Stochastic Processes and their Applications 42 (1992) 195-214
North-Holland

A fixed point theorem for distributions

Uwe Rösler
Institut für Mathematische Stochastik, Göttingen, Germany

Received 6 November 1990
Revised 13 May 1991

We study in a systematic form the contractive behavior of the map S of distributions to distributions $S(F) \stackrel{\mathscr{D}}{=} \sum_i T_i X_i + C$, $(C, T = (T_1, T_2, \ldots))$, X_i are independent r.v., $L(X_i) = F$. Further we show higher and exponential moments of the fixed point. Applications of this structure are given for (a) weighted branching processes, (b) the Hausdorff dimension of random Cantor sets and (c) the sorting algorithm Quicksort.

1. Introduction

Our main concern is the existence of fixed points for the map S of distribution functions to distribution functions defined by

$$S(F) \stackrel{\mathscr{D}}{=} \sum_i T_i X_i + C.$$

Here $\mathscr{D}$ denotes equality in distribution, $(C, T = (T_1, T_2, \ldots))$, X_i, $i \in \mathbb{N}$, are independent random variables, $X. \stackrel{\mathscr{D}}{=} F$.

These fixed points appear in several quite different examples. They were not considered so far under this point of view. Let us start with some examples.

The normal N(0, 1) distribution is a fixed point for

$$S(F) \stackrel{\mathscr{D}}{=} 2^{-1/2} X_1 + 2^{-1/2} X_2.$$

For another example choose T_1 a uniformly on [0, 1] distributed r.v. and $T_2 = 1 - T_1$, $T_3 = 0 = T_4 = \cdots$, $C(x) = 2x \ln x + 2(1-x)\ln(1-x) + 1$. Then the map

$$S(F) \stackrel{\mathscr{D}}{=} T_1 X_1 + (1 - T_1) X_2 + C(T_1)$$

has a (unique up to translation) fixed point.

The number of comparisons, used by the random sorting algorithm Quicksort to sort a list, correctly normalized converges weakly to this fixed point. Basically the

Correspondence to: Prof. Dr. Uwe Rösler, Institut für Mathematische Stochastik, Lotzestrasse 13, W-3400 Göttingen, Germany.

0304-4149/92/$05.00

number of comparisons is proportional to the time spent by Quicksort to sort a list. The convergence result is therefore a statement on the asymptotic distribution of the time used to sort large lists by Quicksort (Rösler [8]).

Another example is the algorithm 65 'Find' (Devroye [4]).

A further example are branching processes (Athreya and Ney [1]). For given offspring distribution p_k, $k \in \mathbb{N}$, let S_n be the number of individuals in the nth generation, $S_0 = 1$. Then for $\sum_k kp_k =: m > 1$ converges the martingale S_n/m^n to a r.v. W a.e. The distribution of W satisfies a fixed point relation.

For this set-up let $C \equiv 0$, $T = (T_1, T_2, \ldots) \in \{0, 1\}^{\mathbb{N}}$, with $P(\sum_i 1_{T_i=1} = k) = p_k$. Define

$$S(F) \stackrel{\mathscr{L}}{=} \sum_i \frac{T_i}{m} X_i.$$

Notice the distribution of S_n/m^n is S^n (pointmass at 1).

A natural generalisation, introduced here for the first time, is the following. Start a process with one plant of some size. Divide this plant into smaller plants and let them grow for a time period. Then repeat this procedure independent and identical, but on a different scale. Let S_n be the total size of all plants in the nth generation. Again the martingale S_n/m^n, $m := E(\sum_i T_i)$, converges under some conditions to a r.v., which distribution is a fixed point for S. More details are in the example section.

