

## LIMIT THEOREMS FOR THE INDUCTIVE MEAN ON METRIC TREES

BOJAN BASRAK,\* *University of Zagreb*

### Abstract

For random variables with values on binary metric trees, the definition of the expected value can be generalized to the notion of a barycenter. To estimate the barycenter from tree-valued data, the so-called inductive mean is constructed recursively using the weighted interpolation between the current mean and a new data point. We show the strong consistency of the inductive mean, but also that it, somewhat peculiarly, converges towards the true barycenter with different rates, and asymptotic distributions depending on the small variations of the underlying distribution.

*Keywords:* Limit theorem; metric tree; inductive mean; convergence in distribution; Lindley process

2010 Mathematics Subject Classification: Primary 60F05

Secondary 92B10; 60K10

### 1. Introduction

Metric trees attract considerable interest in the contemporary scientific literature due to their applications in various fields, most notably in computational biology (see, e.g. [12], [13], and [18]) and theoretical computer science (see, e.g. [3]). For some recent developments about the mathematical properties of metric trees, we refer the reader to [4] or [14]. However, rigorous research on probabilistic modeling and statistical inference for data with values on a metric tree seems to be lagging behind the applications. For some references in this direction, see, e.g. [8] and [15]. The main goal of the present paper is to provide the asymptotic analysis of the inductive sample mean on metric trees. Starting from independent and identically distributed (i.i.d.) observations on a tree, the inductive mean can be used to estimate the center of mass or the barycenter of a probability distribution on a metric tree. This was shown by Sturm [17] in more general metric spaces of nonpositive curvature. His work greatly influenced our approach to the subject and remains our main theoretical inspiration.

Sturm [17] extended the familiar laws of large numbers of probability theory from their usual setting in Euclidean or Banach spaces to the case of metric spaces of global nonpositive curvature (or Hadamard spaces, as they are also called). Among other examples, the list of such spaces includes all Hilbert spaces and all complete, simply connected Riemannian manifolds of nonpositive curvature. The barycenters on Riemannian manifolds (as well as the corresponding notions of martingale) were also discussed in [5], [7], and [11].

This class contains two types of metric space that we are particularly interested in: the metric trees and the so-called tripods. The tripod is basically a leafless binary tree which consists of only one vertex and three unbounded edges (rays) connected to it (see Section 2). In the rest of this section we will, somewhat imprecisely, use the term metric tree to refer to both tripods

---

Received 17 August 2009; revision received 16 July 2010.

\* Postal address: Mathematics Department, University of Zagreb, Bijenička 30, Zagreb, Croatia.

Email address: bbasrak@math.hr

and metric trees. We will assume that a metric tree  $T$  consists of the (at most countable) set of vertices  $V$  and the set of edges  $E$ , where each edge in  $E$  can be represented as a line segment of nonzero length (cf. [14]). For simplicity, we only consider locally finite trees. For every two points  $x_0, x_1 \in T$ , there exists a simple path in  $T$  which starts at  $x_0$  and ends at  $x_1$ . In particular, the metric on such trees (and tripods) allows any two points  $x_0, x_1$  to be connected by the unique *geodesic*, that is, a continuous map  $\gamma : [0, 1] \rightarrow NT$  such that  $\gamma(0) = x_0, \gamma(1) = x_1$ , and

$$d(\gamma(r), \gamma(t)) = d(\gamma(r), \gamma(s)) + d(\gamma(s), \gamma(t)) \quad \text{for all } 0 \leq r < s < t \leq 1.$$

Given a probability space  $(\Omega, \mathcal{F}, P)$  and a separable metric space  $N$  with its Borel  $\sigma$ -algebra  $\mathcal{N}$ , a random variable on  $N$  is defined as any measurable map  $X : \Omega \rightarrow N$ . If  $N$  is also a metric tree or a tripod and  $E d(x, X) < \infty$  for at least one  $x \in N$ , it turns out that we can define the expectation of  $X$  as a point  $E X \in N$  such that function

$$z \mapsto E(d^2(z, X) - d^2(y, X)) \tag{1}$$

is minimized at  $z = E X$  for some fixed  $y \in N$ . We can show (see Proposition 4.3 of [17]) that such a point  $z$  always exists and is unique. Moreover, the value  $z$  which minimizes (1) for one  $y$ , does so for all  $y \in N$ . The expectation of the random variable  $X$  is also called the *barycenter* of its associated probability distribution on the metric tree  $(N, \mathcal{N})$ .

The variance of such a random variable is defined as

$$\text{var } X = \inf_{z \in N} E d^2(z, X).$$

If  $\text{var } X < \infty$  then the expectation of  $X$  can be defined as a point  $E X$  which minimizes

$$z \mapsto E d^2(z, X),$$

exactly as in the case of random variables. Hence,  $\text{var } X = E d^2(E X, X)$ .

We could also try to generalize the notion of expectation to the metric trees through the law of large numbers. In other words, we could view the expectation as the limit of the sample mean for an i.i.d. sequence of  $N$ -valued random variables with the same distribution as  $X$ , if such a limit exists. However, it is not completely straightforward to see what a sample mean might be in this context.

For a sequence of  $N$ -valued random variables  $(X_i)_{i \in \mathbb{N}}$ , we can define the sequence of their means  $(\vec{X}_n)_{n \in \mathbb{N}}$  recursively as

$$\vec{X}_1 = X_1, \quad \vec{X}_n = \left(1 - \frac{1}{n}\right)\vec{X}_{n-1} + \frac{1}{n}X_n,$$

where the right-hand side denotes the  $\gamma(1/n)$ -point on the geodesic connecting  $\gamma(0) = \vec{X}_{n-1}$  and  $\gamma(1) = X_n$ . The point  $\vec{X}_n$  is also denoted by  $\vec{X}_n = n^{-1} \sum_{i=1, \dots, n} X_i$  and called the *inductive mean value*  $X_1, \dots, X_n$ . In general, its value depends on the order in the sequence  $(X_i)$ .

There are other, arguably less intuitive or less practical, choices for the sample mean in the global metric spaces of nonpositive curvature. One such alternative was considered in Es-Sahib and Heinich [6], who also showed the corresponding strong law of large numbers and the ergodic theorem. Sturm [17] proved that, for a sequence of i.i.d. random variables with a finite variance and values in any such space  $(N, d)$ , the inductive mean satisfies

$$\vec{X}_n \xrightarrow{L_2} E X_1 \quad \text{as } n \rightarrow \infty. \tag{2}$$

Moreover, if the  $X_i$ s have finite support, he also showed that

$$\vec{X}_n \rightarrow E X_1 \quad \text{almost surely (a.s.) as } n \rightarrow \infty. \quad (3)$$

For a sequence of i.i.d. real-valued random variables  $(X_i)$  with finite variance  $\sigma^2 = \text{var } X_i$ , the standard central limit theorem provides a more subtle result. We know that the difference between the sample means and the expectation multiplied by the square root of the sample size has asymptotically normal distribution. In the remaining part of the paper we show that something rather different happens in the case of metric trees. On binary trees, both the limiting distribution, as well as the rate of convergence, can change substantially.

