RANDOM FRACTALS AND PROBABILITY METRICS

JOHN E. HUTCHINSON AND LUDGER RÜSCHENDORF

Abstract. New metrics are introduced in the space of random measures and are applied, with various modifications of the contraction method, to prove existence and uniqueness results for self-similar random fractal measures. We obtain exponential convergence, both in distribution and almost surely, of an iterative sequence of random measures (defined by means of the scaling operator) to a unique self-similar random measure. The assumptions are quite weak, and correspond to similar conditions in the deterministic case.

The fixed mass case is handled in a direct way based on regularity properties of the metrics and the properties of a natural probability space. To prove convergence in the random mass case needs additional tools such as a specially adapted choice of the space of random measures and of the space of probability distributions on measures, the introduction of reweighted sequences of random measures and a comparison technique.

1. Introduction

A theory of self-similar fractal sets and measures was developed in Hutchinson (1981). Further results and applications to image compression were obtained by Barnsley and Demko (1985) and Barnsley (1988).

Falconer (1986), Graf (1987) and Mauldin and Williams (1986) randomized each step in the approximation process to obtain self-similar random fractal sets. Arbeiter (1991) introduced and studied self-similar random fractal measures, see also Olsen (1994). For further material see Zähle (1988), Patzschke and Zähle (1990), the survey in Hutchinson (1995), and the references in all of these.

In this paper we first introduce metrics ℓ_p^* and ℓ_p^{**}, for $0 < p < \infty$, on the spaces of random measures (with random supports) and their distributions, respectively. Unlike the much simpler case for the L_p and ℓ_p metrics on real random variables (see Rachev and Rüschendorf (1995) and Rösler (1992)), the contraction properties of these new metrics arise from the linear structure of the set of measures rather than any independence properties. As a consequence it is possible to handle non-independent sums of measures, a surprising fact in the theory of probability metrics (c.f. the construction of Brownian motion in Hutchinson and Rüschendorf (1999)).

Based on contraction properties of random scaling operators with respect to ℓ_p^* and ℓ_p^{**} we establish existence, uniqueness and approximation properties of self-similar random fractal measures under very general conditions concerning the scaling system. We obtain exponential rates of convergence a.s. and in distribution, as well as convergence of moments, for the usual approximating sequences of random fractal measures. The major hypotheses are that $E\sum p_i r_i^p < 1$ for some $p > 0$ and $E\sum p_i = 1$ (where r_i are the Lipschitz constants for the functions S_i determining the random scaling operator, and p_i are the random weights, see Definition 2.1). Passing to the limit $p \to 0$ gives results under the very weak hypotheses of Corollaries 2.7 and 3.4. (In the deterministic case one obtains a very short proof of the

---

1991 *Mathematics Subject Classification.* Primary: 60G57; Secondary: 28A80, 60D05, 60G18.

Key words and phrases. random measure, fractal, self-similar, probability metric, random fractal, Monge Kantorovich metric, scaling operator, iterated function system, minimal metric.

Barnsley-Elton Theorem, originally established by Markov process arguments — see Barnsley and Elton (1985, 1988).)

The proof of existence, uniqueness and convergence in *distribution* in case the masses of the random measures are *constant* follows from the fact that the random scaling operator induces a contraction with respect to ℓ_p^{**} (see Theorem 2.6). Note, however, that the argument is not merely an application of the deterministic arguments at the individual realisation level — such arguments lead to much weaker results. Essentially one then needs to assume $\sum p_i r_i^p < 1$ a.s., rather than $E \sum p_i r_i^p < 1$.

In order to prove a.s. convergence of approximating sequences at the *random measure* level, as opposed to convergence in distribution, one needs to use the compound metric ℓ_p^* and the natural sample space $\widetilde{\Omega}$ of "construction trees", as well as a "non-constructive" extension of the random scaling operator from the distribution level to the random measure level (see (3.5), Remark 3.1 and Theorem 3.2). The tree construction $\widetilde{\Omega}$ was used in a crucial manner in the work of Kahane and Peyrière (1976) (see also the references therein) and in the papers by Falconer (1986), Graf (1987), Mauldin and Williams (1986), and Arbeiter (1991). But apart from our earlier paper Hutchinson and Rüschendorf (1998), it is nowhere else used to our knowledge in the critical manner of the present paper. More precisely, by means of the shift operators $\omega \mapsto \omega^{(i)}$, we avoid martingale arguments and are thus able to improve a.s. convergence to exponential convergence (for the random measures and masses, not just their distributions). See Theorem 3.2 and Lemma 4.4.

The *variable* mass case in Section 4 requires a number of additional new ideas.

The first is the introduction of sets $\mathbb{M}_p^*$ and $\mathcal{M}_p^*$ of random measures and their associated distributions in (4.7) and (4.3). This is needed to handle the difficulty caused by the fact that our metrics lead to infinite distance between two random measures, unless they have the same random mass a.s. (as opposed to just having the same expected mass). The second major point is the introduction of renormalised random measures $\overline{\mu}_n$. In this manner we are able to handle the interplay between weights with $E \sum p_i = 1$ (rather than $\sum p_i = 1$ a.s.) and random transformations $(S_1, \dots S_n)$.

As a consequence we are able to apply probability metric arguments to random measures with variable mass, random supports and even non-independent sums. Results similar to those for the fixed mass case are then obtained.

By way of comparison with previous work, we note that the results in Kahane and Peyrière (1976), Rachev and Rüschendorf (1995) and Rösler (1992) refer only to random masses (i.e. random real variables) and not random transformations or random measures in general. In Hutchinson and Rüschendorf (1998) we introduced "Monge-Kantorovich type" metrics on random measures and their distributions. Although those metrics are equivalent to ℓ_1^* and ℓ_1^{**}, they cannot be directly extended to the present setting. Moreover, we did not treat the variable mass case in the earlier paper.

The corresponding results of Falconer, Graf, Mauldin and Williams, and Olsen, follow from the case $r_i < 1$ a.s. and $\sum p_i = 1$ a.s., although the arguments in Olsen can be modified to give stronger results which then follow from the case $\sum p_i r_i < 1$ a.s. and $\sum p_i = 1$ a.s.. Arbeiter (using martingale arguments) assumes $E \sum p_i = 1$ and restricts considerations to contraction maps, although much of his work can be extended and then follows from the additional assumption $E \sum p_i r_i < 1$. In all cases we obtain stronger quantitative convergence results. But we also obtain our results under much weaker hypotheses with a simpler and more general approach. Finally, the techniques introduced here can be applied in other situations, in particular to the construction of various stochastic processes as noted in Hutchinson and Rüschendorf (1999).

We next discuss in a somewhat informal manner the main results and techniques in the paper. We refer to the example at the end of this section and those in Section 2 of Hutchinson and Rüschendorf (1998) as motivation for the following.

Let $(\mathfrak{X}, d)$ be a complete metric space; the reader should think of the case $\mathfrak{X} = \mathbb{R}^2$ with the Euclidean metric. A *scaling law* $(p_1, S_1, \dots, p_N, S_N)$ is a $2N$-tuple of real numbers $p_i \geq 0$ with $\sum p_i = 1$, and Lipschitz maps $S_i : \mathfrak{X} \to \mathfrak{X}$ with Lipschitz constants denoted by r_i. A *random scaling law* $\mathbf{S}$ is a random variable whose values are scaling laws, except that the condition $\sum p_i = 1$ is replaced by $E \sum p_i = 1$. The distribution induced by $\mathbf{S}$ is denoted by $\mathcal{S}$. We now fix $\mathbf{S}$ and $\mathcal{S}$.

We are interested in random measures μ with values in the set of Radon measures on $\mathfrak{X}$, and the corresponding probability distributions $\mathcal{P}$ (on the set of Radon measures on $\mathfrak{X}$) induced by such random measures.

Given a random measure μ, one defines the random measure

$$\mathbf{S}\mu = \sum_{i=1}^{N} p_i S_i \mu^{(i)},$$

where $(p_1, S_1, \dots, p_N, S_N)$ is chosen with distribution $\mathcal{S}$, and the $\mu^{(i)}$ are iid copies of μ, independent of $(p_1, S_1, \dots, p_N, S_N)$. This is only defined up to distribution; the corresponding probability distribution on measures is denoted by

$$\mathcal{S}\mathcal{P} = \operatorname{dist} \sum_{i=1}^{N} p_i S_i \mu^{(i)},$$

where $\mathcal{P} = \operatorname{dist} \mu$. Thus we can regard $\mathbf{S}$ and $\mathcal{S}$ as *scaling operators* which operate respectively on random measures or on the corresponding probability distributions on measures. Equivalently, $\mathbf{S}$ and $\mathcal{S}$ can be regarded as *random iterated function systems*, using an extension of the well-known terminology of Barnsley to the random setting.

One can iterate this construction to obtain sequences

$$\mu_0 = \mu,\ \mu_1 = \mathbf{S}\mu,\ \mu_2 = \mathbf{S}^2\mu, \dots,\ \mu_n = \mathbf{S}^n\mu, \dots\ , \tag{1.1}$$

$$\mathcal{P}_0 = \mathcal{P},\ \mathcal{P}_1 = \mathcal{S}\mathcal{P},\ \mathcal{P}_2 = \mathcal{S}^2\mathcal{P}, \dots,\ \mathcal{P}_n = \mathcal{S}^n\mathcal{P}, \dots\ . \tag{1.2}$$

Under quite general conditions one has convergence of these sequences to a random measure μ^* and probability distribution $\mathcal{P}^*$ respectively, where $\mathcal{P}^* = \operatorname{dist} \mu^*$. Moreover, $\mathcal{P}^*$ is *self-similar*, in the sense that $\mathcal{S}\mathcal{P}^* = \mathcal{P}^*$, and is the unique probability distribution on measures with this property.

In Sections 2 and 3 we restrict to the constant mass condition $\sum p_i = 1$ a.s., but consider the case $E \sum p_i r_i^p < 1$ for arbitrary $p > 0$. Note that by taking the limit as $p \to 0$ this is the random analogue of the condition of Barnsley et. al., namely $E \sum p_i \log r_i < 0$, see Remark 2.8. We first define extensions ℓ_p^* and ℓ_p^{**} of the usual minimal ℓ_p metric (on the space of unit mass measures on $\mathfrak{X}$) to the spaces of random measures and of induced probability distributions respectively, see (2.13) and (2.16). We then show that $\mathcal{S}$ and $\mathbf{S}$ are contraction maps in the appropriate spaces under the metrics ℓ_p^{**} and ℓ_p^*, thus leading to existence, uniqueness, exponential convergence and convergence of moments; see Theorems 2.6 and 3.2. As in the case of Hutchinson and Rüschendorf (1998), in order to establish the a.s. convergence (as opposed to convergence in distribution), one needs to carefully extend the scaling operator $\mathbf{S}$ to the random measure level on the space of "construction trees", see (3.5). In fact, one gets exponential a.s. convergence, see Remark 3.3.

In Section 4 we replace the condition $\sum p_i = 1$ a.s. by $E \sum p_i = 1$. This latter is necessary if the expected masses of terms in sequence (1.1) are to converge. But there are now difficulties in applying contraction methods and thus establishing exponential rates of convergence. The problem is that the extension of ℓ_p^{**} and ℓ_p^* to pairs of random measures whose masses are not a.s. equal, yields the value ∞; see the discussion at the beginning of Section 4.

However, the problem can be resolved as follows.

One first notes that *if* $\mathcal{P}^*$ is a self-similar probability distribution on measures, then the corresponding probability distribution of masses (a probability distribution on $\mathbb{R}$) is self-similar in a natural sense, see (4.1) and (4.2). The existence and uniqueness of a self-similar probability distribution P^* on $\mathbb{R}$ is established by a contraction mapping argument in Lemma 4.1. We then define a certain class $\mathcal{M}_p^*$ of probability distributions on measures for which the corresponding masses have distribution P^*, see (4.7). One shows that ℓ_p^{**} is a complete (finite) metric on this space and that $\mathcal{S}$ is a contraction map (Theorem 4.3), thus establishing existence and uniqueness of a self-similar $\mathcal{P}^*$, and in fact in a larger class obtained by dropping the mass restriction.