Another example concerns the Hausdorff dimension of random Cantor sets. Divide by random the interval $[0, 1]$ into disjoint intervals, keep some, discard some. The T_i gives the length of the ith interval. Then repeat this procedure for each interval independent and identical on a different scale. The limit set is a random Cantor set. The Hausdorff dimension is the β with $E(\sum_i T_i^\beta) = 1$. This is a result due to Mauldin and Williams [7]. Some of their results are simplified. The distribution of the total mass of the Hausdorff measure is a fixed point for

$$S(F) \stackrel{\mathscr{L}}{=} \sum_i T_i^\beta X_i.$$

In this paper we treat S in a general and systematic form. The map S is under suitable conditions a contraction on the space of distribution functions endowed with the Mallow metric (see Bickel and Freedman [2] for an account). The Mallow distance of two distribution functions is given by the minimal L_p, $1 \leq p < \infty$, distance of two r.v. with those distributions. The existence and uniqueness of a fixed point is then obvious by the contractive behavior. We further discuss higher and exponential moments of the fixed point in the Sections 4 and 5.

2. Mathematical notation

Let $(\Omega, \mathscr{A}, P)$ be a probability space. By L_p, $1 \leq p < \infty$, we denote as usual the set of all random variables $X : \Omega \to \mathbb{R}$ with finite L_p-norm $\|X\|_p = (E(|X|^p))^{1/p}$, identified if they are P a.e. identical. The distribution function, expectation and variance of a r.v. X is denoted by $L(X)$, $\mathrm{Var}(X)$ and $E(X)$. We use also sloppy $E(X) = E(L(X))$

and $\operatorname{Var}(X)=\operatorname{Var}(L(X))$. The set of distribution functions for $X \in L_p$ is called M_p, $1 \leq p<\infty$. The set M_p^b is the subspace of M_p, such that $E(X)=b$.

Throughout the paper the distribution of the r.v. $T: \Omega \rightarrow \mathbb{R}^{\mathbb{N}}$ will be fixed. We allow arbitrary dependence of the coordinates $T_i: \Omega \rightarrow \mathbb{R}$, $T=(T_1, T_2, \ldots)$, $i \in \mathbb{N}$.

For a r.v. $C: \Omega \rightarrow \mathbb{R}$ define a map S of distribution functions to distribution functions, if possible, by

$$S(F)=L\left(\sum_{i \in \mathbb{N}} T_i X_i+C\right).$$

The r.v. $X_i: \Omega \rightarrow \mathbb{R}$, $i \in \mathbb{N}$, (C, T) are independent. The $X_\cdot$ r.v. have all the same distribution function F.

A few words on the existence of S and the meaning of $\sum_i T_i$ are necessary. If all r.v. are positive (in the sense non-negative) then S is always well defined (in a.e. sense). But if we deal with real valued r.v. in general then we have to be more careful. We use here exclusively $\sum_i T_i \in L_p$ if $\sum_{i=1}^N T_i$ is a sequence converging to $\sum_{i=1}^{\infty} T_i$ in L_p-norm if N tends to infinity. Also $\|\sum_i T_i\|_p$ has the above L_p-meaning as a limit.

If $\sum_i |T_i|$, $C \in L_1$, then $S: M_1 \rightarrow M_1$ is well defined (in L_1-sense). If $E(\sum_i T_i^2)<\infty$, $C \in L_2$, then $S: M_2^0 \rightarrow M_2$ is well defined (in L_2-sense). If $\sum_i T_i$, $C \in L_2$, then $S: M_2 \rightarrow M_2$ is well defined (in L_2-sense).

Two examples shall show the difficulties. For the first example take $T_i:=(-1)^i / i^{1/2}$, $C:=0$, F a normal $N(0,1)$ distribution. Then $\sum_i T_i$ is finite in a.e. and L_1-sense and $\sum_i T_i X_i$ is not a proper r.v. For example, the Fourier transform of this formal expression is not the Fourier transform of a r.v.