The rest of the paper is organized as follows. In Section 2 we remove the assumption of finite support needed for (3) and prove the strong law of large numbers on tripods. We also find the asymptotic distribution for inductive means in this case. In Section 3 we first prove the strong law of large numbers for barycenters on general metric binary trees. Finally, by generalizing the results from the tripods, we determine our main result—the asymptotic distribution of the inductive mean on an arbitrary binary tree. Interestingly, we discover a phase transition phenomenon when the distribution changes in such a way that the barycenter becomes one of the vertices. The basic tool used in most of the proofs is the comparison of the paths covered by the inductive sample means with the sample path behaviour of the standard random walk on the real line, and the random walk with negative drift and the barrier at 0. The latter is also well known as the Lindley process in queueing theory, where it is used in conjunction with the popular GI/G/1 model (see Remark 1 below and cf. [1, Chapter X]).

We note that some motivation for our study also comes from the applied perspective. Both rooted and unrooted trees are used as a convenient representation of the evolutionary history in biology. Biological sequences are currently routinely used to reconstruct the phylogenetic relationship among different organisms. An important problem is to find the correct branching point for a new organism in a previously established tree topology connecting a set of organisms. However, different parts of the sequence (or different genes, say) often point towards different locations of the branching point (see [9] or [16]). Sometimes it is desirable to summarize data from different genes by building a so-called consensus tree (see [13, Section 3.6]). By studying our problem we hope to uncover some of the many difficult issues surrounding phylogenetic inference from real-life biological sequence data.

## 2. Auxiliary results on tripods

A tripod is a special case of a *spider*. To define a spider, consider an arbitrary nonempty set  $K$  and, for each of its elements,  $i$  say, define the ray  $N_i = \{(i, x) : x \in [0, \infty)\}$ . The spider is formed by joining the rays together at the origin. In other words, we consider the following set of equivalence classes:

$$N = \{(i, x) : i \in K, x \in [0, \infty)\} / \sim,$$

where  $(i, 0) \sim (j, 0)$  for all  $i, j \in K$ . Therefore, the equivalence class in  $N$  consisting of all points of the form  $(i, 0)$ ,  $i \in K$ , we denote by  $\mathbf{0}$ , and call it the origin, while all the other points  $x = (i, x)$ ,  $x > 0$ , form an equivalence class on their own. If  $K$  has  $k \in \mathbb{N}$  elements, we say that  $N$  is a  $k$ -spider. The tripod is just the 3-spider, while the 2-spider is isomorphic to the real line.

The distance between the points on the tripod (or any spider in general) is calculated as

$$d((i, x), (j, y)) = \begin{cases} |x - y| & i = j, \\ |x| + |y| & \text{otherwise.} \end{cases}$$

Moreover, a binary operation can be added to a spider in the following way:

$$(i, x) + (j, y) = \begin{cases} (i, x + y) & i = j, \\ (i, x - y) & x \geq y, i \neq j, \\ (j, y - x) & \text{otherwise.} \end{cases} \tag{4}$$

On the 2-spider, this binary operation corresponds to the addition of real numbers. However, on general  $k$ -spiders,  $k \geq 3$ , this operation is commutative but not associative. The origin  $\mathbf{0}$  is clearly the neutral element for this operation, but each other point has exactly  $k - 1$  inverses on the  $k$ -spider. In the sequel, as it turns out, it will be important to understand the behaviour of the random walk on this unusual space.

For  $(i, x) \in N$  and  $a \geq 0$ , it is convenient to write  $a(i, x) = (i, ax)$ . The binary operation ‘+’ introduced above allows us to define the inductive mean of the i.i.d. sequence  $(X_i)$  with values on the tripod recursively as

$$\vec{X}_0 = \mathbf{0}, \quad \vec{X}_n = n^{-1}[(n - 1)\vec{X}_{n-1} + X_n], \quad n \geq 1.$$

Another useful process in the sequel is the sequence

$$\vec{S}_n = n\vec{X}_n = \vec{S}_{n-1} + X_n, \quad n \geq 0.$$

We could view  $(\vec{S}_n)$  as the corresponding random walk on the space  $N$  with respect to the nonassociative addition (4).

Suppose that there exists  $E X_1 \in N$ . For each ray  $N_j$ ,  $j = 1, 2, 3$ , we define  $\mu_j$  to be the expectation of  $X_1$  restricted to  $N_j$ . More precisely, consider the projections

$$f_j(\mathbf{x}) = d(\mathbf{0}, \mathbf{x})\mathbf{1}_{N_j}(\mathbf{x}). \tag{5}$$

That is, for  $\mathbf{x} = (i, x)$ , we set  $f_j(\mathbf{x}) = f_j((i, x)) = x$  for  $i = j$  and 0 otherwise, and then  $\mu_j = E f_j(X_1)$ . The values  $\mu_j$  are finite since  $E d(\mathbf{0}, X) = \mu_1 + \mu_2 + \mu_3 < \infty$ . Observe that the three projections  $f_j$ ,  $j = 1, 2, 3$ , uniquely determine the position of the point on the tripod. We will also need the following real functions on  $N$  in the arguments later on:

$$g_j(\mathbf{x}) = g_j((i, x)) = (-1)^{1+\mathbf{1}_{N_j}(\mathbf{x})}d(\mathbf{0}, \mathbf{x}) = \begin{cases} x & i = j, \\ -x & \text{otherwise,} \end{cases} \tag{6}$$

for  $j = 1, 2, 3$ . Without loss of generality, we assume in the sequel that

$$\mu_1 \geq \mu_2 \geq \mu_3 > 0.$$

The case  $\mu_3 = 0$  is not considered here, because in that case  $P(X_1 \in N_3 \setminus \{\mathbf{0}\}) = 0$ , and, therefore, the  $X_i$ s are essentially concentrated on the real line, where the usual strong law of large numbers and the central limit theorem apply. Define  $a_+ = \max\{a, 0\}$  for any real number  $a$ .