This still does not establish convergence of the sequence (1.2), unless $\mathcal{P}_0$ has mass distribution P^*. But this mass distribution is not known apriori in any constructive sense, and in any case we would like to have (exponential) convergence from any initial *constant* unit mass measure (which would not have mass distribution P^* unless the latter is constant) .

The next step is to switch to the space of "construction trees" and again use the extended operator $\mathbf{S}$ defined in Section 3. Analogous to before, one notes that *if* the random measure μ^* is a fixed point of $\mathbf{S}$ then the corresponding real random variable given by the mass is self-similar, see (4.6). The existence and uniqueness of such a self-similar real random variable X^* is established, again by a contraction argument, in Lemma 4.4. Working in the class of random measures with mass given by X^*, a contraction argument gives the existence and uniqueness of a fixed point μ^* for $\mathbf{S}$, and thence in a larger class obtained by dropping the mass restriction, see Theorem 4.5, Step 1.

One next modifies the sequence (1.1) by reweighting each of the N^n "components" of μ_n in such a way that the new sequence of random measures $\overline{\mu}_n$ have their masses given by X^*. This allows one to show $\overline{\mu}_n \to \mu^*$ in the ℓ_p^* metric. A separate argument shows that $\overline{\mu}_n \to \mu^*$ in the weak sense of measures, in a uniformly exponential manner against certain classes of Lipschitz functions, see Theorem 4.5, Step 2. Finally, one shows that $\overline{\mu}_n - \mu_n \to 0$ in a similar sense, analogous to an argument in Arbeiter (1991), see Theorem 4.5, Step 3.

The conclusion is convergence a.s. of the sequence (1.1) in a uniformly exponential manner, and uniform exponential convergence of the probability distributions in sequence (1.2) is also a consequence, see Remarks 4.6 and 4.7.

The following example comes from the Diploma thesis of N. Müller (1995). It indicates that by using random weights the random fractal measure is able to avoid the mass concentration phenomenon which is typical for deterministic fractal measures (see also the detailed discussion in D. Saupe (1988)). This allows more realistic models for the simulation of natural objects. It should be noted, however, that this is a rather special case of the results treated here. In particular, the supports of the measures are not random and also $\sum p_i = 1$ a.s..

In both Figures 2 and 3, $N = 4$ and the contraction maps S_i map the unit square Q to the corresponding square Q_i. Figure 2 shows the self-similar deterministic fractal measure given by certain fixed weights $p_i = m_i$ satisfying $\sum p_i = 1$ (as indicated schematically in Figure 1). Figure 3 shows one realisation of the self-similar random fractal measure in case the given p_1, p_2, p_3, p_4 are randomly permuted.

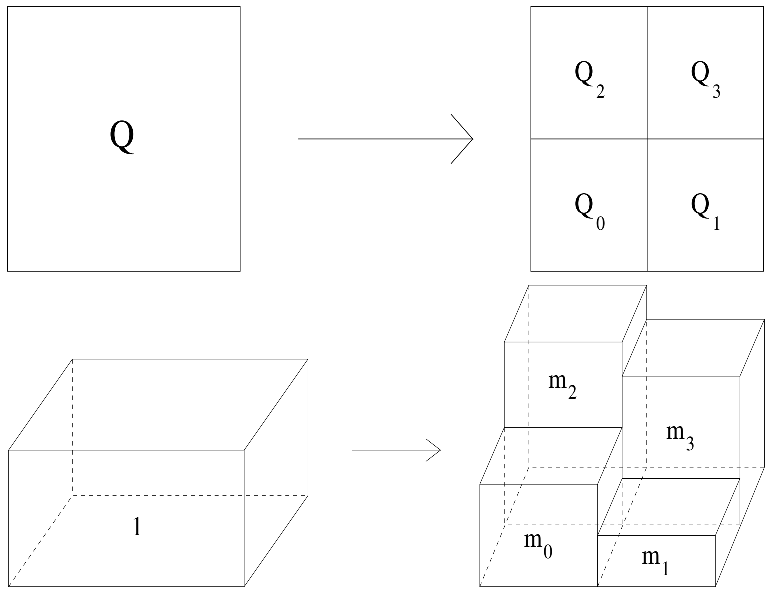

Figure 1: Partition of the support and development of mass, $m_i = p_i$

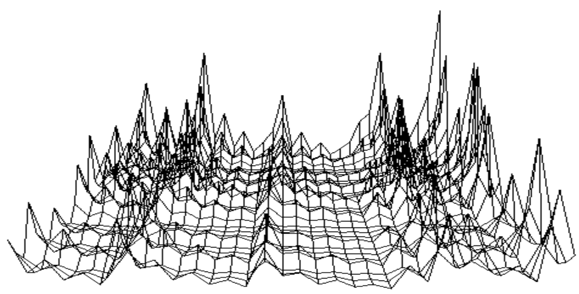

Figure 2: Deterministic fractal measure (showing the mass concentration phenomenon)

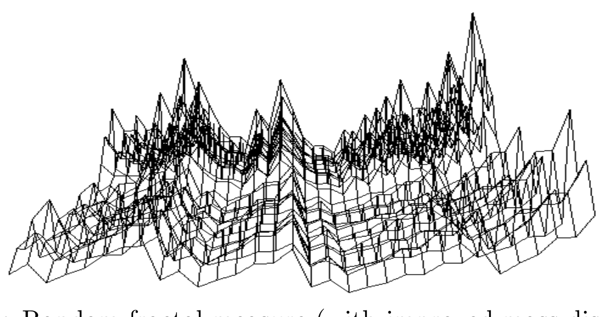

Figure 3: Random fractal measure (with improved mass distribution)

The authors thank D. Saupe and N. Müller for discussions on the subject and for providing the simulations. Part of this work was supported by grants from the Australian Research Council, and done while the first author was a visitor at the University of Freiburg supported by the Deutsche Forschungsgemeinschaft.

2. Probability metrics and self-similar random fractals

Let $(\mathfrak{X}, d)$ be a complete separable metric space and let $M = M(\mathfrak{X})$ denote the set of finite mass Radon measures on $\mathfrak{X}$ with the weak topology and corresponding Borel σ-algebra. Denote the mass of $\mu \in M$ by $|\mu| := \mu(\mathfrak{X})$.

Typically, $\mathfrak{X} = \mathbb{R}^n$ for $1 \leq n \leq 3$.

Let $(\Omega, \mathcal{A}, Q)$ be an underlying probability space, and let $\mathbb{M}$ denote the set of all random measures μ with values in M, i.e. random variables $\mu : \Omega \to M$. Let $\mathcal{M}$ denote the corresponding class of probability distributions on M, i.e.

$$\mathcal{M} = \{\, \mathcal{P} = \operatorname{dist} \mu \mid \mu \in \mathbb{M} \,\}.$$

The scaling properties of random fractal measures are described by scaling laws.

Definition 2.1. A *scaling law* $(p_1, S_1, \dots, p_N, S_N)$ is a $2N$-tuple of real numbers $p_i \geq 0$ with $\sum p_i = 1$, and Lipschitz maps $S_i : \mathfrak{X} \to \mathfrak{X}$. A *random scaling law* $\mathbf{S}$ is a random variable whose values are scaling laws, but with the condition $\sum p_i = 1$ replaced by $E \sum p_i = 1$. We will usually consider a fixed $\mathbf{S}$, and let $\mathcal{S}$ denote the corresponding distribution induced by $\mathbf{S}$.

The associated *random scaling operators* $\mathbf{S} : \mathbb{M} \to \mathbb{M}$ and $\mathcal{S} : \mathcal{M} \to \mathcal{M}$ are defined by

$$\mathbf{S}\mu = \sum_{i=1}^{N} p_i S_i \mu^{(i)}, \tag{2.1}$$

$$\mathcal{S}\mathcal{P} = \operatorname{dist} \sum_{i=1}^{N} p_i S_i \mu^{(i)}, \tag{2.2}$$

where $\mu^{(i)} \overset{d}{=} \mu \overset{d}{=} \mathcal{P}$, $\mathbf{S} = (p_1, S_1, \dots, p_N, S_N) \overset{d}{=} \mathcal{S}$ and $\mu^{(1)}, \dots, \mu^{(N)}, \mathbf{S}$ are independent.

Here $S_i \mu^{(i)}$ is the image of $\mu^{(i)}$ under S_i (or the push forward measure) and $\overset{d}{=}$ always denotes equality in distribution.

We denote the Lipschitz constant of S_i by r_i, i.e.

$$r_i = \operatorname{Lip} S_i.$$

Remark 2.2. The random variables $\mu^{(i)}(\omega)$ are only determined up to their distributions, and in particular are not determined pointwise by the random variables $\mathbf{S}(\omega)$ and $\mu(\omega)$. However, in Section 3 we see how to define a natural probability space $\widetilde{\Omega}$, in which case we do have canonical representatives for $\mu^{(i)}(\omega)$, see (3.3) In this setting we define the random measure $\mathbf{S}\mu$ pointwise so that $\mathbf{S}\mu \overset{d}{=} \mathcal{S}\mathcal{P}$, see (3.5). See also Remark 3.1.

Definition 2.3. Let $\mathbf{S}$ be a random scaling law with distribution $\mathcal{S}$, and let $\mu \in \mathbb{M}$ be a random measure with distribution $\mathcal{P} \in \mathcal{M}$. If

$$\mathcal{S}\mathcal{P} = \mathcal{P},$$

then μ is called a *random fractal measure self-similar w.r.t.* $\mathbf{S}$, and $\mathcal{P}$ is called a *random fractal measure distribution self-similar w.r.t.* $\mathcal{S}$.

The following recursive construction of a sequence of random measures will be shown to converge under quite general conditions to a self-similar measure.

Definition 2.4. Beginning with an initial measure $\mu_0 \in M$ (or more generally a random measure $\mu_0 \in \mathbb{M}$) one iteratively applies iid scaling laws with distribution $\mathcal{S}$ to obtain a sequence μ_n of random measures in $\mathbb{M}$, and a corresponding sequence $\mathcal{P}_n$ of distributions in $\mathcal{M}$, as follows.

(1) Select a scaling law $\mathbf{S} = (p_1, S_1, \ldots, p_N, S_N)$ via the distribution $\mathcal{S}$ and define

$$\mu_1 = \sum_{i=1}^{N} p_i S_i \mu_0, \text{ i.e. } \mu_1(\omega) = \sum_{i=1}^{N} p_i(\omega) S_i(\omega) \mu_0, \quad \mathcal{P}_1 \stackrel{d}{=} \mu_1,$$

(2) Select $\mathbf{S}^1, \ldots, \mathbf{S}^N$ via $\mathcal{S}$ with $\mathbf{S}^i = (p_1^i, S_1^i, \ldots, p_N^i, S_N^i)$ independent of each other and of $\mathbf{S}$ and define

$$\mu_2 = \sum_{i,j} p_i p_j^i S_i \circ S_j^i \, \mu_0, \quad \mathcal{P}_2 \stackrel{d}{=} \mu_2,$$

(3) Select $\mathbf{S}^{ij} = (p_1^{ij}, S_1^{ij}, \ldots, p_N^{ij}, S_N^{ij})$ via $\mathcal{S}$ independent of one another and of $\mathbf{S}^1, \ldots, \mathbf{S}^N, \mathbf{S}$ and define

$$\mu_3 = \sum_{i,j,k} p_i p_j^i p_k^{ij} S_i \circ S_j^i \circ S_k^{ij} \, \mu_0, \quad \mathcal{P}_3 \stackrel{d}{=} \mu_3,$$

(4) etc.