For the second example let the probability space be the unit interval with the Borel σ-field and the Lebesgue measure. Then define $T_1=1$, $T_2=-1$ on $[0, \frac{1}{2}]$ and $T_2=1$ on $(\frac{1}{2}, 1]$, $T_3=0$ on $[0, \frac{1}{2}]$, $T_3=-2$ on $(\frac{1}{2}, \frac{3}{4}]$ and $T_3=2$ otherwise. In general $T_i=0$ on $[0,1-2^{-i+2}]$, $T_i=-2^{i-1}$ on $(1-2^{-i+2}, 1-2^{-i+1}]$, $T_i=2^{i-1}$ otherwise, $i \geq 3$. Then by construction $\sum_{i=1}^N T_i=0$ on $[0,1-2^{-N+1}]$ and 2^{N-1} otherwise. Therefore $\sum_i T_i=0$ a.e. but $\sum_{i=1}^N T_i$ does not converge to $\sum_i T_i$ in L_2-norm. (This is a standard example for the difference between a.e. and L_2-convergence.) Consequently $W= \sum_i T_i W_i$ has the solution $W=1$ in a.e. sense, but not in L_2-norm (or L_1-norm). (It is easy to give variants with, e.g., $E(\sum_i T_i^2<1)$.)

By our imposed assumptions we will deal basically with $S: M_p \rightarrow M_p$ or $S: M_p^0 \rightarrow M_p^0$, $1 \leq p<\infty$, in L_p-sense. The existence of S follows by an assumption of the form $\|\sum_i T_i\|_2<\infty$. We could weaken this assumption, e.g., by requiring a local L_2-estimate $P(\limsup_M \limsup_N |\sum_{i=M}^N T_i|^2=0)=1$.

On M_p, $1 \leq p<\infty$, we use the Mallow metric d_p,

$$d_p(F, G)=\inf \|X-Y\|_p.$$

The infimum is taken over all r.v. X and Y on any probability space, but with distribution function F and G. The infimum is attained for $X=F^{-1}(U)$, $Y= G^{-1}(U)$, where U is uniformly distributed on $[0,1]$ and F^{-1} is the (left-continuous) inverse of the (right-continuous) distribution function F,

$$F^{-1}(x)=\inf\{y \mid F(y)\geq x\} \quad \text{for } x\in[0,1], \quad \inf\emptyset=\infty.$$

Thus

$$d_p(F,G)=\|F^{-1}(U)-G^{-1}(U)\|_p=\left(\int_0^1 |F^{-1}(u)-G^{-1}(u)|^p \,\mathrm{d}u\right)^{1/p}.$$

The tupel (M_p, d_p) is a polish space, precisely M_p is a complete separable metric space under d_p for each $1\leq p<\infty$.

The metric d_p on M_p is in a natural way a pseudo-metric for r.v. in L_p, $d_p(X,Y):=d_p(L(X),L(Y))$. Notice that d_p-convergence is the same as weak convergence plus convergence of the absolute moment of the order p (Denker and Rösler [3] and see also Bickel and Freedman [2]).

3. S is a contraction

Our main concern in this section is the convergence of the sequence F, $S(F)$, $S^2(F),\ldots$ in d_2-metric. Let us first state an easy lemma on the first moment.

Lemma 1. *Assume $\sum_i |T_i|$, $C\in L_1$, $F\in M_1$. Then $S: M_1\to M_1$ is well defined and the following is true*:

(1) *If $E(C)=0$ then $S: M_1^0\to M_1^0$.*

(2) *If $|E(\sum_i T_i)|<1$ then $E(S^n(F))$ converges exponentially fast to $E(C)(1-E(\sum_i T_i))^{-1}$ as n tends to infinity.*

(3) *If $|E(\sum_i T_i)|>1$ and $E(C)E(F)\neq 0$ then $\lim_{n\to\infty}|E(S^n(F))|=\infty$.*

(4) *If $E(\sum_i T_i)=1$ and $E(C)=0$ then $E(F)=E(S^n(F))$ for all n.*

(5) *If $E(\sum_i T_i)=1$ and $E(C)\neq 0$ then $\lim_{n\to\infty}|E(S^n(F))|=\infty$.*

(6) *If $E(\sum_i T_i)=-1$ then for all $n\in\mathbb{N}$, $E(S^{2n}(F))=E(F)$, $E(S^{2n+1}(F))=E(C)-E(F)$.*