**Lemma 1.** *For every two elements  $\mathbf{x}$  and  $\mathbf{y}$  of the tripod  $N$  and any  $j = 1, 2, 3$ , the following inequalities hold:*

$$g_j(\mathbf{x}) + g_j(\mathbf{y}) \leq g_j(\mathbf{x} + \mathbf{y}) \leq f_j(\mathbf{x} + \mathbf{y}) \leq (f_j(\mathbf{x}) + g_j(\mathbf{y}))_+.$$

*Proof.* The statement follows directly from (5) and (6) and the definition of the binary operation ‘+’ on the tripod; see (4). To prove the rightmost inequality, observe that

$$\begin{aligned}
 f_j(\mathbf{x} + \mathbf{y}) &= d(\mathbf{0}, \mathbf{x} + \mathbf{y})\mathbf{1}_{N_j}(\mathbf{x} + \mathbf{y}) \\
 &= \begin{cases} d(\mathbf{0}, \mathbf{x}) + d(\mathbf{0}, \mathbf{y}), & \mathbf{x}, \mathbf{y} \in N_j, \\ (d(\mathbf{0}, \mathbf{x}) - d(\mathbf{0}, \mathbf{y}))_+, & \mathbf{x} \in N_j, \mathbf{y} \notin N_j, \\ (d(\mathbf{0}, \mathbf{y}) - d(\mathbf{0}, \mathbf{x}))_+, & \mathbf{x} \notin N_j, \mathbf{y} \in N_j, \\ 0, & \text{otherwise.} \end{cases}
 \end{aligned}$$

Each of these four expressions is less than or equal to

$$(d(\mathbf{0}, \mathbf{x})\mathbf{1}_{N_j}(\mathbf{x}) + d(\mathbf{0}, \mathbf{y})(-1)^{1+\mathbf{1}_{N_j}(\mathbf{y})})_+ = (f_j(\mathbf{x}) + g_j(\mathbf{y}))_+.$$

This proves the third inequality, and the other two inequalities are proved in a similar way.

**Remark 1.** For a given i.i.d. sequence of random variables  $(Y_n)$ , by an associated Lindley process, we understand a stochastic process of the form

$$W_0 = 0, \quad W_n = (W_{n-1} + Y_n)_+, \quad n \in \mathbb{N}.$$

It is known that  $W_n \stackrel{D}{=} \max_{k=0, \dots, n} Y_1 + \dots + Y_k$ , where ‘ $\stackrel{D}{=}$ ’ denotes equality in distribution. Also, the long-term behaviour of the process  $(W_n)$  is well understood (see [1, Chapter 10, Propositions 1.1 and 1.2]). For instance, if  $E Y_1 < 0$  then a limiting stationary distribution exists for the sequence  $(W_n)$ . On the other hand, if  $\mu = E Y_1 > 0$  then  $W_n/n \rightarrow \mu$  a.s. Finally, if  $E Y_1 = 0$  and  $\sigma^2 = \text{var } Y_1$  is finite and greater than 0, then the sequence  $W_n/\sqrt{n}$  converges in distribution to a random variable of the form  $|W|$ , where  $W$  represents a normal random variable with mean 0 and variance  $\sigma^2$  (cf. the discussion of the reflection principle in [2]).

As a consequence of Lemma 1, the random sequences  $(f_j(\vec{S}_n)_n)$  are dominated by the Lindley processes (or the half-line random walks)

$$S_0^j = 0, \quad S_n^j = (S_{n-1}^j + g_j(X_n))_+, \quad n \geq 0, \tag{7}$$

for each  $j = 1, 2, 3$  and  $n \geq 0$ .

Our first goal in this section is to prove the strong law of large numbers under the condition that the first moment  $E d(z, X_i)$  is finite for some value  $z \in N$ , thus extending result (3) established by Sturm [17]. It is well known that in the case of real-valued random variables (and, therefore, 2-spiders) this condition is both necessary and sufficient. If we suppose that the expectation  $E X_1$  exists on  $N$ , then, by Proposition 5.7 of [17], it follows that  $E X_1 = (1, \mu_1 - \mu_2 - \mu_3)$  if  $\mu_1 > \mu_2 + \mu_3$ , and  $E X_1 = \mathbf{0}$  otherwise.

**Proposition 1.** *Suppose that  $(X_i)$  is a sequence of i.i.d.  $N$ -valued random variables with finite expectation  $E X_1 \in N$ . Then, as  $n \rightarrow \infty$ ,*

$$\vec{X}_n \rightarrow E X_1 \quad \text{a.s.}$$

*Proof.* We split the proof into two cases depending on the value  $\mu = \mu_1 - \mu_2 - \mu_3 = E g_1(X_i)$ . Recall that we assumed that  $\mu_1 \geq \mu_2 \geq \mu_3 > 0$ , and consider the processes  $f_1(\vec{S}_n)$  and  $(S_n^1)$  (see (7)).

Suppose first that  $\mu = E g_1(X_i) > 0$ . By the inequality in Lemma 1,

$$\frac{1}{n} \sum_{i=1}^n g_1(X_i) \leq \frac{1}{n} f_1(\vec{S}_n) \leq \frac{1}{n} S_n^1.$$

The first and the last sequences in the inequality above converge a.s. to  $\mu$  by standard results on the (half-line) random walks (see Remark 1). Hence,  $\vec{X}_n = \vec{S}_n/n \rightarrow E X_1 = (1, \mu)$  a.s. as well.

Suppose that  $\mu \leq 0$ . Then it is known that  $S_n^j/n \rightarrow 0$  a.s. for  $j = 1, 2, 3$ ; see Remark 1. Since, for each  $n \in \mathbb{N}$ ,

$$0 \leq \frac{1}{n} f_j(\vec{S}_n) \leq \frac{1}{n} S_n^j,$$

the same holds for the sequence  $f_j(\vec{S}_n)$ . In particular, it follows that

$$d(\vec{X}_n, \mathbf{0}) = \frac{d(\vec{S}_n, \mathbf{0})}{n} = \frac{1}{n}(f_1(\vec{S}_n) + f_2(\vec{S}_n) + f_3(\vec{S}_n)) \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

This completes the proof.

Denote by  $Z$  a standard normal random variable, and by  $\Phi$  its distribution function. The following theorem describes the asymptotic behaviour of the inductive mean on tripods.