Thus $\mu_{n+1} = \sum p_i S_i \mu_n^{(i)}$ where $\mu_n^{(i)} \stackrel{d}{=} \mu_n \stackrel{d}{=} \mathcal{P}_n$, $\mathbf{S} = (p_1, S_1, \ldots, p_N, S_N) \stackrel{d}{=} \mathcal{S}$, and the $\mu_n^{(i)}$ and $\mathbf{S}$ are independent. It follows that

$$\mathcal{P}_n = \mathcal{S}\mathcal{P}_{n-1} = \mathcal{S}^n \mathcal{P}_0,$$

where $\mathcal{P}_0$ is the distribution corresponding to μ_0 (so $\mathcal{P}_0$ is constant if $\mu_0 \in M$).

In future, $p > 0$ is any positive number.

We next introduce some of the various spaces and metrics which we will use in the paper.

Let $M_p = M_p(\mathfrak{X})$ denote the set of unit mass Radon measures μ on $\mathfrak{X}$ with finite p-th moment. That is

$$M_p = \left\{ \mu \in M \;\middle|\; |\mu| = 1, \int d^p(x,a) \, d\mu(x) < \infty \right\} \tag{2.3}$$

for some (and hence any) $a \in \mathfrak{X}$. In particular, $M_p \subset M_q$ if $q \le p$. Note that μ can be considered as a probability distribution on $\mathfrak{X}$, in which case the moment condition becomes

$$Ed^p(X,a) < \infty \quad \text{if } X \stackrel{d}{=} \mu. \tag{2.4}$$

The minimal metric ℓ_p on M_p is defined by

$$\ell_p(\mu,\nu) = \inf \left\{ (Ed^p(X,Y))^{\frac{1}{p} \wedge 1} \;\middle|\; X \stackrel{d}{=} \mu, \, Y \stackrel{d}{=} \nu \right\}, \tag{2.5}$$

where $\wedge$ denotes the minimum of the relevant numbers. Equivalently,

$$\ell_p(\mu,\nu) = \inf \left\{ \left(\int d^p(x,y) \, d\gamma(x,y) \right)^{\frac{1}{p} \wedge 1} \;\middle|\; \pi_1\gamma = \mu, \; \pi_2\gamma = \nu \right\}, \tag{2.6}$$

where $\pi_i\gamma$ denotes the i-th marginal of γ, i.e. projection of the measure γ on $\mathfrak{X} \times \mathfrak{X}$ onto the i-th component.

It will be convenient in Section 4 to extend the definition of ℓ_p to arbitrary $\mu, \nu \in M$. Version (2.6) immediately carries over to this setting; version (2.5) is valid if we allow "random variables" for which the underlying probability measure need not have unit mass. Consistent with (2.6), *we define $\ell_p(\mu,\nu) = \infty$ if μ and ν have unequal masses.*

Note that if δ_a is the Dirac measure at $a \in \mathfrak{X}$ then

$$\begin{aligned} \ell_p(\mu, |\mu|\,\delta_a) &= \left(\int d^p(x,a)\,d\mu(x)\right)^{\frac{1}{p}\wedge 1}, \\ \ell_p(\delta_a, \delta_b) &= d^{1\wedge p}(a,b). \end{aligned} \tag{2.7}$$

Remark 2.5.
1. (M_p, ℓ_p) is a complete separable metric space and $\ell_p(\mu_n, \mu) \to 0$ if and only if
(1) $\mu_n \stackrel{w}{\to} \mu$ (weak convergence) and
(2) $\int d^p(x,a)\,d\mu_n(x) \to \int d^p(x,a)\,d\mu(x)$ (convergence of p-th moments).

2. The metric ℓ_1 is identical to the Monge Kantorovich metric defined by

$$d_{MK}(\mu,\nu) = \sup\left\{ \left|\int f\,d\mu - \int f\,d\nu\right| \;\middle|\; \operatorname{Lip} f \le 1 \right\}, \tag{2.8}$$

where $f : \mathfrak{X} \to \mathbb{R}$. Hutchinson (1981) used this metric for the construction of fractal measures. An extension using ℓ_p was given in Rachev and Rüschendorf (1995).

3. For measures μ, ν not necessarily of unit mass, and for $p \ge 1$, one has

$$\begin{aligned} \ell_p^p(\alpha\mu, \alpha\nu) &= \alpha \ell_p^p(\mu,\nu), \\ \ell_p^p(\mu_1+\mu_2, \nu_1+\nu_2) &\le \ell_p^p(\mu_1,\nu_1) + \ell_p^p(\mu_2,\nu_2). \end{aligned} \tag{2.9}$$

The first follows from (2.6) by setting $\gamma = c\widetilde{\gamma}$ where $\widetilde{\gamma}$ is optimal for (μ,ν). The second follows be setting $\gamma = \gamma_1 + \gamma_2$ where γ_i is optimal for (μ_i, ν_i). A similar result holds for $0 < p < 1$ if ℓ_p^p is replaced by ℓ_p, by noting $(a+b)^p \le a^p + b^p$ if $a, b \ge 0$ and $0 < p < 1$.

For Lipschitz functions $S : \mathfrak{X} \to \mathfrak{X}$, one has

$$\ell_p(S\mu, S\nu) \le (\operatorname{Lip} S)^{1\wedge p} \ell_p(\mu,\nu). \tag{2.10}$$

This follows from (2.6) by setting $\gamma = S\widetilde{\gamma}$, the pushforward of $\widetilde{\gamma}$ by S, where $\widetilde{\gamma}$ is optimal for (μ,ν).

A detailed discussion of the properties of ℓ_p can be found in the book Rachev (1991).

Let M_p be supplied with the Borel σ-algebra induced by ℓ_p. Let $\mathbb{M}_p$ denote the space of random measures $\mu : \Omega \to M_p$ with finite expected p-th moment. That is,

$$\mathbb{M}_p = \left\{ \mu \in \mathbb{M} \;\middle|\; |\mu^\omega| = 1 \text{ a.s., } E_\omega \int d^p(x,a)\,d\mu^\omega(x) < \infty \right\}. \tag{2.11}$$

It follows from (2.11) that $\mu^\omega \in M_p$ a.s. Note that $\mathbb{M}_p \subset \mathbb{M}_q$ if $q \le p$. Moreover, since $E^{1/p}|f|^p \to \exp(E \log |f|)$ as $p \to 0$,

$$\begin{aligned} \mathbb{M}_0 &:= \bigcup_{p>0} \mathbb{M}_p \\ &= \left\{ \mu \in \mathbb{M} \;\middle|\; |\mu^\omega| = 1 \text{ a.s., } E_\omega \int \log d(x,a)\,d\mu^\omega(x) < \infty \right\}. \end{aligned} \tag{2.12}$$

For random measures $\mu, \nu \in \mathbb{M}_p$, define

$$\ell_p^*(\mu,\nu) = \begin{cases} E_\omega^{\frac{1}{p}} \ell_p^p(\mu^\omega, \nu^\omega) & p \ge 1 \\ E_\omega \ell_p(\mu^\omega, \nu^\omega) & 0 < p < 1. \end{cases} \tag{2.13}$$

Compare this with (2.5) and note the formal difference in case $0 < p < 1$.

Note that ℓ_p^* is a compound metric on $\mathbb{M}_p(X)$, i.e. $\ell_p^*(\mu,\nu)$ depends on the joint distribution of μ and ν. Moreover, $(\mathbb{M}_p, \ell_p^*)$ is a complete separable metric space. Note also that $\ell_p^*(\mu,\nu) = \ell_p(\mu,\nu)$ if μ and ν are constant random measures.

Let $\mathcal{M}_p$ be the set of probability distributions of random measures $\mu \in \mathbb{M}_p$, i.e.

$$
\begin{aligned}
\mathcal{M}_p &= \{ \operatorname{dist} \mu \mid \mu \in \mathbb{M}_p \} \\
&= \Big\{ \mathcal{P} \in \mathcal{M} \Bigm| \mu \overset{d}{=} \mathcal{P} \ \Rightarrow \ |\mu^\omega| = 1 \text{ a.s.}, E_\omega \!\int d^p(x,a)\, d\mu^\omega(x) < \infty \Big\}.
\end{aligned} \tag{2.14}
$$

Note that $\mathcal{M}_p \subset \mathcal{M}_q$ for $q \le p$, and

$$
\begin{aligned}
\mathcal{M}_0 &:= \bigcup_{p>0} \mathcal{M}_p \\
&= \Big\{ \mathcal{P} \in \mathcal{M} \Bigm| \mu \overset{d}{=} \mathcal{P} \ \Rightarrow \ |\mu^\omega| = 1 \text{ a.s.}, E_\omega \!\int \log d(x,a)\, d\mu^\omega(x) < \infty \Big\}.
\end{aligned} \tag{2.15}
$$

The minimal metric on $\mathcal{M}_p$ is defined by

$$
\ell_p^{**}(\mathcal{P}, \mathcal{Q}) = \inf \{ \ell_p^*(\mu,\nu) \mid \mu \overset{d}{=} \mathcal{P}, \ \nu \overset{d}{=} \mathcal{Q} \}. \tag{2.16}
$$

Thus

$$
\ell_p^{**}(\mathcal{P}, \mathcal{Q}) = \begin{cases} \inf \Big\{ E^{\frac{1}{p}} \ell_p^p(\mu,\nu) \Bigm| \mu \overset{d}{=} \mathcal{P}, \ \nu \overset{d}{=} \mathcal{Q} \Big\} & p \ge 1 \\ \inf \Big\{ E\ell_p(\mu,\nu) \Bigm| \mu \overset{d}{=} \mathcal{P}, \ \nu \overset{d}{=} \mathcal{Q} \Big\} & 0 < p < 1. \end{cases} \tag{2.17}
$$

It follows that $(\mathcal{M}_p, \ell_p^{**})$ is a complete separable metric space with properties analogous to those in Remark 2.5, the proofs being essentially the same. In fact, similarly to Remark 2.5.1, $\ell_p^{**}(\mathcal{P}_n, \mathcal{P}) \to 0$ if and only if

(1) $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$ (weak convergence of distributions), and
(2) $E_\omega \int d^p(x,a)\, d\mu_n^\omega(x) \to E_\omega \int d^p(x,a)\, d\mu^\omega(x)$ for $\mu_n \overset{d}{=} \mathcal{P}_n$, $\mu \overset{d}{=} \mathcal{P}$ (convergence of p-th moments).

We can now prove the first existence, uniqueness and convergence result for random fractal distributions.