Proof. All statements follow by the obvious relation

$$E(S^{n+1}(F))=E(S^n(F))E\left(\sum_i T_i\right)+E(C).$$

For the second establish first

$$E(S^{n+1}(F)-S^n(F))=E\left(\sum_i T_i\right)(E(S^n(F))-E(S^{n-1}(F)))$$

and argue for $m<n$,

$$\begin{aligned}E(S^m(F))-E(S^n(F))&=\sum_{j=m}^{n-1}(E(S^{j+1}(F))-E(S^j(F)))\\&=\sum_{j=m}^{n-1}c^j(E(S(F))-E(F)),\end{aligned}$$

with $c := E(\sum_i T_i)$. Therefore the Cauchy sequence $E(S^n(F))$, $n \in \mathbb{N}$, converges to some value v. This convergence is exponentially fast

$$|E(S^n(F)) - v| \leq \left| \sum_{j=n}^{\infty} c^j (E(S(F)) - E(F)) \right|$$
$$\leq |c|^n (1-|c|)^{-1} (E(S(F)) - E(F)).$$

The value v can be calculated by $v = E(\sum_i T_i)v + E(C)$. □

Denote by F^b the translate $F(\cdot + a)$ for some $a \in \mathbb{R}$, of a distribution $F \in M_1$, such that $E(F^b) = b$.

Lemma 2. *Assume $C \in L_2$, F, $G \in M_2$.*

(1) *If $E(\sum_i T_i^2) < \infty$ then $S : M_2^0 \to M_2$ is well defined and*

$$d_2^2((S(F^0)), (S(G^0))) \leq E\left(\sum_i T_i^2\right) d_2^2(F^0, G^0).$$

(2) *If $\sum_i T_i \in L_2$ then $S : M_2 \to M_2$ is well defined and*

$$d_2^2(S(F), S(G)) \leq E\left(\sum_i T_i^2\right) d_2^2(F^0, G^0) + \left\| \sum_i T_i \right\|_2^2 (E(F) - E(G))^2.$$

(3) *In general*

$$d_2(F^0, G^0) \leq d_2(F, G) \leq d_2(F^0, G^0) + |E(F) - E(G)|.$$

Proof. We first have to show $S : M_2^0 \to M_2$ or $S : M_2 \to M_2$ is well defined. This follows by $\sum_{i=1}^N T_i X_i$, X_i independent r.v. with distribution function $F \in M_2^0$, M_2, is a Cauchy sequence in L_2.

Choose independent r.v. (C, T), (X_i, Y_i), $i \in \mathbb{N}$, with distribution function $L(X_i) = F$, $L(Y_i) = G$, $\|X_i - Y_i\|_2 = d_2(F, G)$. Then

$$\begin{aligned} d_2^2(S(F), S(G)) &\leq \left\| \sum_i T_i X_i + C - \sum_i T_i Y_i - C \right\|_2^2 \\ &= E\left(\sum_i T_i (X_i - Y_i)\right)^2 \\ &= E\left(\sum_i T_i^2\right) E[(X_1 - E(X_1) - Y_1 + E(Y_1))^2] \\ &\quad + E\left[\left(\sum_i T_i\right)^2\right] (E(X_1) - E(Y_1))^2 \\ &= E\left(\sum_i T_i^2\right) d_2^2(F^0, G^0) + E\left[\left(\sum_i T_i\right)^2\right] (E(F) - E(G))^2. \end{aligned}$$

In case $E(F) = E(G)$ we obtain the first inequality.

The last statement is standard. □

Theorem 3. *Assume*

$$E\Big(\sum_i T_i^2\Big)<1, \quad E(C)=0, \quad \|C\|_2<\infty.$$

Then $S: M_2^0 \to M_2^0$ is well defined and has a unique fixed point. The sequence F, $S(F)$, $S^2(F), \ldots$ converges for every $F\in M_2^0$ in d_2-metric exponentially fast to the fixed point of $S: M_2^0\to M_2^0$.