**Theorem 1.** *Suppose that  $(X_i)$  is a sequence of i.i.d.  $N$ -valued random variables with finite expectation  $E X_1$ .*

(i) *If  $\mu_1 > \mu_2 + \mu_3$  and the  $X_i$  are of finite variance  $\sigma^2 = \text{var } X_1 > 0$ , then*

$$\frac{\sqrt{n}}{\sigma} (f_1(\vec{X}_n) - f_1(E X_1)) \xrightarrow{D} Z, \tag{8}$$

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

(ii) *If  $\mu_1 = \mu_2 + \mu_3$  and the  $X_i$  are of finite variance  $\sigma^2 = \text{var } X_1 > 0$ , then*

$$\frac{\sqrt{n}}{\sigma} f_1(\vec{X}_n) \xrightarrow{D} |Z|. \tag{9}$$

(iii) *If  $\mu_1 < \mu_2 + \mu_3$  then the sequence  $(n \vec{X}_n)_n$  is tight in distribution.*

*Proof.* To show that (8) holds, observe that if, for some  $x, y \in N$  and  $a, b \geq 0$ , both  $x$  and  $ax + by$  belong to  $N_1$ , then

$$f_1(ax + by) = af_1(x) + bf_1(y). \tag{10}$$

The sequence  $Y_n = n^{-1} \sum_{i=1}^n g_1(X_i)$ ,  $n \geq 1$ , clearly satisfies  $Y_n \rightarrow \mu$  a.s. Since, by Proposition 1,  $\vec{X}_n \rightarrow E X_1$  a.s., with probability 1, there exists a random integer  $n_0$  such that, for all

$n \geq n_0, Y_n \geq 0$  and  $\vec{X}_n \in N_1$ . For all such  $n$ , applying (10) iteratively we obtain

$$\begin{aligned} & \frac{\sqrt{n}}{\sigma} (f_1(\vec{X}_n) - Y_n) \\ &= \frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} f_1(\vec{S}_{n_0}) + \frac{1}{n} \sum_{i=n_0+1}^n g_1(X_i) - \frac{1}{n} n_0 Y_{n_0} - \frac{1}{n} \sum_{i=n_0+1}^n g_1(X_i) \right) \\ &= \frac{1}{\sqrt{n}\sigma} (f_1(\vec{S}_{n_0}) - n_0 Y_{n_0}) \\ &\rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty. \end{aligned}$$

Furthermore, the sequence  $(Y_n)$  satisfies the standard central limit theorem

$$\frac{\sqrt{n}}{\sigma} (Y_n - \mu) \xrightarrow{D} N(0, 1) \quad \text{as } n \rightarrow \infty.$$

Therefore, the first statement of the theorem follows.

(ii) To prove (9), consider again the random walk  $\vec{S}_n = n \vec{X}_n, n \geq 0$ . For  $j = 1, 2, 3$ , recall that the sequences

$$f_j(\vec{S}_n) = d(\vec{S}_n, \mathbf{0}) \mathbf{1}_{N_j}(\vec{S}_n), \quad n \geq 0,$$

are respectively dominated by the Lindley processes  $(S_n^j)$ , since  $E g_j(X_n) < 0$  for  $j = 2$  or  $3$  and the corresponding Markov chains  $(S_n^j)_n$  converge in distribution; see Remark 1. This implies that the sequences  $(f_j(\vec{S}_n))_n$  and  $(S_n^j)_n$  for  $j = 2$  or  $3$  are all tight in distribution. In particular, it follows that

$$\frac{1}{\sqrt{n}} f_j(\vec{S}_n) \leq \frac{1}{\sqrt{n}} S_n^j \xrightarrow{P} 0$$

as  $n \rightarrow \infty$  for  $j = 2, 3$ , where ‘ $\xrightarrow{P}$ ’ denotes convergence in distribution.

Now consider the sequence  $(f_1(\vec{S}_n))_n$ . By Remark 1,

$$\limsup_{n \rightarrow \infty} P(f_1(\vec{S}_n) > a\sigma\sqrt{n}) \leq \limsup_{n \rightarrow \infty} P(S_n^1 > a\sigma\sqrt{n}) = P(|Z| > a) = 2(1 - \Phi(a)).$$

By the same remark, the process  $(S_n^1)_n$  satisfies

$$S_n^1 \stackrel{D}{=} \max \left\{ 0, \sum_{i=k}^n g_1(X_i) : k = 1, \dots, n \right\}. \tag{11}$$

To prove the appropriate lower bound, note that, for  $\mathbf{x} + \mathbf{y} \in N_1$ ,

$$f_1(\mathbf{x} + \mathbf{y}) = \begin{cases} -\max_{j=2,3} f_j(\mathbf{x}) + g_1(\mathbf{y}), & \mathbf{x} \notin N_1, \\ f_1(\mathbf{x}) + g_1(\mathbf{y}), & \mathbf{x} \in N_1. \end{cases} \tag{12}$$

For any  $\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_n \in N$ ,

$$f_1(\mathbf{x} + \mathbf{y}_1 + \dots + \mathbf{y}_n) \geq -\max_{j=2,3} f_j(\mathbf{x}) + \sum_{i=1}^n g_1(\mathbf{y}_i),$$

which can be shown by induction: the case  $n = 1$  follows from (12), while the induction step follows from

$$-\max_{j=2,3} f_j(\mathbf{x} + \mathbf{y}) \geq -\max_{j=2,3} f_j(\mathbf{x}) + g_1(\mathbf{y}).$$

Now take an arbitrary  $\varepsilon > 0$ , and observe that

$$P(f_1(\vec{S}_n) > a\sigma\sqrt{n}) \tag{13}$$

$$\begin{aligned} &\geq P\left(\bigcup_{k=1}^n \left\{ \sum_{i=k}^n g_1(X_i) > (a + \varepsilon)\sigma\sqrt{n}, \max_{j=2,3} f_j(\vec{S}_{k-1}) < \varepsilon\sigma\sqrt{n} \right\}\right) \\ &\geq P\left(\bigcup_{k=1}^n \left\{ \sum_{i=k}^n g_1(X_i) > (a + \varepsilon)\sigma\sqrt{n}, \max_{j=2,3} S_{k-1}^j < \varepsilon\sigma\sqrt{n} \right\}\right). \end{aligned} \tag{14}$$

Define  $\kappa = \sup\{j : j = 1, \dots, n, \sum_{i=j}^n g_1(X_i) > (a + \varepsilon)\sigma\sqrt{n}\}$ . Now, the probability in (14) is greater than or equal to

$$P\left(\bigcup_{j=1}^n \{\kappa = j\} \cap \left\{ \max_{j=2,3} S_{j-1}^j < \varepsilon\sigma\sqrt{n} \right\}\right) = \sum_{j=1}^n P\left(\{\kappa = j\} \cap \left\{ \max_{j=2,3} S_{j-1}^j < \varepsilon\sigma\sqrt{n} \right\}\right).$$

Since the events on the right-hand side are independent for each  $j$ , the last sum is equal to

$$\begin{aligned} &\sum_{j=1}^n P(\kappa = j) \left[ 1 - P\left(\max_{j=2,3} S_{j-1}^j > \varepsilon\sigma\sqrt{n}\right) \right] \\ &\geq \sum_{j=1}^n P(\kappa = j) - P(S_n^2 > \varepsilon\sigma\sqrt{n}) - P(S_n^3 > \varepsilon\sigma\sqrt{n}) \\ &= P(S_n^1 > (a + \varepsilon)\sigma\sqrt{n}) - P(S_n^2 > \varepsilon\sigma\sqrt{n}) - P(S_n^3 > \varepsilon\sigma\sqrt{n}), \end{aligned}$$

where the equality follows from (11).