Theorem 2.6. *Let* $\mathbf{S}$ *be a random scaling law with corresponding scaling operator* $\mathcal{S}$ *and* $\sum p_i = 1$ *a.s. Assume* $\lambda_p := E \sum p_i r_i^p < 1$ *and* $E \sum p_i d^p(S_i a, a) < \infty$ *for some* $p > 0$ *and some (and hence any)* $a \in \mathfrak{X}$.
Then

(1) *The scaling operator* $\mathcal{S} : \mathcal{M}_p \to \mathcal{M}_p$ *is a contraction map w.r.t.* ℓ_p^{**}.
(2) *There exists a unique fractal measure distribution* $\mathcal{P}^* \in \mathcal{M}_p$ *which is self-similar w.r.t.* $\mathcal{S}$.
(3) $\mathcal{P}_n := \mathcal{S}^n \mathcal{P}_0 \to \mathcal{P}^*$ *exponentially fast w.r.t.* ℓ_p^{**} *for any* $\mathcal{P}_0 \in \mathcal{M}_p$*; more precisely*

$$
\ell_p^{**}(\mathcal{P}_n, \mathcal{P}^*) \le \frac{\lambda_p^{n(\frac{1}{p} \wedge 1)}}{1 - \lambda_p^{\frac{1}{p} \wedge 1}} \, \ell_p^{**}(\mathcal{P}_0, \mathcal{P}_1).
$$

Proof. We first claim that if $\mathcal{P} \in \mathcal{M}_p$ then $\mathcal{S}\mathcal{P} \in \mathcal{M}_p$. For this, choose iid $\mu^{(i)} \overset{d}{=} \mathcal{P}$ and $(p_1, S_1, \dots, p_N, S_N) \overset{d}{=} \mathcal{S}$ independent of the $\mu^{(i)}$. Then $\sum p_i S_i \mu^{(i)} \overset{d}{=} \mathcal{S}\mathcal{P}$. For

$p \geq 1$ we compute from (2.7), Remark 2.5.3 and independence properties,

$$\begin{aligned}
E \int & d^p(x,a)\, d\left(\sum p_i S_i \mu^{(i)}\right) \\
&= E\ell_p^p\left(\sum p_i S_i \mu^{(i)}, \delta_a\right) \\
&\leq 2^p E\ell_p^p\left(\sum p_i S_i \mu^{(i)}, \sum p_i S_i \delta_a\right) + 2^p E\ell_p^p\left(\sum p_i S_i \delta_a, \sum p_i \delta_a\right) \\
&\leq 2^p E \sum p_i r_i^p \ell_p^p(\mu^{(i)}, \delta_a) + 2^p E \sum p_i \ell_p^p\left(S_i \delta_a, \delta_a\right) \\
&= 2^p \lambda_p E\ell_p^p(\mu, \delta_a) + 2^p E \sum p_i d^p(S_i a, a) \\
&< \infty.
\end{aligned}$$

The case $0 < p < 1$ is dealt with similarly, replacing ℓ_p^p by ℓ_p and 2^p by 1.

To establish the contraction property let $\mathcal{P}, \mathcal{Q} \in \mathcal{M}_p$. Choose $\mu_i \stackrel{d}{=} \mathcal{P}$ and $\nu_i \stackrel{d}{=} \mathcal{Q}$ for $1 \leq i \leq N$ so that the pairs (μ_i, ν_i) are independent of one another and so that $\ell_p^{**}(\mathcal{P}, \mathcal{Q}) = \ell_p^*(\mu_i, \nu_i)$. Choose $(p_1, S_1, \ldots, p_N, S_N) \stackrel{d}{=} \mathcal{S}$ and independent of the (μ_i, ν_i).

For $p \geq 1$, one has from Remark 2.5.3 and independence properties that

$$\begin{aligned}
\ell_p^{**p}(\mathcal{S}\mathcal{P}, \mathcal{S}\mathcal{Q}) &\leq E\ell_p^p\left(\sum p_i S_i \mu_i, \sum p_i S_i \nu_i\right) \\
&\leq E \sum p_i r_i^p\, \ell_p^p(\mu_i, \nu_i) \\
&= \sum E(p_i r_i^p)\, E\ell_p^p(\mu_i, \nu_i) \\
&= \sum E(p_i r_i^p)\, \ell_p^{**p}(\mathcal{P}, \mathcal{Q}) \\
&= \lambda_p \ell_p^{**p}(\mathcal{P}, \mathcal{Q}).
\end{aligned}$$

In case $0 < p < 1$, one replaces ℓ_p^{**p} and ℓ_p^p throughout by ℓ_p^{**} and ℓ_p respectively.

Hence $\mathcal{S}$ is a contraction map on $\mathcal{M}_p$ with contraction constant $\lambda_p^{\frac{1}{p} \wedge 1}$.

Parts 2. and 3. are consequences of 1. □

The next result establishes existence, uniqueness and distributional convergence for random fractal measure distributions in the class $\mathcal{M}_0$.

Corollary 2.7. *Let* $\mathbf{S}$ *be a random scaling law with corresponding scaling operator* $\mathcal{S}$ *and with* $\sum p_i = 1$ *a.s. Assume* $E \sum p_i \log r_i < 0$ *and* $E \sum p_i \log d(S_i a, a) < \infty$.

Then for some $p > 0$ *the hypotheses, and hence the conclusions, of Theorem 2.6 are true. In particular, there is a unique fractal measure distribution* $\mathcal{P}^* \in \mathcal{M}_0$ *which is self-similar w.r.t.* $\mathcal{S}$*. Moreover,* $\mathcal{P}_n = \mathcal{S}^n \mathcal{P}_0 \to \mathcal{P}^*$ *in the distributional sense for any* $\mathcal{P}_0 \in \mathcal{M}_0$.

Proof. The contraction coefficient in Theorem 2.6 is $a_p = \lambda_p^{\frac{1}{p} \wedge 1}$, where $\lambda_p = E \sum p_i r_i^p$. To investigate the behaviour of a_p, note that

$$\lambda_p = \int r_i^p(\omega)\, dQ^*(\omega, i),$$

where Q^* is the probability measure on $\Omega \times \{1, \ldots, N\}$ uniquely defined by

$$\int f(\omega, i)\, dQ^* = \int \sum p_i(\omega) f(\omega, i)\, dQ(\omega) = E_\omega \sum p_i f(\omega, i).$$

By standard properties of L^p metrics it follows that $\lambda_p^{\frac{1}{p}} = \|r(i,\omega)\|_{L^p(Q^*)}$ is monotonically nondecreasing and continuous on $(0,\infty)$ and

$$\lambda_p^{\frac{1}{p}} \to \operatorname*{ess\,sup}_{\omega} \max_i \{\, r_i(\omega) \mid p_i(\omega) > 0 \,\} \qquad \text{as } p \to \infty,$$

$$\lambda_p^{\frac{1}{p}} \to \exp E \sum p_i \log r_i \qquad \text{as } p \to 0.$$

In particular, $a_p < 1$ implies $a_q < 1$ for all $0 < q < p$ (although a_p is not itself monotone). Hence $\mathcal{S}$ is a contraction on $(\mathcal{M}_p, \ell_p^{**})$ implies $\mathcal{S}$ is a contraction on $(\mathcal{M}_q, \ell_q^{**})$ for all $0 < q < p$. Moreover, $\exp E \sum p_i \log r_i < 1$ iff $\mathcal{S}$ is a contraction on $\mathcal{M}_p$ for some $p > 0$.

Similarly, $E \sum p_i \log d(S_i a, a) < \infty$ iff $E \sum p_i d^p(S_i a, a) < \infty$ for some $p > 0$.

The result now follows from Theorem 2.6. □

Remark 2.8. The main hypothesis of the Corollary reduces to $\prod r_i^{p_i} < 1$ in the deterministic case. This is the condition used in Barnsley, Demko, Elton and Geronimo (1988, 1989) and in Barnsley and Elton (1985), provided p_i is independent of $x \in \mathfrak{X}$. Note that by Jensen's inequality $E \log \prod r_i^{p_i} \leq \log E \prod r_i^{p_i}$, and so the assumption $E \prod r_i^{p_i} < 1$ is stronger than that used in the Corollary.

Remark 2.9. In the opposite direction to the Corollary, if one assumes $r_{\max} := \max_i \{\, r_i \mid p_i \neq 0 \,\} \leq \gamma < 1$ a.s., then it follows from Theorem 2.6 that one obtains convergence of moments of all orders. This is even true if $\gamma = 1$ a.s., provided that $\min_i \{\, r_i \mid p_i \neq 0 \,\} < 1$ with non-zero probability.

Remark 2.10. In the deterministic case, the arguments in Theorem 2.6 and Corollary 2.7 simplify greatly. In particular, this simplifies the arguments in Barnsley, Demko, Elton and Geronimo (1988, 1989) and in Elton (1987) and provides further information concerning the type of convergence.

More precisely, let $\mathbf{S} = (p_1, S_1, \ldots, p_n, S_N)$ be a fixed scaling law such that $\sum p_i = 1$, $\sum p_i r_i^p < 1$ and $\sum p_i d^p(S_i a, a) < \infty$. One argues as in Theorem 2.6 except that one simply uses the ℓ_p metric throughout, so it is not necessary to take expectations and there is no need to consider optimal pairings (μ_i, ν_i).

It follows that: *If* $\mathbf{S}\mu := \sum p_i S_i \mu$ *then there is a unique Radon measure* $\mu^* \in M_p$ *such that* $\mathbf{S}\mu^* = \mu^*$. *We say* μ^* *is* self-similar with respect to $\mathbf{S}$. *Moreover, for any* $\mu_0 \in M_p$ *the sequence* $\mathbf{S}^n \mu_0$ *converges exponentially in the* ℓ_p *metric, and in particular in the weak sense of measures, to* μ^*. *Existence, uniqueness and weak convergence follow under the weaker condition* $\prod r_i^{p_i} < 1$ *as in the proof of Corollary 2.7.*

Remark 2.11. Let $\mathbf{S} = (p_1, S_1, \ldots, p_n, S_N)$ be a fixed scaling law as in the previous Remark. Suppose $\mu \in M_p$ and consider μ as a probability distribution on $\mathfrak{X}$. Let X be a random variable with $\operatorname{dist} X = \mu$ and let I be an independent random variable with $P(I = i) = p_i$. It follows that

$$\mathbf{S}\mu = \operatorname{dist}(S_I X).$$

This is the point of view taken in Rachev and Rüschendorf (1995), Section 2.4, and leads to an instructive alternative approach to some of the results here in Section 2.

For example, to establish the contraction properties of ℓ_p^{**} in Theorem 2.6 one can use a probabilistic argument to show for fixed weights $\sum p_i = 1$, for $\mu_i, \nu_i \in \mathbb{M}_p(\mathfrak{X})$ and for $p \geq 1$, that

$$\ell_p^p \left(\sum p_i S_i \mu_i, \sum p_i S_i \nu_i \right) \leq \sum p_i r_i^p \ell_p^p(\mu_i, \nu_i),$$

by arguing as follows:

Let (X_i, Y_i) be independent optimal couplings for (μ_i, ν_i) and let I be a random variable independent of the (X_i, Y_i) with $P(I = i) = p_i$. Conditioning on I we have

$$\operatorname{dist} S_I X_I = \sum p_i \operatorname{dist} S_i X_i = \sum p_i S_i \mu_i,$$

and so

$$\begin{aligned}\ell_p^p \Big(\sum p_i S_i \mu_i, \sum p_i S_i \nu_i \Big) &\le E d^p (S_I X_I, S_I Y_I) \\ &= \sum p_i E|_I d^p (S_i X_i, S_i Y_i) = \sum p_i E|_I r_i^p d^p (X_i, Y_i) = \sum p_i r_i^p \ell_p^p (\mu_i, \nu_i).\end{aligned}$$

In case $0 < p < 1$ a similar argument applies.

Remark 2.12. Neither the definition of the scaling operator $\mathcal{S}$ on $\mathcal{M}_p$, nor the proof of Theorem 2.6, are the obvious analogue of the deterministic case.

More precisely, motivated by the previous Remark and since $\mathcal{P} \in \mathcal{M}_p$ is the probability distribution of a random measure, one might try to define

$$\mathcal{S}\mathcal{P} = \operatorname{dist}(S_I \mu),$$

where μ is a random measure with $\operatorname{dist} \mu = \mathcal{P}$, $\mathbf{S} = (p_1, S_1, \ldots, p_N, S_M)$ is a random scaling law with $\operatorname{dist} \mathbf{S} = \mathcal{S}$, I is a real random variable with $P(I = i) = p_i$, and μ, $\mathbf{S}$ and I are independent.

But a scaling operator defined in this manner does not lead to the correct notion. In particular, one does not here select N independent realisations of $\mathcal{P}$ in defining $\mathcal{S}\mathcal{P}$. Moreover, $E \mapsto \bigcup S_i E$ is a contraction map on compact sets in case all $r_i < 1$. Thus realisations of $\mathcal{S}^n \mathcal{P}$ would here converge to Dirac measures (not a very interesting situation).