Proof. The previous lemma states S is a contraction. For $c:=(E(\sum_i T_i^2))^{1/2}$ we obtain, $m\leqslant n$,

$$\begin{aligned} d_2(S^m(F), S^n(F)) &\leqslant \sum_{j=m}^{n-1} d_2(S^j(F), S^{j+1}(F)) \\ &\leqslant \sum_{j=m}^{n-1} c^j d_2(F, S(F)) \leqslant \frac{c^m}{1-c} d_2(F, S(F)). \end{aligned}$$

Therefore $S^n(F)$, $n\in\mathbb{N}$, is a Cauchy sequence and converges exponentially fast to some limit in M_2^0. This limit is a fixed point. The fixed point is unique for the contraction is strict. □

Theorem 4. *Assume*

$$E\Big(\sum_i T_i^2\Big)<1, \quad \Big\|\sum_i T_i\Big\|_2<\infty, \quad \Big|E\Big(\sum_i T_i\Big)\Big|<1, \quad \|C\|_2<\infty.$$

Then $S: M_2\to M_2$ is well defined. For every $F\in M_2$ the sequence $S^n(F)$, $n\in\mathbb{N}$, converges in d_2-metric exponentially fast to the unique fixed point of S.

Proof. The map $S: M_2\to M_2$ is well defined. Then for $b_n^0 := d_2^2[(S^{n+1}(F))^0, (S^n(F))^0]$ we obtain $b_n^0\leqslant c b_{n-1}^0$ with $c:=(E(\sum_i T_i^2))^{1/2}$ by Lemma 2. Therefore b_n^0 converges exponentially fast to zero as n tends to infinity. By the second inequality in Lemma 2 and the second statement of Lemma 1 we obtain also that $d_2^2[(S^{n+1}(F)), (S^n(F))]$ converges exponentially fast to zero as n tends to infinity.

The sequence F, $S(F), S^2(F), \ldots$ is a Cauchy sequence in M_2 and therefore convergent. The limit distribution G is in M_2 and a fixed point of S. The convergence is exponentially fast. The fixed point is unique, because the contraction is strict. □

Remark. If we replace in Theorem 4 the assumption $|E(\sum_i T_i)|<1$ by $E(\sum_i T_i)=1$ and $E(C)=0$, then we obtain exponentially fast convergence of F, $S(F)$, $S^2(F), \ldots$

to a fixed point of $S: M_2 \to M_2$ which has the same expectation as the starting distribution F. Especially for every $b \in \mathbb{R}$ we have a unique fixed point G^b of S given the first moment b. If $\sum_i T_i$ is identical one then these fixed points G^b are translates of each other.

4. Representation of the fixed point

In this section we shall represent the fixed point as an infinite sum in a.e. sense.

Let $I = \emptyset \cup \bigcup_{n=1}^{\infty} \mathbb{N}^n$, $\mathbb{N} = \{1, 2, \ldots\}$. We shall use the notation

$$\sigma | i = (\sigma_1, \sigma_2, \ldots, \sigma_i) \quad \text{for } \sigma \in \bigcup_{n \geq i} \mathbb{N}^n, \ i = 1, 2, \ldots, \ \sigma | 0 = \emptyset,$$

$$|\sigma| = n \quad \text{the length of } \sigma, \ \sigma = (\sigma_1, \ldots, \sigma_n), \ |\emptyset| = 0.$$

Let $(C(\sigma), T(\sigma))$, $\sigma \in I$, be i.i.d. r.v. Let $X(\sigma)$, $\sigma \in I$, be i.i.d. r.v. with distribution function F and independent of $(C(\sigma), T(\sigma))$, $\sigma \in I$.