Since, for  $j = 2, 3$ , the Markov chains  $(S_n^j)_n$  are tight, it holds that, as  $n \rightarrow \infty$ ,

$$P(S_n^j > \varepsilon\sigma\sqrt{n}) \rightarrow 0.$$

Coming back to the probability in (13) and calculating  $\liminf_{n \rightarrow \infty}$ , we obtain the bound  $P(|Z| > a + \varepsilon) = 2(1 - \Phi(a + \varepsilon))$  (again we refer to Remark 1). The statement is proved once we note that  $\varepsilon > 0$  was arbitrary and  $\Phi$  is continuous.

(iii) Since  $E g_j(X_n) < 0$  for  $j = 1, 2, 3$ , all three Markov chains  $(S_n^j)_n$ ,  $j = 1, 2, 3$ , in (7) are tight by the argument above. Hence, the same holds for the sequences  $(f_j(\vec{S}_n))_n$  and, therefore, for the sequence  $(n\vec{X}_n)_n$  as well, just note that  $n\vec{X}_n = \vec{S}_n = \sum_{j=1}^3 (j, f_j(\vec{S}_n))$ . This completes the proof.

There is an alternative, and perhaps more illustrative way, to express the first two statements of the theorem. For any  $a \in \mathbb{R}$ , define

$$(i, x) + a = (i, x + a) \quad \text{if } -a \leq x,$$

where, for convenience, we can put  $(i, x) + a = \mathbf{0}$  if  $-a > x$ . If  $\mathbf{x} = (i, x)$  and  $\mathbf{y} = (i, y)$  for some  $0 \leq x < y \leq \infty$ , we denote the open interval on  $N_i$  by

$$(\mathbf{x}, \mathbf{y}) = \{(i, z) : x < z < y\}.$$

Now, we can reformulate (8) by saying that, for each  $a < b \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} P\left(\vec{X}_n \in \left(E X_1 + \frac{a\sigma}{\sqrt{n}}, E X_1 + \frac{b\sigma}{\sqrt{n}}\right)\right) = \Phi(b) - \Phi(a)$$

as  $n \rightarrow \infty$ , and (9) as

$$\lim_{n \rightarrow \infty} P\left(\vec{X}_n \in \left(\left(1, \frac{a\sigma}{\sqrt{n}}\right), (1, \infty)\right)\right) = 2(1 - \Phi(a)) \quad \text{for each } a \geq 0.$$

It is actually the third statement of the theorem that is, maybe, the most surprising, since in that case the rate of convergence drastically improves and the limiting distribution, if it exists, changes as well (cf. Proposition 2 and the Example 1, below).

Note that, in general, we cannot expect to have a limiting distribution in Theorem 1(iii). Consider, for instance, the case when the law of random variables  $X_i$  puts equal mass at points  $(j, 1)$ ,  $j = 1, 2, 3$ . Then the sequence  $(\vec{S}_n)_n = (n \vec{X}_n)_n$  is just a simple random walk on  $N$ . Like the standard simple symmetric random walk on  $\mathbb{Z}$ ,  $(n \vec{X}_n)_n$  is clearly a periodic Markov chain with period 2 and, therefore, it does not have a limiting distribution.

In the rest of the section we consider Theorem 1(iii) in detail. We rely on the results in Chapter 14 of [10] to establish the existence of the stationary distribution and ergodic properties of the process  $(\vec{S}_n)$  (cf. [10, Subsection 14.4.1] in particular). We write  $\vec{S}_n = (i_{S_n}, S_n) \in N$ . Observe that the sequence  $(\vec{S}_n)$  is a Markov chain satisfying

$$\begin{aligned} \vec{S}_n &= \vec{S}_{n-1} + X_n \\ &= (i_{S_{n-1}}, (S_{n-1} + g_{i_{S_{n-1}}}(X_n))_+) + (i_{X_n}, (S_{n-1} + g_{i_{S_{n-1}}}(X_n))_-) \end{aligned} \tag{15}$$

for  $n \geq 0$ , with  $\vec{S}_0 = \mathbf{0}$ . We denote by  $(P(\mathbf{x}, \cdot)_{\mathbf{x} \in N})$  its transition probability kernel, and by  $P_{\mathbf{x}}$  the law of this process, conditionally that it starts at the value  $\mathbf{x}$ . Finally, by  $\tau(C)$  we denote the return time to the set  $C$  of the Markov chain  $(\vec{S}_n)$ , that is,

$$\tau(C) = \inf\{t \geq 1: \vec{S}_t \in C\}.$$

Recall that the Markov chain  $(\vec{S}_n)$  is  $\psi$ -irreducible for a measure  $\psi$  on a measurable space  $(N, \mathcal{N})$ , if  $P_{\mathbf{x}}(\tau(A) < \infty) > 0$  for each  $\mathbf{x}$  and set  $A$  such that  $\psi(A) > 0$ , and  $\psi$  is the maximal irreducibility measure.

It is immediate from (15) that, for  $n \geq 1$ ,

$$S_n = |S_{n-1} + g_{i_{S_{n-1}}}(X_n)|.$$

Recall that, for a signed measure  $\nu$  and a measurable function  $f: N \rightarrow [1, \infty)$ , the  $f$ -norm of  $\nu$  is defined as

$$\|\nu\|_f = \sup_{\{g: g \leq f\}} |\nu(g)|.$$

The following proposition follows from the Lyapunov–Foster-type drift condition applied to the Markov chain  $(\vec{S}_n)$ .