Another analogue of the deterministic case $\mathbf{S}\mu = \sum p_i S_i \mu$ is to define

$$\mathcal{S}\mathcal{P} = \operatorname{dist} \sum p_i S_i \mu,$$

where $\mu \overset{d}{=} \mathcal{P}$ is independent of $(p_1, S_1, \ldots, p_N, S_N) \overset{d}{=} \mathcal{S}$. This corresponds to a variation of the previous iterative procedure where in Step 2 one selects $\mathbf{S}^1, \ldots, \mathbf{S}^N$ with distribution $\mathcal{S}$ but all equal to one another a.s., in step 3 one selects $\mathbf{S}^{ij}$ equal to one another, etc. Again, this is not the correct notion.

3. Construction trees and almost sure convergence

The minimal L_p-metric ℓ_p^{**} on the set of random measures introduced in Section 2 describes weak convergence of the iterative sequence of distributions $\mathcal{P}_n$ of random measures to a random fractal distribution.

In this section we consider a natural probability space $\widetilde{\Omega}$, the space of construction trees, on which the corresponding sequence μ_n converges almost surely. The argument for a.s. convergence is based on a contraction argument for the compound version ℓ_p^* of the L_p-metric as defined in (2.13).

In order to introduce the space of construction trees let $C = C_N$ denote the N-fold tree of all finite sequences from $\{1, \ldots, N\}$, including the empty sequence $\emptyset$. For $\sigma = \sigma_1 \ldots \sigma_n \in C$ define the length $|\sigma| = n$, and for $\tau = \tau_1 \ldots \tau_m \in C$ denote the concatenated sequence $\sigma_1 \ldots \sigma_n \tau_1 \cdots \tau_m$ by $\sigma \star \tau$.

A *construction tree* (or tree of scaling laws) is a map $\omega : C \to \Upsilon$, where Υ is the set of scaling laws of $2N$-tuples. Let

$$\widetilde{\Omega} = \{\, \omega \mid \omega : C \to \Upsilon \,\}$$

denote the space of all construction trees. Denote the scaling law at the node $\sigma \in C$ of ω by

$$\mathbf{S}^\sigma(\omega) = \omega^\sigma = \big(p_1^\sigma(\omega), S_1^\sigma(\omega), \ldots, p_N^\sigma(\omega), S_N^\sigma(\omega) \big) = \omega(\sigma).$$

Consider the probability measure on $\widetilde{\Omega}$ obtained by selecting iid scaling laws $\mathbf{S}^\sigma \stackrel{d}{=} \mathcal{S}$ for each $\sigma \in C$. Note that the distributions and independencies of the $\mathbf{S}^\sigma$ are the same as in the iterative procedure described in Definition 2.4.

We use the notation

$$\begin{aligned}\overline{p}^\sigma &= p_{\sigma_1} p^{\sigma_1}_{\sigma_2} p^{\sigma_1\sigma_2}_{\sigma_3} \cdots p^{\sigma_1\cdots\sigma_{n-1}}_{\sigma_n}, \\ \overline{S}^\sigma &= S_{\sigma_1} \circ S^{\sigma_1}_{\sigma_2} \circ \cdots \circ S^{\sigma_1\cdots\sigma_{n-1}}_{\sigma_n},\end{aligned} \tag{3.1}$$

where $|\sigma| = n$. In particular $\overline{p}^i = p_i$ and $\overline{S}^i = S_i$ for $1 \le i \le N$. The motivation for this notation is the following definition.

For a fixed measure $\mu_0 \in M$ define

$$\mu_n = \mu_n(\omega) = \sum_{|\sigma|=n} \overline{p}^\sigma(\omega) \overline{S}^\sigma(\omega) \mu_0 \tag{3.2}$$

for $n \ge 1$. This is just the sequence defined in Definition 2.4 with underlying space $\Omega = \widetilde{\Omega}$. Note that μ_n is the sum of the N^n measures naturally associated with μ_0 and the N^n nodes at level n of the construction tree ω.

For $\omega \in \widetilde{\Omega}$ and $1 \le i \le N$ let $\omega^{(i)} \in \widetilde{\Omega}$, corresponding to the i-th branch of ω, be defined by

$$\omega^{(i)}(\sigma) = \omega(i \star \sigma) \tag{3.3}$$

for $\sigma \in C$. Then

$$\begin{aligned}\overline{p}^{i\star\sigma}(\omega) &= p_i(\omega)\, \overline{p}^\sigma(\omega^{(i)}), \\ \overline{S}^{i\star\sigma}(\omega) &= S_i(\omega) \circ \overline{S}^\sigma(\omega^{(i)}).\end{aligned} \tag{3.4}$$

By construction, the "branches" $\omega^{(1)}, \dots, \omega^{(N)}$ of ω are iid with the same distribution as ω and are independent of $\big(p_1(\omega), S_1(\omega), \dots, p_N(\omega), S_N(\omega)\big)$. For more details see Hutchinson and Rüschendorf (1998).

Corresponding to the random scaling law $\mathbf{S}$ we define the *scaling operator* $\mathbf{S} : \mathbb{M} \to \mathbb{M}$, where $\mathbb{M} = \mathbb{M}(\widetilde{\Omega})$ is the class of random measures on $\widetilde{\Omega}$, by

$$\mathbf{S}\mu(\omega) = \sum_i p_i(\omega) S_i(\omega) \mu(\omega^{(i)}). \tag{3.5}$$

Since $\mu(\omega^{(i)}) \stackrel{d}{=} \mu$ are iid, $\mathbf{S}\mu$ is identical in distribution to the scaling operator $\mathcal{S}$ applied to $\operatorname{dist}\mu$, see (2.2). Moreover, if $\mu_n(\omega)$ is as in (3.2) then

$$\mu_{n+1}(\omega) = \mathbf{S}\mu_n(\omega). \tag{3.6}$$

To see this, note from (3.2) and (3.4) that

$$\begin{aligned}\mu_{n+1}(\omega) &= \sum_{i=1}^N \sum_{|\sigma|=n} \overline{p}^{i\star\sigma} \overline{S}^{i\star\sigma}(\omega)\mu_0 \\ &= \sum_{i=1}^N p_i(\omega) S_i(\omega) \sum_{|\sigma|=n} \overline{p}^\sigma(\omega^{(i)})\, \overline{S}^\sigma(\omega^{(i)})\mu_0 \\ &= \sum_{i=1}^N p_i(\omega) S_i(\omega) \mu_n(\omega^{(i)}) = \mathbf{S}\mu_n(\omega).\end{aligned}$$

Thus if we take $\Omega = \widetilde{\Omega}$ then the sequence $\mu_n(\omega) = \mathbf{S}^n \mu_0$ is the same as that given in Definition 2.4.

Remark 3.1. It follows from (3.5) that $\mu_{n+1}(\omega) = \mathbf{S}\mu_n(\omega)$ is completely determined by a knowledge of the construction tree ω up to level n. This is clear in any case from Definition 2.4, and so in a certain sense we might think of the scaling operator $\mathbf{S}$ as being constructive. However, in order to apply a contraction mapping argument we need to extend the definition of $\mathbf{S}$ to *all* of $\mathbb{M}$ (or at least $\mathbb{M}_p$), and this extension is not in general constructive, see (3.5). More precisely, $\mathbf{S}$ is of necessity a type of "shift operator" and $\mathbf{S}\mu(\omega)$ generally depends on knowledge of the complete tree ω.

We now prove almost sure convergence of μ_n, with respect to weak convergence of measures, to a fixed point μ^* of $\mathbf{S}$. An immediate consequence is that $\mu^* \overset{d}{=} \mathcal{P}^*$, i.e. by Theorem 2.6 the distribution of μ^* is the unique distribution in $\mathcal{M}_p$ which is self-similar with respect to $\mathcal{S}$.

Theorem 3.2. *Let* $\mathbf{S} = (p_1, S_1, \dots, p_N, S_N)$ *be a random scaling law with* $\sum p_i = 1$ *a.s. Assume* $\lambda_p := E\left(\sum p_i r_i^p\right) < 1$ *and* $E\sum p_i d^p(S_i a, a) < \infty$ *for some* $p > 0$. *Then*

1. *The operator* $\mathbf{S} : \mathbb{M}_p \to \mathbb{M}_p$ *is a contraction map w.r.t.* ℓ_p^*.
2. *If* μ^* *is the unique fixed point of* $\mathbf{S}$ *and* $\mu_0 \in M_p$ *(or more generally* $\mathbb{M}_p$*), then* $\mu_n = \mathbf{S}^n \mu_0 \to \mu^*$ *exponentially fast w.r.t.* ℓ_p^**, and hence a.s. in the sense of weak convergence of measures.*

Moreover, $\operatorname{dist}\mu^* = \mathcal{P}^*$.

Proof. The fact $\mathbf{S} : \mathbb{M}_p \to \mathbb{M}_p$ can be seen as in Theorem 2.6.

Further, for $\mu, \nu \in \mathbb{M}_p$ and $p \geq 1$, we have from Remark 2.5.3 and the independence of $(p_1, S_1, \dots, p_N, S_N)$ and the $\omega^{(i)}$, that

$$\begin{aligned}
\ell_p^{*p}(\mathbf{S}\mu, \mathbf{S}\nu) &= E_\omega \ell_p^p \left(\sum p_i^\omega S_i^\omega \mu(\omega^{(i)}), \sum p_i^\omega S_i^\omega \nu(\omega^{(i)}) \right) \\
&\leq E_\omega \sum p_i^\omega (r_i^\omega)^p \, \ell_p^p \left(\mu(\omega^{(i)}), \nu(\omega^{(i)}) \right) \\
&= \sum E_\omega p_i^\omega (r_i^\omega)^p \, E_\omega \ell_p^p \left(\mu(\omega^{(i)}), \nu(\omega^{(i)}) \right) \\
&= \lambda_p \ell_p^{*p}(\mu, \nu),
\end{aligned}$$

as $\left(\mu(\omega^{(i)}), \nu(\omega^{(i)}\right) \overset{d}{=} (\mu(\omega), \nu(\omega))$. In case $0 < p < 1$, one replaces ℓ_p^{*p} and ℓ_p^p throughout by ℓ_p^* and ℓ_p respectively.

Thus $\mathbf{S}$ is a contraction map with contraction ratio $\lambda_p^{\frac{1}{p}\wedge 1}$, establishing 1., and hence giving exponential convergence as in 2. to the unique fixed point μ^* of $\mathbf{S}$.

Moreover, for $p \geq 1$

$$\sum_{n=1}^{\infty} Q\left(\ell_p^p(\mathbf{S}^n \mu_0, \mu^*) \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{E\ell_p^p(\mathbf{S}^n \mu_0, \mu^*)}{\varepsilon} \leq c \sum_{n=1}^{\infty} \frac{\lambda_p^n}{\varepsilon} < \infty.$$

This implies that $\ell_p(\mu_n, \mu^*) \to 0$ almost surely, and similarly for $0 < p < 1$ with ℓ_p^p replaced by ℓ_p.

Finally, since $\mathbf{S}\mu^* = \mu^*$, taking distributions of both sides and using the uniqueness of $\mathcal{P}^*$ from Theorem 2.6, it follows that $\operatorname{dist}\mu^* = \mathcal{P}^*$. □

Remark 3.3. In fact from the above argument

$$\frac{\ell_p(\mu_n, \mu^*)}{\tau^n} \to 0 \text{ a.s., for all } 0 < \tau < \lambda_p^{\frac{1}{p}\wedge 1}.$$

That is, we have an exponential a.s. convergence rate.

Analogously to Corollary 2.7 we have:

Corollary 3.4. *Let* $\mathbf{S}$ *be a random scaling law with* $\sum p_i = 1$ *a.s. Assume that* $E\sum p_i \log r_i < 0$ *and* $E\sum p_i \log d(S_i a, a) < \infty$.

Then for some $p > 0$ *the hypotheses, and hence the conclusions, of Theorem 3.2 are true. In particular,* $\mathbf{S} : \mathbb{M}_0 \to \mathbb{M}_0$ *has a unique fixed point* μ^**, and* $\mu_n = \mathbf{S}^n \mu_0 \to \mu^*$ *a.s. in the sense of weak convergence of measures, for any* $\mu_0 \in \mathbb{M}_0$. *Moreover,* $\operatorname{dist}\mu^* = \mathcal{P}^*$.