Define iterative Y_n, $n \in \mathbb{N} \cup \{0\}$, by $Y_0 := X(\emptyset)$,

$$Y_1 := C(\emptyset) + \sum_{|\sigma|=1} T_{\sigma_1}(\emptyset) X(\sigma),$$

$$Y_2 := C(\emptyset) + \sum_{|\sigma|=1} T_{\sigma_1}(\emptyset) C(\sigma_1) + \sum_{|\sigma|=2} T_{\sigma_1}(\emptyset) T_{\sigma_2}(\sigma | 1) X(\sigma),$$

and so on. In general

$$Y_n := C(\emptyset) + \sum_{k=1}^{n-1} \sum_{\substack{\sigma \in I \\ |\sigma| = k}} C(\sigma) \prod_{j=1}^{k} T_{\sigma_j}(\sigma | (j-1))$$

$$+ \sum_{\substack{\sigma \in I \\ |\sigma| = n}} X(\sigma) \prod_{j=1}^{n} T_{\sigma_j}(\sigma | (j-1)).$$

Rewriting Y_n in the form

$$Y_n = \sum_{i=1}^{\infty} T_i(\emptyset) \Bigg[C(i) + \sum_{k=2}^{n-1} \sum_{\substack{\sigma_1 = i \\ |\sigma| = k}} C(\sigma) \prod_{j=2}^{k} T_{\sigma_j}(\sigma | (j-1)).$$

$$+ \sum_{\substack{\sigma_1 = i \\ |\sigma| = n}} X(\sigma) \prod_{j=2}^{n} T_{\sigma_j}(\sigma | (j-1)) \Bigg] + C(\emptyset),$$

the recursive structure of the sequence of distributions $L(Y_n)$ and the relation $L(Y_n) = S^n(F)$ is obvious.

Theorem 5. *Assume*

$$F \in M_2, \quad E\Big(\sum_i T_i^2\Big) < 1, \quad \Big\|\sum_i T_i\Big\|_2 < \infty,$$

$$\Big|E\Big(\sum_i T_i\Big)\Big| < 1, \quad \|C\|_2 < \infty.$$

Then the process Y_n converges a.e. and in d_2-metric to the r.v.

$$C(\emptyset) + \sum_{k=1}^{\infty} \sum_{\substack{\sigma\in I\\ |\sigma|=k}} C(\sigma) \prod_{j=1}^{k} T_{\sigma_j}(\sigma|(j-1)).$$

The distribution of this r.v. is the unique fixed point of $S: M_2 \to M_2$.

The process

$$\sum_{\substack{\sigma\in I\\ |\sigma|=n}} X(\sigma) \prod_{j=1}^{n} T_{\sigma_j}(\sigma|(j-1)),$$

$n\in\mathbb{N}$, converges a.e. to 0. If F corresponds to the point distribution in $b = E(C)(1-E(\sum_i T_i))^{-1}$ then Y_n, $n\in\mathbb{N}$, is a uniformly square integrable martingale with respect to the σ-fields $\mathcal{F}_n$ generated by all r.v. $C(\sigma)$, $T(\sigma)$, $|\sigma|<n$ and $X(\sigma)$, $\sigma\in I$, $|\sigma|\leqslant n$.

Proof. If F corresponds to the point distribution in b, then Y_n, $n\in\mathbb{N}$, is a martingale.

$$\begin{aligned} E(Y_{n+1}|\mathcal{F}_n) &= C(\emptyset) + \sum_{k=1}^{n-1} \sum_{\substack{\sigma\in I\\ |\sigma|=k}} C(\sigma) \prod_{j=1}^{k} T_{\sigma_j}(\sigma|(j-1)) \\ &\quad + \sum_{\substack{\sigma\in I\\ |\sigma|=n}} E(C(\sigma)) \prod_{j=1}^{n} T_{\sigma_j}(\sigma|(j-1)) \\ &\quad + \sum_{\substack{\sigma\in I\\ |\sigma|=n+1}} E(X(\sigma))E(T_{\sigma_{n+1}}(\sigma|n)) \prod_{j=1}^{n} T_{\sigma_j}(\sigma|(j-1)) \\ &= Y_n. \end{aligned}$$

This martingale is uniformly square integrable by $L(Y_n) = S^n(F)$ and $\lim_{n\to\infty} \mathrm{Var}(Y_n) = \lim_{n\to\infty} \mathrm{Var}(S^n(F)) < \infty$. By the martingale convergence theorem converges Y_n, $n\in\mathbb{N}$, a.e. (and in L_2) to some r.v., which is given in the statement of the theorem.