**Proposition 2.** *Assume that  $\mu_1 < \mu_2 + \mu_3$  and that  $E d(\mathbf{0}, X_i)^{k+1} < \infty$  for some  $k \in \mathbb{N}_0$ . If the Markov chain  $(\vec{S}_n)$  is aperiodic and  $\psi$ -irreducible, then*

- (a) *it admits a stationary measure  $\pi$  such that  $\int x^k d\pi(\mathbf{x}) < \infty$ ;*
- (b) *for  $f(\mathbf{x}) = x^k$ , as  $n \rightarrow \infty$ ,  $\|P(\vec{S}_n \in \cdot) - \pi(\cdot)\|_f \rightarrow 0$ .*

*Proof.* Note that if  $E d(\mathbf{0}, X_i)^l < \infty$  for some  $l \in \mathbb{N}$  then  $E |g_j(X_1)|^l < \infty$  for  $j = 1, 2, 3$ , since  $|g_j(\mathbf{x})| = d(\mathbf{0}, \mathbf{x})$ . Moreover, since  $\mu_1 < \mu_2 + \mu_3$ , it holds that  $E g_j(X_1) < 0$  for  $j = 1, 2, 3$ . Using the binomial theorem and the function  $V : N \rightarrow \mathbb{R}$  such that  $V(\mathbf{x}) = x^{k+1}$  for  $\mathbf{x} = (i, x) \in N$ , we can see that these two assumptions yield the following drift inequality:

$$\int P(\mathbf{x}, d\mathbf{y})V(\mathbf{y}) = E |x + g_i(X_1)|^{k+1} \leq x^{k+1} - cx^k = V(\mathbf{x}) - cx^k$$

for some constant  $c > 0$  and all  $\mathbf{x} = (i, x)$  with the coordinate  $x$  large enough.

In particular, there exist constants  $0 < c_0, b_0, \delta_0 < \infty$  such that

$$\int P(\mathbf{x}, d\mathbf{y})y^{k+1} - x^{k+1} \leq -c_0x^k + b_0\mathbf{1}_{B(\mathbf{0}, \delta)}(\mathbf{x}).$$

Statements (a) and (b) now follow through respective application of Theorem 14.0.1 (i.e. the  $f$ -norm ergodic theorem) and Theorem 14.3.7 of [10].

In particular, under the conditions of Proposition 2,  $\vec{S}_n \xrightarrow{D} \pi$  as  $n \rightarrow \infty$ . Denote by  $\lambda_N$  the Lebesgue measure on  $N$ . To show  $\psi$ -irreducibility, we could assume that the distribution of the  $X_i$ s has a component with a strictly positive density with respect to the measure  $\lambda_N$  on some ball of radius  $\delta > 0$  around the origin, denoted by  $B(\mathbf{0}, \delta)$ . Then  $(\vec{S}_n)_n$  is easily seen to be aperiodic and  $\lambda_N$ -irreducible (cf. [10, Section 4.2]).

**Example 1.** (*Triple exponential distribution.*) We can write each  $N$ -valued random variable  $X$  as  $(i_X, d(\mathbf{0}, X))$ , where  $i_X$  denotes the leg of the spider containing  $X$ . Suppose that  $i_X$  and  $d(\mathbf{0}, X)$  are independent random variables. Assume further that the first of these is uniformly distributed on the set  $\{1, 2, 3\}$ , while the latter has the exponential distribution on  $[0, \infty)$  with a parameter  $\lambda > 0$ . Assume that the random variables in the i.i.d. sequence  $(X_n)_n$  have the same distribution as  $X$ . We could call this distribution triple exponential, thereby generalizing the well-known double exponential or Laplace distribution.

By Proposition 2, the Markov chain  $(\vec{S}_n)$  has an invariant distribution in the triple exponential case. From (15), it is clear that if a random variable  $\vec{S}$  on  $N$  has this invariant distribution then it necessarily holds that

$$\vec{S} \stackrel{D}{=} \vec{S} + X,$$

where  $\vec{S}$  and  $X$  on the right-hand side are assumed to be independent,  $N$ -valued random variables. It is straightforward to check that  $\vec{S} = (i_S, S)$ , with a triple exponential distribution and parameter  $\alpha = \lambda/3$ , satisfies this distributional equation. Therefore, this is the limiting distribution of the sequence  $(\vec{S}_n) = (n\vec{X}_n)$  from Theorem 1(iii).

### 3. Binary trees

In [17] the laws of large numbers are proved for the inductive mean  $\vec{X}_n$  in the global metric spaces of nonpositive curvature (see (2) and (3)). However, the conditions behind these results can be relaxed in the case of  $k$ -spiders or metric trees, as we have already seen in the case of tripods. We extend the strong law of large numbers to binary trees under the condition that the first moment is finite. In the sequel, we denote by  $T$  an arbitrary binary metric tree (unrooted and possibly unbounded). Observe that any two points on  $T$  are connected by a unique path. It is easy to see that  $T$  is also a global metric space of nonpositive curvature (cf. [17]).

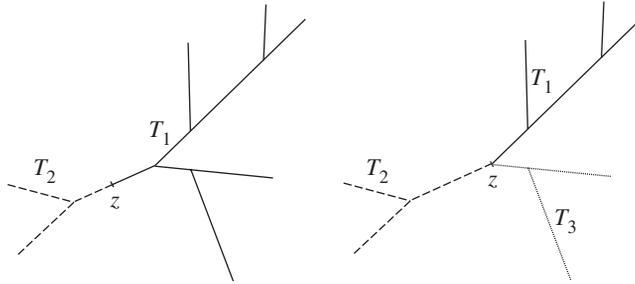


FIGURE 1: Two possible and essentially different positions of the barycenter on a binary tree.

**Theorem 2.** *Suppose that  $(X_i)$  is a sequence of i.i.d.  $T$ -valued random variables with finite expectation  $E X_1 \in T$  (see (1)). Then, as  $n \rightarrow \infty$ ,*

$$\vec{X}_n \rightarrow E X_1 \text{ a.s.}$$

*Proof.* Denote by  $T$  an arbitrary unrooted binary tree. Without loss of generality, we assume that  $z = E X_1$  is not placed at the leaf of the tree, since this happens if and only if  $X_1$  is a.s. equal to a constant  $z$ . In this case, however, the statement of the theorem clearly holds. By splicing up the tree at  $z$  we can end up with two or three binary trees rooted at  $z$  (cf. Figure 1).

(i) First, suppose that we end up with two trees, say  $T_1$  and  $T_2$ , rooted at  $z$ , as in the left-hand diagram of Figure 1. Define  $\mu_j = E d(X_1, z) \mathbf{1}_{T_j}(X_1)$ ,  $j = 1, 2$ . It necessarily holds that  $\mu_1 = \mu_2$ .