Proof. The Corollary follows from Theorem 2.6 by similar arguments to those used to prove Corollary 2.7. □

4. Self-similar fractals in the general mass case

The aim of this section is to extend the contraction technique in case the condition $\sum p_i = 1$ a.s. is replaced by the assumption $E\sum p_i = 1$. This allows for fractal measures and distributions in case the masses are not a.s. constant. The condition $E\sum p_i = 1$ is necessary if the expected mass of measures in the iterative procedure is to converge.

The methods from Sections 2 and 3, based on contraction properties of the ℓ_p, ℓ_p^* and ℓ_p^{**} metrics, appear at first to only work in the fixed mass case. More precisely, if $\mu, \nu \in M(\mathfrak{X})$ and $|\mu| \neq |\nu|$ then $\ell_p(\mu,\nu) = \infty$. This implies that if the definition of ℓ_p^* in (2.13) is extended from $\mathbb{M}_p$ to $\mathbb{M}$ (defined at the beginning of Section 2) then $\ell_p^*(\mu,\nu) = \infty$ unless $|\mu^\omega| = |\mu^\omega|$ a.s. Hence, if the definition of ℓ_p^{**} in (2.16) is extended from $\mathcal{M}_p$ to $\mathcal{M}$ then $\ell_p^{**}(\mathcal{P},\mathcal{Q}) = \infty$ unless there exist $\mu,\nu \in \mathbb{M}$ with $\mu \stackrel{d}{=} \mathcal{P}$, $\nu \stackrel{d}{=} \mathcal{Q}$ *and* $|\mu^\omega| = |\mu^\omega|$ a.s. This would seem to restrict us to the constant mass case.

However, we can avoid this problem in the following manner. First define for any $\mu \in \mathbb{M}$ the corresponding real random variable $|\mu|$ determined by the masses $|\mu^\omega|$. Similarly, for any $\mathcal{P} \in \mathcal{M}$ define

$$|\mathcal{P}| = \left\{ \operatorname{dist}|\mu| \;\middle|\; \mu \stackrel{d}{=} \mathcal{P} \right\}.$$

If $\mathcal{P}^*$ is self-similar w.r.t. $\mathcal{S}$, then on taking masses of each side of $\mathcal{P}^* = \mathcal{S}\mathcal{P}^*$ we obtain

$$|\mathcal{P}^*| = |\mathcal{S}\mathcal{P}^*| = |\mathcal{S}|\,|\mathcal{P}^*|. \tag{4.1}$$

Here $|\mathcal{S}|$ is the distribution on $(p_1,\dots,p_N)$ induced from the distribution $\mathcal{S}$ on $(p_1,S_1,\dots,p_N,S_N)$. Also, for any probability distribution P on $\mathbb{R}$, $|\mathcal{S}|P$ is the distribution defined by

$$|\mathcal{S}|P = \operatorname{dist}\sum p_i X_i, \tag{4.2}$$

where $X_i \stackrel{d}{=} P$ are iid with distribution P and $(p_1,\dots,p_N) \stackrel{d}{=} |\mathcal{S}|$ is independent of the X_i.

In Lemma 4.1 we show, under the natural assumption $E\sum p_i^2 < 1$ (c.f. Remark 4.2.1), that there is a *unique* probability distribution P^* on $[0,\infty)$, with expectation normalised to be one and finite variance, such that

$$P^* = |\mathcal{S}|P^*.$$

Thus *if* there exists $\mathcal{P}^*$ which is self-similar with respect to $\mathcal{S}$ and which has expected mass one and finite mass variance, then it follows from (4.1) that $|\mathcal{P}^*| = P^*$.

For this reason, assuming Lemma 4.1, we define

(4.3)
$$\overline{\mathcal{M}}_p = \left\{ \mathcal{P} \in \mathcal{M} \;\middle|\; \mu \stackrel{d}{=} \mathcal{P} \;\Rightarrow\; E|\mu| = 1,\ E|\mu|^2 < \infty,\ E_\omega \int d^p(x,a)\,d\mu^\omega(x) < \infty \right\}$$
$$\mathcal{M}_p^* = \left\{ \mathcal{P} \in \mathcal{M} \;\middle|\; \mu \stackrel{d}{=} \mathcal{P} \;\Rightarrow\; |\mu| \stackrel{d}{=} P^*,\ E_\omega \int d^p(x,a)\,d\mu^\omega(x) < \infty \right\},$$

where P^* is as in Lemma 4.1, c.f. (2.14). Then $\mathcal{M}_p^* \subset \overline{\mathcal{M}}_p$ by the previous discussion. If $\mathcal{P} \in \overline{\mathcal{M}}_p$ is self-similar with respect to $\mathcal{S}$, then $\mathcal{P} \in \mathcal{M}_p^*$. (We will later define an analogous spaces $\overline{\mathbb{M}}_p$ and $\mathbb{M}_p^*$ at the random measure level, see (4.7); this is most naturally done in the context of the special probability space $\widetilde{\Omega}$.)

Define ℓ_p^{**} as in (2.16). Although ℓ_p^{**} takes infinite values on $\overline{\mathcal{M}}_p$ (as noted previously), $(\mathcal{M}_p^*, \ell_p^{**})$ *is a complete metric space.*

The main new point is to show $\ell_p^{**}(\mathcal{P}, \mathcal{Q}) < \infty$ for $\mathcal{P}, \mathcal{Q} \in \mathcal{M}_p^*$. To see this, choose $\mu \overset{d}{=} \mathcal{P}$ and $\nu \overset{d}{=} \mathcal{Q}$. Then $|\mu| \overset{d}{=} |\nu|$. A well known result from measure theory says that for real random variables X, Y with $P^X = P^Y$ there exists a measure preserving map ϕ on Ω (i.e. $P^\phi = P$) such that $X = Y \circ \phi$ a.s. So $|\mu| = |\nu| \circ \phi$ a.s.. Define the random measure $\overline{\nu} = \nu \circ \phi$; then $|\overline{\nu}| = |\mu|$ a.s. Therefore for $p \geq 1$ one has from (2.7) that

$$\begin{aligned}
\ell_p^{**p}(\mathcal{P}, \mathcal{Q}) &\leq E\ell_p^p(\mu, \nu) \leq 2^p E\ell_p^p(\mu, |\mu|\delta_a) + 2^p E\ell_p^p(\nu, |\nu|\delta_a) \\
&= 2^p E \int d^p(x, a)\, d\mu(x) + 2^p E \int d^p(x, a)\, d\nu(x) < \infty.
\end{aligned}$$

A similar argument applies if $0 < p < 1$.

For the following Lemma define the set $\mathcal{D}_2$ of probability measures P on $\mathbb{R}$ by

$$\mathcal{D}_2 = \left\{ P \mid X \overset{d}{=} P \Rightarrow EX = 1,\ EX^2 < \infty \right\}. \tag{4.4}$$

The requirement $EX = 1$ is a normalisation condition and involves no loss of generality.

Lemma 4.1. *If $E\sum p_i = 1$ and $E\sum p_i^2 < 1$ then there is a unique probability distribution $P^* \in \mathcal{D}_2$ such that*

$$P^* = |\mathcal{S}|P^*. \tag{4.5}$$

Proof. We claim $|\mathcal{S}| : \mathcal{D}_2 \to \mathcal{D}_2$ is a contraction map in the ℓ_2 metric.

For this purpose suppose $P_1, P_2 \in \mathcal{D}_2$ and let X, Y be optimal ℓ_2-couplings, i.e. $\ell_2^2(P_1, P_2) = E(X - Y)^2$. Choose iid copies $(X^{(i)}, Y^{(i)})$ of (X, Y) and choose $(p_1, \ldots, p_N) \overset{d}{=} |\mathcal{S}|$ independent of the $(X^{(i)}, Y^{(i)})$. Then

$$\begin{aligned}
\ell_2^2\big(|S|P_1, |S|P_2\big) &\leq E\bigg(\sum_i p_i \Big(X^{(i)} - Y^{(i)}\Big)\bigg)^2 \\
&= \sum_i Ep_i^2\, E\left(X^{(i)} - Y^{(i)}\right)^2
\end{aligned}$$

(by the independence properties and since $E(X^{(i)} - Y^{(i)}) = 0$)

$$= \left(E\sum p_i^2\right) \ell_2^2(P_1, P_2).$$

The contraction mapping principle now implies the Lemma. □

Remark 4.2.
1. The condition $E\sum p_i^2 < 1$ is also necessary for the existence of a fixed point with finite second moments. (The only exception is the trivial case where, almost surely, all but one p_i equals zero, in which case the exceptional p_i must equal 1, and every $P \in \mathcal{D}_2$ is then clearly a fixed point.)

To see this, suppose P^* is a fixed point and let $Z \overset{d}{=} P^*$. Then from (4.5) we have

$$EZ^2 = \sum_i Ep_i^2 EZ^2 + 2\sum_{i \neq j} Ep_i p_j.$$

If $EZ^2 < \infty$ this implies $E\sum p_i^2 < 1$, apart from the exceptional case mentioned above.

2. For $1 < p \le 2$ one can replace the condition $E\sum p_i^2 < 1$ by

$$E\sum p_i^p < \frac{(p-1)^{1/2}}{18\,p^{3/2}}.$$

This comes from working with the ℓ_p-metric (instead of ℓ_2) and using the Marcinciewicz-Zygmund inequality (c.f. Rachev and Rüschendorf (1995)). A proof of the existence of a solution of (4.5) without an additional moment assumption (of the form $E\sum p_i^2 < 1$) can be obtained from a martingale argument.

We are now in a position to give a proof in the variable mass case of the existence and uniqueness of a self-similar probability distribution on fractal nmeasures.

Theorem 4.3. *Let* $\mathbf{S}$ *be a random scaling law with corresponding scaling operator* $\mathcal{S}$*, such that* $E\sum p_i = 1$ *and* $E\sum p_i^2 < 1$*. Assume also that* $E\sum p_i r_i^p < 1$ *and* $E\sum p_i d^p(S_i a, a) < \infty$ *for some* $p > 0$*.*

Then the scaling operator $\mathcal{S} : \mathcal{M}_p^* \to \mathcal{M}_p^*$ *is a contraction w.r.t.* ℓ_p^{**}*, and there exists a unique distribution* $\mathcal{P}^* \in \mathcal{M}_p^*$ *which is self-similar w.r.t.* $\mathcal{S}$*.*

Proof. First note that for $\mathcal{P} \in \mathcal{M}_p^*$,

$$|\mathcal{S}\mathcal{P}| = |\mathcal{S}|\,|\mathcal{P}| = |\mathcal{S}|\,P^* = P^*.$$

As in the proof of Theorem 2.6 one shows that if $\mathcal{P} \in \mathcal{M}_p^*$ then $\mathcal{S}\mathcal{P}$ has finite expected p-moment, except that one instead estimates $E\ell_p^p\left(\sum p_i S_i \mu^{(i)}, \sum p_i \left|\mu^{(i)}\right| \delta_a\right)$ or $E\ell_p\left(\sum p_i S_i \mu^{(i)}, \sum p_i \left|\mu^{(i)}\right| \delta_a\right)$, according as $p \ge 1$ or $0 < p < 1$. It follows that $\mathcal{S} : \mathcal{M}_p^* \to \mathcal{M}_p^*$.

The proof of the contraction property of $\mathcal{S}$ is essentially unchanged, and one obtains the same contraction constant as before. The fixed point of $\mathcal{S}$ is the required $\mathcal{P}^*$. □

Since any fixed point of $\mathcal{S}$ in $\overline{\mathcal{M}}_p$ necessarily has mass distribution P^* and so belongs to $\mathcal{M}_p^*$, uniqueness of a fixed point holds also in $\overline{\mathcal{M}}_p$.