Next we show

$$Z_n := \sum_{\substack{\sigma\in I\\ |\sigma|=n}} X(\sigma) \prod_{j=1}^{n} T_{\sigma_j}(\sigma|(j-1)),$$

$n \in \mathbb{N}$, converges a.e. to 0. The process $L(Z_n)$ has a recursive structure given by

$$L(Z_n) = \bar{S}^n(F), \quad \bar{S}: M_2 \to M_2,$$

with $\bar{S}(F) = L(\sum_i T_i V_i)$, $L(V_i) = F$ and (C, T), V_i, $i \in \mathbb{N}$, independent. Therefore

$$E(\bar{S}^n(F)) = E\Big(\sum_i T_i\Big) E(\bar{S}^{n-1}(F)) = \Big(E\Big(\sum_i T_i\Big)\Big)^n E(F) \to 0$$

and

$$\operatorname{Var}(\bar{S}^n(F)) = E\Big(\sum_i T_i^2\Big) \operatorname{Var}(\bar{S}^{n-1}(F)) = \Big(E\Big(\sum_i T_i^2\Big)\Big)^n \operatorname{Var}(F) \to 0$$

converges exponentially fast to 0. Estimate for large n,

$$P(|Z_n| > 1/n) \leq P(|Z_n - E(Z_n)| \geq 1/n - E(Z_n))$$
$$\leq \operatorname{Var}(Z_n)(1/n - E(Z_n))^{-2} \leq 1/n^2.$$

The lemma of Borel-Cantelli provides now Z_n converges a.e. to 0.

The rest of the theorem is easy to show. □

5. Exponential moments

In this section we consider exponential moments of F, $S(F)$, $S^2(F), \ldots$ and also of the limit distribution, which is the fixed point. The assumption $\|\sum_i T_i\|_2 < \infty$ ensures the existence of S and could be weakened. The main theorem is the following.

Theorem 6. *Assume*

$$\Big\|\sum_i T_i\Big\|_2 < \infty, \quad E\Big(\sum_i T_i^2\Big) < 1, \quad |T_i| \leq 1,$$

$i \in \mathbb{N}$, *and for* $\lambda \in \mathbb{R}$ *in some open neighborhood of* 0,

$$E(e^{\lambda \sum_i T_i^2}) < \infty, \quad \lim_{N \to \infty} E(e^{\lambda \sum_{i=1}^N T_i}) < 1, \quad E(e^{\lambda C}) < \infty,$$

and either $E(\sum_i T_i) \neq 1$ *or* $E(C) = 0$. *Define* b *by* $E(C) = b(1 - E(\sum_i T_i))$ *if* $E(\sum_i T_i) \neq 1$ *and* b *arbitrary otherwise. Let* F *be a distribution function with*

$$E(e^{\lambda X}) \leq e^{b\lambda + K\lambda^2}.$$

$L(X) = F$, *for some positive* K *in an open neighborhood of* $\lambda = 0$. *Then there exists an open neighborhood of* 0 *and a* $K_0 \in \mathbb{R}$, *such that for all* λ *in this neighborhood and all r.v.* Z *with a distribution* $S^n(F)$ *for some* $n \in \mathbb{N}$ *or the fixed point distribution*

$$E(e^{\lambda Z}) \leq e^{b\lambda + K_0\lambda^2}.$$

Proof. The proof runs by induction on n. Let X be as above. Then for $Y=\sum_i T_iX_i+C$ with (C,T), X_i, $i\in\mathbb{N}$, independent, $L(X_i)=F$,

$$E(\mathrm{e}^{\lambda Y})\leqslant \lim_N E\left(\mathrm{e}^{\lambda C}\prod_{i=1}^N E(\mathrm{e}^{\lambda T_iX_i}\mid T)\right)\leqslant \mathrm{e}^{b\lambda+K\lambda^2}f_K(\lambda),$$

$$f_K(\lambda):=\lim_N E(\mathrm{e}^{\lambda C+b\lambda(\sum_{i=1}^N T_i-1)+K\lambda^2(\sum_{i=1}^N T_i^2-1)}).$$

We will show there exists a K_0, $L_0>0$, such that $f_{K_0}(\lambda)\leqslant 1$ for $|\lambda|\leqslant L_0$.

