \le L_2(\frac{1}{2}(\overset{\circ}{T}_n - n), \frac{1}{2}(\overset{\circ}{S}_n - n)) = \frac{1}{2}L_2(\overset{\circ}{T}_n, \overset{\circ}{S}_n)$. Thus

$$\begin{aligned}
\ell_2\left(\frac{\overset{\circ}{C}_n}{n^{3/2}}, \frac{\overset{\circ}{D}_n}{n^{3/2}}\right) &\le \frac{1}{2}L_2\left(\frac{\overset{\circ}{T}_n}{n^{3/2}}, \frac{\overset{\circ}{S}_n}{n^{3/2}}\right) \\
&= \frac{1}{2}L_2((2\sum_i\sum_j \overset{\circ}{\zeta}_i\overset{\circ}{\zeta}_j + n^2 - 2n\sum \overset{\circ}{\zeta}_i)n^{-3/2}, \\
&\qquad (2\sum_i\sum_j \overset{\circ}{\eta}_i\overset{\circ}{\eta}_j + n^2 - 2n\sum \overset{\circ}{\eta}_i)n^{-3/2}) \\
&\le L_2(n^{-3/2}\sum_i\sum_j \overset{\circ}{\zeta}_i\overset{\circ}{\zeta}_j, n^{-3/2}\sum_i\sum_j \overset{\circ}{\eta}_i\overset{\circ}{\eta}_j) \\
&+ L_2(n^{-1/2}\sum \overset{\circ}{\zeta}_i, n^{-1/2}\sum \overset{\circ}{\eta}_i) =: I_1 + I_2.
\end{aligned}$$

Assuming that the pairs $(\overset{\circ}{\zeta}_i, \overset{\circ}{\eta}_i)$ are independent, we obtain

$$\begin{aligned}
I_1 &= n^{-3/2}L_2(\sum_i \overset{\circ}{\zeta}{}_i^2, \sum_i \overset{\circ}{\eta}{}_i^2) + n^{-3/2}L_2(\sum_{i\neq j}\overset{\circ}{\zeta}_i\overset{\circ}{\zeta}_j, \sum_{i\neq j}\overset{\circ}{\eta}_i\overset{\circ}{\eta}_j) \\
&\le n^{-1/2}L_2(\overset{\circ}{\zeta}{}_1^2, \overset{\circ}{\eta}{}_1^2) + n^{-3/2}(E(\sum_{i\neq j}(\overset{\circ}{\zeta}_i\overset{\circ}{\zeta}_j - \overset{\circ}{\eta}_i\overset{\circ}{\eta}_j))^2)^{1/2} \\
&= n^{-1/2}L_2(\overset{\circ}{\zeta}{}_1^2, \overset{\circ}{\eta}{}_1^2) + n^{-3/2}(\sum_{i\neq j}(E(\overset{\circ}{\zeta}_i\overset{\circ}{\zeta}_j - \overset{\circ}{\eta}_i\overset{\circ}{\eta}_j))^2)^{1/2} \\
&= n^{-1/2}L_2(\overset{\circ}{\zeta}{}_1^2, \overset{\circ}{\eta}{}_1^2) + n^{-3/2}(n(n-1)L_1(\overset{\circ}{\zeta}_1\overset{\circ}{\zeta}_2, \overset{\circ}{\eta}_1\overset{\circ}{\eta}_2))^{1/2} \\
&\le n^{-1/2}(\mathrm{Var}(\overset{\circ}{\zeta}_1)^2 + \mathrm{Var}(\overset{\circ}{\eta}_1)^2) + n^{-1/2}(\mathrm{Var}(\overset{\circ}{\zeta}_1\overset{\circ}{\zeta}_2 + \mathrm{Var}(\overset{\circ}{\eta}_1\overset{\circ}{\eta}_2)) \le c\,n^{-1/2},
\end{aligned} \tag{3.69}$$

where here and in sequel c stands for an absolute constant, which may be different at different places. Similarly,