Indeed, assume that the opposite holds, for instance,  $\mu_1 - \mu_2 = \varepsilon > 0$ . Denote by  $B(z, \delta)$  the open ball of radius  $\delta > 0$  centered at  $z$ , and define  $T_{j,\delta} = T_j \cap B(z, \delta)$ ,  $T_{j,\delta}^c = T_j \setminus T_{j,\delta}$ , and  $\mu_{j,\delta} = E(d(X_1, z) \mathbf{1}_{T_{j,\delta}}(X_1))$ . By the dominated convergence theorem,  $\mu_{1,\delta} - \mu_2 \rightarrow \mu_1 - \mu_2$  as  $\delta \rightarrow 0$ . In particular, there is a  $\delta_0 > 0$  such that  $\mu_{1,\delta} - \mu_2 > \varepsilon/2$  for all  $\delta < \delta_0$ . Hence, we can find a point  $z' \in T_1$  on the same edge with  $z$  such that  $\delta = d(z, z') < \min\{\delta_0, \varepsilon\}$ . Observe that, for this  $z'$ , the following holds:

$$\begin{aligned} & E(d^2(X_1, z') - d^2(X_1, y)) \\ &= E(d^2(X_1, z') \mathbf{1}_{T_{1,\delta}^c}(X_1) + d^2(X_1, z') \mathbf{1}_{T_{1,\delta}}(X_1) + d^2(X_1, z') \mathbf{1}_{T_2}(X_1) - d^2(X_1, y)) \\ &\leq E((d(X_1, z) - \delta)^2 \mathbf{1}_{T_{1,\delta}^c}(X_1) + \delta^2 \mathbf{1}_{T_{1,\delta}}(X_1) + (d(X_1, z) + \delta)^2 \mathbf{1}_{T_2}(X_1) - d^2(X_1, y)) \\ &\leq E(d^2(X_1, z) - d^2(X_1, y)) + \delta^2 - 2\delta(\mu_{1,\delta} - \mu_2) \\ &< E(d^2(X_1, z) - d^2(X_1, y)). \end{aligned}$$

However, the last inequality contradicts the assumption that  $z$  is the barycenter; cf. (1) and the discussion thereafter. Thus,  $\mu_1 = \mu_2$ .

The following real functions on  $T$  are used repeatedly in the sequel:

$$f'_j(\mathbf{x}) = \mathbf{1}_{T_j}(\mathbf{x})d(\mathbf{x}, z), \tag{16}$$

$$g'_j(\mathbf{x}) = (-1)^{1+\mathbf{1}_{T_j}(\mathbf{x})}d(\mathbf{x}, z). \tag{17}$$

Note that, since  $\mu_1 = \mu_2$ ,  $E g'_j(X_1) = 0$  for  $j = 1$  and  $2$ . Consider the sequence  $(f'_j(\vec{X}_n))$ . It is straightforward to check that

$$f'_j(\vec{X}_n) \leq \left( \frac{n-1}{n} f'_j(\vec{X}_{n-1}) + \frac{1}{n} g'_j(X_n) \right)_+$$

by considering all possible positions of the points  $\vec{X}_{n-1}$  and  $X_n$  with respect to  $z$  (cf. the proof of Lemma 1). Multiplying the last equality by  $n$ , we obtain

$$nf'_j(\vec{X}_n) \leq ((n - 1)f'_j(\vec{X}_{n-1}) + g'_j(X_n))_+ \quad \text{for each } n \geq 0.$$

In particular, the sequence  $(nf'_j(\vec{X}_n))$  for  $j = 1, 2$  is dominated by the Lindley process  $(S_n^j)$ , given by  $S_0^j = 0$  and

$$S_n^j = (S_{n-1}^j + g'_j(X_n))_+, \quad n \geq 1.$$

Since  $E g'_j(X_n) = 0$ , it follows that  $\hat{S}_n^j/n \rightarrow 0$  a.s., again by Remark 1. Hence,  $d(\vec{X}_n, z) = f'_1(\vec{X}_n) + f'_2(\vec{X}_n) \rightarrow 0$  a.s.

(ii) In the case  $z$  is located at the vertex, it splits  $T$  into three trees, say  $T_1, T_2$ , and  $T_3$ , rooted at  $z$ , as in the right-hand diagram of Figure 1. Again, define  $\mu_j = E d(X_1, z) \mathbf{1}_{T_j}(X_1)$ ,  $j = 1, 2, 3$ , and assume without loss of generality that  $\mu_1 \geq \mu_2 \geq \mu_3$ . It necessarily holds that  $\mu_1 \leq \mu_2 + \mu_3$  or otherwise  $z$  would not be a barycenter, which can be shown as in part (i). If we consider the functions  $g'_j, f'_j$  and the sequences  $(f'_j(\vec{X}_n)), (S_n^j)$  as above for  $j = 1, 2$  and  $3$ , it holds that  $E g'_j(X_n) \leq 0$  for each  $j$ . Therefore,  $S_n^j/n \rightarrow 0$  a.s., and, finally,  $d(\vec{X}_n, z) \rightarrow 0$  a.s. as  $n \rightarrow \infty$ , exactly as in part (i). This completes the proof.

Suppose that there exists the expectation  $z = E X_1 \in T$ . Then  $z$  splits  $T$  into two or three binary trees rooted at  $z$  (see Figure 1). In either case, consider the functions  $f'_j(x)$  and  $g'_j(x)$  from (16) and (17) for  $j = 1$  or  $2$  (or  $3$  if  $z$  is located at one of the vertices), and define

$$\mu_j = E(f'_j(X_1)).$$

Assume again that

$$\mu_1 \geq \mu_2 \geq \mu_3 > 0,$$

if  $z$  is a vertex (since, if  $\mu_3 = 0$ , the random variables  $X_n$  do not take values at subtree  $T_3$  with probability 1, and  $z$  can be considered an interior point of the edge connecting the subtrees  $T_1$  and  $T_2$ ).

**Theorem 3.** *Suppose that  $(X_i)$  is a sequence of i.i.d.  $T$ -valued random variables with finite expectation  $z = E X_1$ .*

(i) *If the  $X_i$  have finite variance  $\sigma^2 = \text{var } X_1 > 0$  and  $z$  is not a vertex of  $T$ , then, as  $n \rightarrow \infty$ ,*

$$\frac{\sqrt{n}}{\sigma} g'_1(\vec{X}_n) \xrightarrow{D} Z. \tag{18}$$

(ii) *If the  $X_i$  have finite variance  $\sigma^2 = \text{var } X_1 > 0$ ,  $z$  is a vertex of  $T$ , and  $\mu_1 = \mu_2 + \mu_3$ , then, as  $n \rightarrow \infty$ ,*

$$\frac{\sqrt{n}}{\sigma} f'_1(\vec{X}_n) \xrightarrow{D} |Z|. \tag{19}$$

(iii) *If  $z$  is a vertex of  $T$  and  $\mu_1 < \mu_2 + \mu_3$ , then the sequence  $(n d(z, \vec{X}_n))_n$  is tight in distribution.*

*Proof.* (i) Let  $e$  be the edge of  $T$  containing  $z$ . For some  $\delta > 0$ , clearly  $B(z, \delta) \subseteq e$ . By Theorem 2, with probability 1, there is a number  $n_0 = n_0(\omega)$  such that, for all  $n \geq n_0$ ,  $\vec{X}_n \in B(z, \delta)$ . Since, for  $n > n_0$ ,

$$g'_1(\vec{X}_n) = (-1)^{1+1T_1(\vec{X}_n)} d(z, \vec{X}_n) = (-1)^{1+1T_1(\vec{X}_{n-1})} d\left(\mathbf{0}, \frac{n-1}{n} \vec{X}_{n-1}\right) + g'_1\left(\frac{1}{n} X_n\right),$$

for  $V_n = ng'_1(\vec{X}_n)$ , it holds that

$$V_n = V_{n-1} + g'_1(X_n).$$

In other words, for  $n > n_0$ , the sequence  $(V_n)$  behaves as a random walk on the real line with i.i.d. steps  $(g'_1(X_n))$ , since  $E g'_1(X_1) = 0$  and  $\text{var } g'_1(X_1) = \sigma^2$ , and relation (18) follows.