The contraction argument in Theorem 4.3 shows that for any $\mathcal{P}_0$ in $\mathcal{M}_p^*$ the sequence of probability distributions $\mathcal{S}^n\mathcal{P}_0$ converges exponentially fast with respect to ℓ_p^{**} to the unique self-similar distribution $\mathcal{P}^*$ in $\mathcal{M}_p^*$. But the mass distribution P^* of $\mathcal{P}^*$ is not normally known apriori, and so we do not necessarily have a candidate for $\mathcal{P}_0$ (which must also have mass distribution P^*). In order to obtain an effective method to approximate $\mathcal{P}^*$ we extend Theorem 4.3 to allow the sequence of random measures to start with *any* initial measure in $\overline{\mathcal{M}}_p$.

Our strategy will be the following. As in Section 3 we change to the special probability space $\widetilde{\Omega}$ of construction trees which allowed us to define the scaling operator $\mathbf{S}$ at the random measure level. For any real random variable X we define

$$|\mathbf{S}|\,X(\omega) = \sum p_i(\omega)\,X(\omega^{(i)}). \tag{4.6}$$

If the random measure μ^* is a fixed point of $\mathbf{S}$, i.e. $\mathbf{S}\mu^* = \mu^*$, then on taking the mass of each side it follows

$$|\mathbf{S}|\,|\mu^*| = |\mu^*|.$$

In Lemma 4.4 we see that there is a unique real random variable X^* with $EX^* = 1$ and $EX^{*2} < \infty$ such that

$$X^* = |\mathbf{S}|X^*.$$

For these reasons, assuming Lemma 4.4, define

$$\begin{aligned}
\overline{\mathbb{M}}_p &= \left\{\, \mu \in \mathbb{M} \;\middle|\; E|\mu| = 1,\ E|\mu|^2 < \infty,\ E_\omega\!\int d^p(x,a)\,d\mu^\omega(x) < \infty \,\right\},\\
\mathbb{M}_p^* &= \left\{\, \mu \in \mathbb{M} \;\middle|\; |\mu| = X^*,\ E_\omega\!\int d^p(x,a)\,d\mu^\omega(x) < \infty \,\right\},
\end{aligned} \tag{4.7}$$

with X^* as in Lemma 4.4. Any fixed point μ^* of $\mathbf{S}$ in $\overline{\mathbb{M}}_p$ belongs to $\mathbb{M}_p^*$. Note that ℓ_p^* is a complete metric on $\mathbb{M}_p^*$; finiteness of ℓ_p^* follows from the fact $\mu, \nu \in \mathbb{M}_p^*$ implies $|\mu| = |\nu| = X^*$ a.s. (see the remarks at the beginning of this section).

The class $\mathcal{X}_2$ of real random variables is now defined by

$$\mathcal{X}_2 = \left\{ X \mid EX = 1,\ EX^2 < \infty \right\}. \tag{4.8}$$

The following Lemma should be compared with Lemma 4.1. Remember that we are now working on the space $\widetilde{\Omega}$.

Lemma 4.4. *If $E\sum p_i = 1$ and $E\sum p_i^2 < 1$ then $|\mathbf{S}|$ is a contraction map on $\mathcal{X}_2$ and so there is a unique $X^* \in \mathcal{X}_2$ such that*

$$X^* = |\mathbf{S}|X^*. \tag{4.9}$$

Proof. One has $|\mathbf{S}| : \mathcal{X}_2 \to \mathcal{X}_2$, since

$$E(|\mathbf{S}|X) = \sum Ep_i(\omega)EX(\omega^{(i)}) = \Big(E\sum p_i\Big)EX = 1,$$

$$E\Big(\sum p_i(\omega)X(\omega^{(i)})\Big)^2 \le E\left(\sum p_i^2(\omega)\sum X^2(\omega^{(i)})\right) < \infty.$$

Moreover, $|\mathbf{S}|$ is a contraction map on $\mathcal{X}_2$ in the L^2-sense, since

$$\begin{aligned} E\big(|\mathbf{S}|X - |\mathbf{S}|Y\big)^2 &= E\Big(\sum_i p_i(\omega)\Big(X(\omega^{(i)}) - Y(\omega^{(i)})\Big)\Big)^2 \\ &= \sum_i Ep_i^2(\omega)E\Big(X(\omega^{(i)}) - Y(\omega^{(i)})\Big)^2 \\ &= \nu E(X-Y)^2, \end{aligned}$$

where $\nu = E\sum p_i^2$, using independence properties and the fact

$$E\big(X(\omega^{(i)}) - Y(\omega^{(i)})\big) = E\big(X(\omega) - Y(\omega)\big) = 1 - 1 = 0.$$

It follows that $|\mathbf{S}|$ has a unique fixed point in $\mathcal{X}_2$ which we denote by X^*. □

We say X^* is a *self-similar random variable*. Note that this notion, unlike the notion of self-similarity of a real probability distribution P^*, depends on the particular sample space $(\widetilde{\Omega}, \widetilde{\Sigma})$.

Since $|\mu_0| \in \mathcal{X}_2$ if μ_0 is a fixed (non-random) unit mass measure, and $|\mu_{n+1}| = |\mathbf{S}|\,|\mu_n|$ from (3.6), it follows from the proof of Lemma 4.4 that $|\mu_n| \in \mathcal{X}_2$ for all n. In particular, $E|\mu_n| = 1$ and $E|\mu_n|^2 < \infty$. Moreover, $|\mu_n| \to X^*$ exponentially fast in the L^2-metric as $n \to \infty$, and hence $|\mu_n| \to X^*$ a.s.

In the next theorem the existence of a unique fixed point $\mu^* \in \mathbb{M}_p^*$ (and hence of a unique fixed point in $\overline{\mathbb{M}}_p$) for $\mathbf{S}$ follows once we show $\mathbf{S}$ is a contraction map on $\mathbb{M}_p^*$. It then follows that $\mu^* \overset{d}{=} \mathcal{P}^*$ where $\mathcal{P}^*$ is as in Theorem 4.3.

Since the random measures $\mu_n = \mathbf{S}^n\mu_0$ need not have random mass X^* a.s. we introduce a reweighted sequence $\overline{\mu}_n$ for which it *is* true that $|\overline{\mu}_n| = X^*$ a.s., and then prove by contraction arguments that $\overline{\mu}_n \to \mu^*$. In the final step the reweighted sequence $\overline{\mu}_n$ and μ_n are compared in order to imply a.s. convergence of μ_n to μ^*.

By $C^{0,1}$ is meant the space of Lipschitz functions $f : \mathfrak{X} \to \mathbb{R}$

Theorem 4.5. *Let $\mathbf{S}$ be a random scaling law such that $E\sum p_i = 1$, $E\sum p_i^2 < 1$, $\lambda_p := E\sum p_i r_i^p < 1$ and $E\sum p_i d^p(S_i a, a) < \infty$, for some $p > 0$.*

Then for any $\mu_0 \in M_p$ the sequence of random measures (μ_n) converges a.s. in the weak sense of measures to some $\mu^* \in \mathbb{M}_p^*$ with $\mu^* \overset{d}{=} \mathcal{P}^*$, where $\mathcal{P}^*$ is the unique self-similar probability distribution in $\mathcal{M}_p^*$. If $p \ge 1$ and $f \in C^{0,1}$, or $0 < p < 1$ and $f \in C^{0,1}$ has bounded support, then $E|\mu_n(f) - \mu^*(f)| \to 0$ exponentially fast.

Proof.
Step 1: Contraction on the space $\mathbb{M}_p^*$.

We claim that

$$\mathbf{S} : \mathbb{M}_p^* \to \mathbb{M}_p^*. \tag{4.10}$$

First note that $|\mu| = X^*$ automatically ensures $E|\mu| = 1$ and $E|\mu|^2 < \infty$ since $X^* \in \mathcal{X}_2$. But $|\mu| = X^*$ also implies $|\mathbf{S}\mu| = |\mathbf{S}|\,|\mu| = |\mathbf{S}|X^* = X^*$, and so to establish (4.10) we need only show that $E\int d^p(x,a)\,d\mathbf{S}\mu < \infty$. This follows by an argument similar to that in the proof of Theorem 2.6 (see also Theorem 4.3).

We also claim

$$\mathbf{S} \text{ is a contraction map on } \mathbb{M}_p^* \text{ with Lipschitz constant } \lambda_p^{\frac{1}{p}\wedge 1}. \tag{4.11}$$

This follows from an argument similar to that in Theorem 2.6 (see also Theorem 4.3). It follows that $\mathbf{S}$ has a unique fixed point in $\mathbb{M}_p^*$, which we denote by μ^*. We say μ^* is a *self-similar random measure*. Again, this notion depends on the particular sample space $(\widetilde{\Omega}, \widetilde{\Sigma})$.

Step 2: The reweighted sequence $\{\overline{\mu}_n(\omega)\}$

As noted before, we cannot directly deduce the convergence of (μ_n) from the contraction property of $\mathbf{S}$, since it is not necessarily the case that $|\mu_n| = X^*$ a.s., i.e. that $\mu_n \in \mathbb{M}_p^*$. For this reason we first consider the "reweighted" sequence

$$\begin{aligned}
&\overline{\mu}_0(\omega) = X^*(\omega)\mu_0,\\
&\overline{\mu}_n(\omega) = \sum_{|\sigma|=n} X^*(\omega^\sigma)\,\overline{p}^\sigma(\omega)\overline{S}^\sigma(\omega)\mu_0 \quad \text{for } n \geq 1.
\end{aligned} \tag{4.12}$$

Comparing (3.2) with (4.12) we see that $\overline{\mu}_n(\omega)$ is obtained by weighting each of the N^n components in $\mu_n(\omega)$, corresponding to the nodes σ with $|\sigma| = n$, with the factor $X^*(\omega^\sigma)$. In particular, μ_n and $\overline{\mu}_n$ have the same support $\bigcup_{|\sigma|=n} \overline{S}^\sigma(\omega)[\text{spt } \mu_0]$ a.s.

We next compute from (4.12), (3.4), (3.3) and (3.5) that

$$\begin{aligned}
\overline{\mu}_{n+1}(\omega) &= \sum_{i=1}^N \sum_{|\sigma|=n} X^*(\omega^{i*\sigma})\,\overline{p}^{i*\sigma}(\omega)\overline{S}^{i*\omega}(\omega)\mu_0\\
&= \sum_{i=1}^N p_i(\omega)S_i(\omega)\left(\sum_{|\sigma|=n} X^*((\omega^{(i)})^\sigma)\,\overline{p}^\sigma(\omega^{(i)})\,\overline{S}^\sigma(\omega^{(i)})\mu_0\right)\\
&= \sum_{i=1}^N p_i(\omega)S_i(\omega)\overline{\mu}_n(\omega^{(i)}) = \mathbf{S}\overline{\mu}_n(\omega).
\end{aligned} \tag{4.13}$$

In particular, $|\overline{\mu}_{n+1}| = |\mathbf{S}|\,|\overline{\mu}_n|$, and since $\overline{\mu}_0 \in \mathbb{M}_p^*$ it follows from (4.10) and (4.13) that $\overline{\mu}_n \in \mathbb{M}_p^*$ for all n, and hence $|\overline{\mu}_n| = X^*$ a.s. Moreover, $\overline{\mu}_n \to \mu^*$ in the ℓ_p^* metric, and from (4.11),

$$\ell_p^*(\overline{\mu}_n, \mu^*) \leq \frac{\lambda_p^{n(\frac{1}{p}\wedge 1)}}{1 - \lambda_p^{\frac{1}{p}\wedge 1}}\,\ell_p^*(\overline{\mu}_0, \overline{\mu}_1). \tag{4.14}$$