$$\begin{aligned}
I_2^2 &= L_2^2(n^{-1/2}\sum \overset{\circ}{\zeta}_i, n^{-1/2}\sum \overset{\circ}{\eta}_i) \\
&\le L_1(n^{-1/2}\sum \overset{\circ}{\zeta}_i, n^{-1/2}\sum \overset{\circ}{\eta}_i)(E|n^{-1/2}\sum \overset{\circ}{\zeta}_i| + E|n^{-1/2}\sum \overset{\circ}{\eta}_i|) \\
&\le L_1(n^{-1/2}\sum \overset{\circ}{\zeta}_i, n^{-1/2}\sum \overset{\circ}{\eta}_i)(\mathrm{Var}\,\overset{\circ}{\zeta}_1 + \mathrm{Var}\,\overset{\circ}{\eta}_1) \\
&= 2\,L_1(n^{-1/2}\sum \overset{\circ}{\zeta}_i, n^{-1/2}\sum \overset{\circ}{\eta}_i). \qquad (3.70)
\end{aligned}$$

Passing to the minimal metric we obtain

$$\ell_2(\overset{\circ}{C}_n, \overset{\circ}{D}_n) \le c\,n^{-1/2} + \sqrt{2}\sqrt{\ell_1(n^{-1/2}\sum \overset{\circ}{\zeta}_i, \overset{\circ}{\eta}_2)}. \qquad (3.71)$$

The rate of convergence in the CLT for the $\ell_1 = \zeta_1$-metric has been discussed by Zolotarev (1986), Theorem 5.4.7 (see also Rachev and Rüschendorf (1990), Lemma 3.3) and is given by

$$\ell_1(n^{-1/2}\sum \overset{\circ}{\zeta}_i, \overset{\circ}{\eta}_1) \le 11.5\max(\ell_1(\overset{\circ}{\zeta}_1, \overset{\circ}{\eta}_1), \zeta_2(\overset{\circ}{\zeta}_1, \overset{\circ}{\eta}_1))n^{-1/2}, \qquad (3.72)$$

(ζ_2 defined in (1.4)). This implies $\ell_2(\overset{\circ}{C}_n, \overset{\circ}{D}_n) \le C\,n^{-1/4}$.

For the proof of (3.68) we can argue similarly and (3.69) can be replaced by

$$I_1 \le n^{-1/2}(E|\overset{\circ}{\zeta}_1|^2 + E|\overset{\circ}{\eta}_1|^2) + n^{-1/2}(E|\overset{\circ}{\zeta}_1\overset{\circ}{\zeta}_2| + E|\overset{\circ}{\eta}_1\overset{\circ}{\eta}_2|) \le c\,n^{-1/2} \qquad (3.73)$$

(now L_2 being replaced by L_1). I_2 is by (3.72) of order $n^{-1/2}$. □

We can extend the results to $m = 3, 4$ by conditioning but the proofs get computationally involved. For some general results on the asymptotic distributions of quadratic forms we refer to [7].

Acknowledgement. We thank the referees for the careful reading of the manuscript and for their suggestions.

References

[1] Aldous, D.: The continuum random tree II: An overview. In: Stochastic Analysis. Eds.: M. T. Barlow and N. H. Bingham, 1991, London Math. Soc. Lecture Notes Series 167, Cambridge University Press, 23-70

[2] Barnsley, M. F. and Elton, J. H.: A new class of of Markov processes for image encoding. Adv. Appl. Prob. 20 (1988), 14 - 32

[3] Bickel, P. and Freedman, D. A.: Some asymptotic theory for the bootstrap. Ann. Statist. 9 (1981), 1196 - 1217

[4] de Haan, L., Rachev, S.: Estimates of the rate of convergence for max-stable processes. Ann. Probab. 17 (1989), 651 - 677

[5] Deheuvels, P. and Pfeifer, D.: On a relationship between Uspensky's theorem and Poisson approximation. Ann. Inst. Statist. Math. 40 (1988), 671 - 681

[6] Devroye, L.: Lecture Notes on Bucket Algorithms. Birkhäuser, Boston, 1986

[7] de Jong, P.: Central Limit Theorems for Generalized Multilinear Forms. CWI Tract 61, 1989, Amsterdam.