(ii) Again, by Theorem 2, with probability 1, there is an integer  $n_0 = n_0(\omega)$  such that, for all  $n \geq n_0$ ,  $\vec{X}_n$  stays on one of the three edges connected to the vertex  $z$ . Hence, for indices larger than  $n_0$ , the sequence  $(\vec{X}_n)$  evolves as if it was defined on a tripod. It is useful for the rest of the proof to embed this sequence in an auxiliary tripod  $N$ . This can be achieved by setting  $\vec{X}'_n \equiv (i, d(z, \vec{X}_n))$  when  $\vec{X}_n \in T_i$ . Define an  $N$ -valued process  $(\vec{S}_n)_{n \geq n_0}$  by

$$\vec{S}_n = n \vec{X}'_n, \quad n \geq n_0.$$

For the sequences for  $j = 1, 2, 3$ , set

$$f_j(\vec{S}_n) := d(\mathbf{0}, \vec{S}_n) \mathbf{1}_{N_j}(\vec{S}_n) = n d(z, \vec{X}_n) \mathbf{1}_{T_j}(\vec{X}_n), \quad n \geq n_0,$$

These three sequences behave exactly as  $(f_j(\vec{S}_n))$  in the proof of Theorem 1, because the random variables  $\vec{X}_n$  stay on the tripod with the origin  $z$  for all  $n \geq n_0$ . In particular, they are respectively dominated by the following Markov chains for  $j = 1, 2, 3$ :

$$S_{n_0}^j = f_j(\vec{S}_{n_0}), \quad S_n^j = (S_{n-1}^j + g'_j(X_n))_+, \quad n > n_0.$$

Since  $E g'_j(X_n) < 0$  for  $j = 2$  or  $3$ , the sequences  $(f'_j(\vec{S}_n))_n$  and  $(S_n^j)$  are tight in distribution for  $j \neq 1$ . Comparison of the sequence  $(f'_1(\vec{S}_n))_n$  with the Markov chain  $(S_n^1)$  implies the desired relation (19) by the fact that  $E g'_1(X_n) = 0$  (cf. arguments in the proof of Theorem 1).

(iii) Again, with probability 1, we can find an  $n_0$  such that, for all  $n \geq n_0$ , the  $\vec{X}_n$ s stay on the tripod centered at  $z$ . The reasoning above and the proof of Theorem 1 are easily adjusted to show that the sequences  $(f'_j(\vec{S}_n))_n$  are tight in distribution for all  $j = 1, 2, 3$ . This completes the proof.

Observe that Theorems 1 and 3 can be used to derive a potentially useful, simple statistical test concerning the hypotheses about the barycenter. As usual, we have to substitute the unknown variance  $\sigma^2$  with its standard sample estimate. Note that the results and proofs above extend directly to arbitrary  $k$ -spiders and locally finite nonbinary metric trees. The strong laws of large numbers hold unaltered in these cases too. Limit theorems will need minor adjustments, since on a general tree, the barycenter can split the tree into more than three subtrees. Nevertheless, asymptotically, the inductive mean will have one of the three types of behaviour described in Theorem 3. In particular, the phase transition in the limiting behaviour is still to be expected.

### Acknowledgement

This research was partially supported by the research grant MZOŠ, project number 037-0372790-2800, of the Croatian government.

### References

[1] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.  
 [2] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.

- [3] CHARIKAR, M. *et al.* (1998). Approximating a finite metric by a small number of tree metrics. In *Proc. 39th Ann. Symp. on Foundations of Computer Science*, IEEE Computer Science, Washington, DC, pp. 379–388.
- [4] DOUST, I. AND WESTON, A. (2008). Enhanced negative type for finite metric trees. *J. Funct. Anal.* **254**, 2336–2364. (Correction: **255** (2008), 532–533.)
- [5] ÉMERY, D. AND MOKOBODZKI, G. (1991). Sur le barycentre d'une probabilité dans une variété. In *Séminaire de Probabilités XXV* (Lecture Notes Math. **1485**), Springer, Berlin, pp. 220–233.
- [6] ES-SAHIB, A. AND HEINICH, H. (1999). Barycentre canonique pour un espace métrique à courbure négative. In *Séminaire de Probabilités XXXIII* (Lecture Notes Math. **1709**), Springer, Berlin, pp. 355–370.
- [7] KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small images. I. Uniqueness and fine existence. *Proc. London Math. Soc.* **61**, 371–406.
- [8] LIGGETT, T. M. AND SCHINAZI, R. B. (2009). A stochastic model for phylogenetic trees. *J. Appl. Prob.* **46**, 601–607.
- [9] LIU, L. *et al.* (2009). Coalescent methods for estimating phylogenetic trees. *Molec. Phylogenet. Evol.* **53**, 320–328.
- [10] MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- [11] PICARD, J. (1994). Barycentres et martingales sur une variété. *Ann. Inst. H. Poincaré Prob. Statist.* **30**, 647–702.
- [12] RZHETSKY, A., KUMAR, S. AND NEI, M. (1995). Four-cluster analysis: a simple method to test phylogenetic hypotheses. *Molec. Biol. Evol.* **12**, 163–167.
- [13] SEMPLE, C. AND STEEL, M. (2003). *Phylogenetics*. Oxford University Press.
- [14] SOLOMYAK, M. (2004). On the spectrum of the Laplacian on regular metric trees. *Waves Random Media* **14**, S155–S171.
- [15] STEEL, M., GOLDSTEIN, L. AND WATERMAN, M. S. (1996). A central limit theorem for the parsimony length of trees. *Adv. Appl. Prob.* **28**, 1051–1071.
- [16] STRIMMER, K. AND VON HAESLER, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Nat. Acad. Sci. USA.* **94**, 6815–6819.
- [17] STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces* (Contemp. Math. **338**), American Mathematical Society, Providence, RI, pp. 357–390.
- [18] WATERMAN, M. S. (1995). *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman and Hall, London.