Next suppose $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz. For future reference we compute, choosing γ optimal for the pair $(\overline{\mu}_n, \mu^*)$, and using the facts $|\gamma| = |\overline{\mu}_n| = |\mu^*| = X^*$ a.s., $EX^* = 1$ and (4.14), that for $p \geq 1$

$$
\begin{aligned}
E\left|\int f\,d\overline{\mu}_n - \int f\,d\mu^*\right| &= E\left|\int (f(x)-f(y))\,d\gamma\right| \le (\mathrm{Lip}\,f)E\int d(x,y)\,d\gamma \\
&\le (\mathrm{Lip}\,f)E\left(\left(\int d^p(x,y)\,d\gamma\right)^{\frac{1}{p}} |X^*|^{1-\frac{1}{p}}\right) \\
&\le (\mathrm{Lip}\,f)\left(E\int d^p(x,y)\,d\gamma\right)^{\frac{1}{p}} |EX^*|^{1-\frac{1}{p}} \\
&= (\mathrm{Lip}\,f)\,\ell_p^*(\overline{\mu}_n,\mu^*) \le \frac{\lambda_p^{\frac{n}{p}}}{1-\lambda_p^{\frac{1}{p}}}(\mathrm{Lip}\,f)\,\ell_p^*(\overline{\mu}_o,\overline{\mu}_1).
\end{aligned} \tag{4.15}
$$

If $0<p<1$ and $d(x,y)\le M$ for x,y in the support of f, then from the first line in (4.15)

$$
\begin{aligned}
E\left|\int f\,d\overline{\mu}_n - \int f\,d\mu^*\right| &\le (\mathrm{Lip}\,f)M^{1-p}E\int d^p(x,y)\,d\gamma \\
&= (\mathrm{Lip}\,f)M^{1-p}\ell_p^*(\overline{\mu}_n,\mu^*) \\
&\le \frac{\lambda_p^n}{1-\lambda_p}(\mathrm{Lip}\,f)M^{1-p}\ell_p^*(\overline{\mu}_0,\overline{\mu}_1).
\end{aligned} \tag{4.16}
$$

Step 3: Comparison of $\{\overline{\mu}_n\}$ *and* $\{\mu_n\}$

Recall that $\overline{\mu}_n$ is obtained from μ_n by weighting each of its N^n components, corresponding to $\sigma\in C$ where $|\sigma|=n$, by $X^*(\omega^\sigma)$. Using the fact that $EX^*(\omega^\sigma)=1$, we next show that $E|\int f\,d\overline{\mu}_n - \int f\,d\mu_n|\to 0$ exponentially fast as $n\to\infty$ for any $f\in C^{0,1}$.

First note that $\nu := E\sum_i p_i^2 < 1$. From (3.1) and independence properties,

$$
E\sum_{|\sigma|=n}(\overline{p}^\sigma)^2 = E\sum_{1\le\sigma_1,\dots,\sigma_n\le N}(\overline{p}^{\sigma_1\dots\sigma_{n-1}})^2(p_{\sigma_n}^{\sigma_1\dots\sigma_{n-1}})^2 = \nu\sum_{|\tau|=n-1}E(\overline{p}^\tau)^2.
$$

Hence

$$
E\sum_{|\sigma|=n}(\overline{p}^\sigma)^2 = \nu^n \tag{4.17}
$$

for all $n\ge 1$.

In the next computation, recall that $\omega(\sigma)$ and $\omega(\tau)$ are independent if $\sigma\ne\tau$. In particular $\overline{p}^\sigma$ and $\overline{S}^\sigma$ are independent of $X^*(\omega^\sigma)$ since the former two depend on certain $\mathbf{S}^\tau(\omega)=\omega(\tau)$ for $|\tau|<|\sigma|$, while the latter depends on $\omega(\tau)$ for $|\tau|\ge\sigma$ since a.s. $X^*(\omega^\sigma)=\lim_{n\to\infty}(|\mathbf{S}|^n 1)(\omega^\sigma)=\lim_{n\to\infty}\sum_{|\tau|=n}\overline{p}^\tau(\omega^\sigma)=\lim_{n\to\infty}\sum_{|\tau|=n}\overline{p}^{\tau\star\sigma}(\omega)$ (the second equality comes from Lemma 4.4 and the last equality comes from the natural extension of (3.4)). Hence from (4.12) and (3.2)

$$
\begin{aligned}
&\left(E\left|\int f\,d\overline{\mu}_n-\int f\,d\mu_n\right|\right)^2\\
&\quad\le E\left|\int f\,d\overline{\mu}_n-\int f\,d\mu_n\right|^2\\
&\quad=E\left|\int f\,d\Big(\sum_{|\sigma|=n}X^*(\omega^\sigma)\overline{p}^\sigma(\omega)\overline{S}^\sigma(\omega)\mu_0\Big)-\int f\,d\Big(\sum_{|\sigma|=n}\overline{p}^\sigma(\omega)\overline{S}^\sigma(\omega)\mu_0\Big)\right|^2\\
&\quad=E\left|\sum_{|\sigma|=n}\overline{p}^\sigma(\omega)(X^*(\omega^\sigma)-1)\int f\circ\overline{S}^\sigma(\omega)\,d\mu_0\right|^2\\
&\quad=\sum_{|\sigma|=n}E\bigg((\overline{p}^\sigma(\omega))^2\Big(\int f\circ\overline{S}^\sigma(\omega)\,d\mu_0\Big)^2\bigg)E(X^*(\omega^\sigma)-1)^2
\end{aligned}
$$

by the independencies noted above and since $E(X^*(\omega^\sigma)-1)=0$,

$$
\le ||f||^2_{C^0}\nu^n E(X^*(\omega)-1)^2
$$

by (4.17).

From (4.15) and (4.16) it now follows that

$$
E\left|\int f\,d\mu_n-\int f\,d\mu^*\right|\le\begin{cases}c\max^n\{\lambda_p^{\frac{1}{p}},\nu^{1/2}\}\|f\|_{C^{0,1}} & p\ge 1\\ cM^{1-p}\max^n\{\lambda_p,\nu^{1/2}\}\|f\|_{C^{0,1}} & 0<p<1\end{cases}\tag{4.18}
$$

where $c=c(\mathbf{S},\mu_0)$, thus proving exponential convergence.

It follows that if $f\in C^{0,1}$ (with bounded support if $0<p<1$) then $\mu_n(f)\to\mu^*(f)$ a.s. Moreover, by choosing a countable set of such f which is dense in C_c^0, it follows by an approximation argument that $\mu_n(f)\to\mu^*(f)$ for any $f\in C_c^0$, i.e. $\mu_n\to\mu^*$ a.s. in the vague sense of measures. Finally, since $|\mu_n|\to X^*=|\mu^*|$ a.s., it follows that $\mu_n\to\mu^*$ a.s. in the weak sense of measures. □

Remark 4.6. The proof above shows that the exponential convergence is uniform on the class of $f\in C^{0,1}$ with $\operatorname{Lip} f\le 1$ and $\|f\|_\infty\le 1$ if $p\ge 1$, and similarly for $0<p<1$ if the f also have uniformly bounded support.

Remark 4.7. The convergence results in Theorem 4.5 on the space of construction trees implies approximation results in distribution on any probability space for the recursive sequence (μ_n). In particular, uniform exponential convergence is a consequence of (4.18).

Remark 4.8. The convergence of the recursive sequence (μ_n) can be extended to random initial measures μ_0 by considering a product space $\widetilde{\Omega}\times\Omega'$ supplied with product measure $\widetilde{P}\otimes P'$ and letting $\mu_0^{(\widetilde{\omega},\omega')}=\mu_0^{\omega'}$ depend only on ω' while the scaling system (p_i,S_i) just depends and operates on $\widetilde{\omega}$.

References

[1] Arbeiter, M.A., *Random recursive constructions of self-similar fractal measures. The non-compact case*, Prob. Th. Rel. Fields **88** (1991), 497–520.

[2] Barnsley, M.F., Demko, S., *Iterated function systems and the global construction of fractals*, Proc. Roy. Soc. London **A399** (1985), 243–275.

[3] Barnsley, M.F., Demko, S.G., Elton, J.H., Geronimo, J.S., *Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*, Ann. Inst. H. Poincaré Probab. Statist. 24 (1988), 367–394.

[4] Barnsley, M.F., Demko, S.G., Elton, J.H., Geronimo, J.S., *Erratum: Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*, Ann. Inst. H. Poincaré Probab. Statist. **25** (1989), 589–590.

[5] Barnsley, M.F., *Fractals Everywhere*, Academic Press, 1988.

[6] Barnsley, M.F., Elton, J.H., *Stationary attractive measures for a class of Markov chains arising from function iteration*, Preprint, 1985.

[7] Barnsley, M.F., Elton, J.H., *A new class of Markov processes for image encoding*, Adv. in Appl. Probab. **20** (1988), 14–32.

[8] Falconer, K.J., *Random fractals*, Math. Proc. Cambridge Philos. Soc. **100** (1986), 559–582.

[9] Graf, S., *Statistically self-similar fractals*, Prob. Th. Rel. Fields **74** (1987), 357–392.

[10] Graf, S., Mauldin, R.D., Williams, S.C., *The exact Hausdorff dimension in random recursive constructions*, Mem. Amer. Math. Soc. **71** (1988), no. 381.

[11] Hutchinson, J.E., *Fractals and self-similarity*, Indiana. Univ. Math. J. **30** (1981), 713–747.

[12] Hutchinson, J.E., *Deterministic and random fractals*, Complex Systems, Ch. 5, Cambridge Univ. Press, 2000, eds. T.R.J. Bossomaier and D.G. Green.

[13] Hutchinson, J.E., Rüschendorf, L., *Random fractal measures via the contraction method*, Indiana Univ. Math. Journal **47** (1998), 471–487.

[14] Hutchinson, J.E., Rüschendorf, L., *Selfsimilar fractals and selfsimilar random fractals*, Aust. Nat. Univ. Math. Res. Rep. MRR 99.010, or Proceedings of Sec. Int. Conf. on Fractal Geom. and Stochastics, Greifswald, Aug. 28–Sept. 3, 1998 (to appear).

[15] Kahane, J.P., Peyrière, J., *Sur certaines martingales de Mandelbrot*, Adv. Math. **22** (1976), 131–145.

[16] Mauldin, R.D., Williams, S.C., *Random recrusive constructions; asymptotic geometric and topological properties*, Trans. Amer. Math. Soc. **295** (1986), 325–346.

[17] Müller, N., *Zufällige und Multifraktale mit Anwendungen inder Computergrafik*, Masterthesis, Univ. Freiburg.

[18] Olsen, R., Random geometrically graph directed self-similar multifractals. Pitman Research Notes **307**, Longman (1994).

[19] Patzschke, N., Zähle, U., *Self-similar random measures. IV — The recursive construction model of Falconer, Graf, and Mauldin and Williams*, Math. Nachr. **149** (1990), 285–302.

[20] Peitgen, H.O., Saupe, D., *The Science of Fractal Images*, Springer, 1988.

[21] Rachev, S.T., *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991.

[22] Rachev, S.T., Rüschendorf, L., *Probability metrics and recursive algorithms*, Adv. in Appl. Probab. **27** (1995), 770–799.

[23] Rösler, U., *A fixed Point theorem for distributions*, Stoch. Proc. and Appl. **42** (1992), 195–214.

[24] Saupe, D., *Algorithms for random fractals*, in The Science of Fractal Images, 1988, eds. H.O. Peitgen and D. Saupe, 71–113.

[25] Zähle, U., *Self-similar random measures. I—Notion, carrying Hausdorff dimension, and hyperbolic distribution*, Prob. Th. Rel. Fields **80** (1988), 79-100.

School of Mathematical Sciences, Australian National University, Canberra, ACT 0200, Australia

E-mail address: `John.Hutchinson@anu.edu.au`

Institut für Mathematische Stochastik, Universität Freiburg, Eckerstr. 1, D-79104 Freiburg, Germany

E-mail address: `ruschen@galton.mathematik.uni-freiburg.de`