[8] Fayolle, G., Flajolet, P., Hofri, M., and Jacquet, P.: Analysis of a stack algorithm for random multiple-access communication. IEEE Trans. Inf. Theory 31 (1985), 244 - 254

[9] Fayolle, G., Flajolet, P. and Hofri, M.: On a functional equation arising in the analysis of a protocol for a multi-access broadcast channel. Adv. Appl. Prob. 18 (1986), 441 - 472

[10] Gutjahr, W. and Pflug, G. Ch.: The asymptotic contour process of a binary tree is a Brownian excursion. Stoch. Process. Appl. 41 (1992), 69 - 89

[11] Hofri, M.: Probabilistic Analysis of Algorithms. Springer, New York, 1987

[12] Jacquet, P., and Regnier, M.: Normal limiting distribution of the size of tries. In: Proc. Performance 87, 209 - 223, North Holland, Amsterdam, 1987

[13] Kemp, R.: Fundamentals of the Average Case Analysis of Particular Algorithms. Wiley, New York, 1984

[14] Knuth, D. E.: The Art of Computer Programming II. Addison-Wesley, 1969

[15] Maejima, M., and Rachev, S. T.: An ideal metric and the rate of convergence to a self similar process. Ann. Probab. 15 (1987), 708 - 727

[16] Mahmoud, H. M.: Evolution of Random Search Trees. Wiley, New York, 1992

[17] Mathar, R. and Pfeifer, D.: Stochastik für Informatiker. Teubner, Stuttgart, 1990

[18] Mehlhorn, K.: Datenstrukturen und effiziente Algorithmen. Vol. I. Teubner, Stuttgart, 1986

[19] Nevzorov, V. B.: Records. Theor. Prob. Appl. 32 (1988), 201 - 228

[20] Pfeifer, D.: Some remarks on Nevzorov's record model. Adv. Appl. Prob. 23 (1991), 823 - 834

[21] Pflug, G.: Stochastische Modelle in der Informatik. Teubner, Stuttgart, 1986

[22] Pittel, B.: Paths in a random digital tree: Limiting distributions. Adv. Appl. Prob. 18 (1986), 139 - 155

[23] Rachev, S. T.: The Monge-Kantorovich mass transference problem and its stochastic applications. Theor. Prob. Appl. 29 (1984), 647 - 676

[24] Rachev, S. T., and Yukich, J. E.: Rates for the CLT via new ideal metrics. Ann. Probab. 17 (1989), 775 - 788

[25] Rachev, S. T., and Rüschendorf, L.: Rate of convergence for sums and maxima and doubly ideal metrics. Theor. Prob. Appl. 37 (1992), 276 - 289

[26] Rachev, S. T., and Rüschendorf, L.: Approximation of sums by compound Poisson distributions with respect to stop-loss distances. Adv. Appl. Prob. 22 (1990), 350 - 374

[27] Rachev, S. T., and Todorovic, P.: On the rate of convergence of some functionals of a stochastic process. J. Appl. Prob. 28 (1990), 805 - 814

[28] Rachev, S. T., and Samorodnitski, G.: Classes of distributions arising in modelling environmental processes. Tech. Report No. 904 (1990), School of Operations Research and Industrial Engineering, Cornell University, Ithaca, N.Y.

[29] Regnier, M., and Jacquet, P.: New results on the size of tries. IEEE Inf. Theory 35 (1989), 203 - 205

[30] Robbins, H. and Siegmund, D.: A convergence theorem for nonnegative almost supermartingales. Optimiz. Meth in Statistics (Rustagi ed). Academic Press, 1971, 233 - 258

[31] Rösler, U.: A limit theorem for Quicksort. Informatique Théorique et Applications 25 (1991), 85 - 100

[32] Ross, S. M.: A simple heuristic approach to simplex efficiency. European J. Oper. Res. 9 (1982), 344 - 346

[33] Ross, S. M.: Stochastic Processes. Wiley, New York, 1983

[34] Seidel, L.: On limit distributions of random symmetric polynomials. Theor. Prob. Appl. 23 (1988), 266 - 278

[35] Vervaat, W.: On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. Adv. Appl. Prob. 11 (1979), 750 - 783

[36] Zolotarev, V. M.: Approximation of distributions of sums of independent random variables with values in infinite dimensional spaces. Theor. Prob. Appl. 21 (1976), 721 - 737

[37] Zolotarev, V. M., and Rachev, S. T.: Rate of convergence in limit theorems for the max scheme. In: Stability Problems for stochastic models. LNM 1155, pp. 415 - 442. Springer-Verlag, Berlin, 1984