Calculate for small $|\lambda|$, taking $\mathrm{d}/\mathrm{d}\lambda$ inside the expectation, suppress the N,

$$f_K(0)=1,$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}f_K(\lambda)=E\Bigg(\mathrm{e}^{\lambda C+b\lambda(\sum_i T_i-1)+K\lambda^2(\sum_i T_i^2-1)}\times\left(C+b\left(\sum_i T_i-1\right)+2K\lambda\left(\sum_i T_i^2-1\right)\right)\Bigg),$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}f_K(0)=0,$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}f_K(\lambda)=E\Bigg(\mathrm{e}^{\lambda C+b\lambda(\sum_i T_i-1)+K\lambda^2(\sum_i T_i^2-1)}\times\left(\left(C+b\left(\sum_i T_i-1\right)+2\lambda K\left(\sum_i T_i^2-1\right)\right)^2+2K\left(\sum_i T_i^2-1\right)\right)\Bigg),$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}f_K(0)=E(\mathrm{e}^{\lambda C+b\lambda(\sum_i T_i-1)+K\lambda^2(\sum_i T_i^2-1)}).$$

Choose $K_0\geqslant K$ large enough to ensure $(\mathrm{d}^2/\mathrm{d}\lambda^2)(f_K(0))<0$. Then $f_{K_0}(\lambda)\leqslant 1$ holds in an open neighborhood of $\lambda=0$. Notice

$$E(\mathrm{e}^{\lambda X})\leqslant \mathrm{e}^{b\lambda+K\lambda^2}\leqslant \mathrm{e}^{b\lambda+K_0\lambda^2}.$$

We also have to ensure that all the derivatives exist and are as given above. This is standard.

The proof runs now by induction. Notice that the choice of the neighborhood is independent of the induction step. We skip the details.

We may pass to the limit, because all s-exponential moments, $s<|\lambda|$, for any weakly convergent sequence F_n, $n\in\mathbb{N}$, of distribution functions with bounded $|\lambda|$-exponential moments converge to the s-exponential moment of the limit distribution. □

We state for further use an extension of the above.

Theorem 7. *Assume additionally for some* $L>0$,

$$\sum_i T_i^2 \leq 1, \quad \sum_i T_i^2 \neq 1, \quad E(\mathrm{e}^{3L|C|})<\infty$$

and also $E(\mathrm{e}^{3|bL\sum_i T_i|})<\infty$ *if* $E(\sum_i T_i)\neq 1$. *Then the conclusion of Theorem* 6 *is true for all* $\lambda\in(-L,+L)$.

Proof. The problem reduces to show $f_K(\lambda)<1$ for all $|\lambda|\leq L$ and K sufficiently large. We know this for $|\lambda|$ smaller than or equal to some L_0 by Theorem 6. Estimate for $|\lambda|\in[L_0, L]$,

$$\begin{aligned} f_K(\lambda) &\leq \|\mathrm{e}^{\lambda C}\|_3 \|\mathrm{e}^{\lambda^2 K(\sum_i T_i^2-1)}\|_3 \|\mathrm{e}^{b\lambda(\sum_i T_i-1)}\|_3 \\ &\leq \|\mathrm{e}^{L|C|}\|_3 \|\mathrm{e}^{L_0^2 K(\sum_i T_i^2-1)}\|_3 \|\mathrm{e}^{bL|\sum_i T_i-1|}\|_3. \end{aligned}$$

Then $f_K(\lambda)$ converges to 0 uniformly for $|\lambda|\in[L_0, L]$ as K converges to ∞. This proves the theorem. □

Two applications are the examples on Quicksort and on random Cantor sets.

Remark. In Theorem 6 we assumed for a r.v. X the condition $E(\mathrm{e}^{\lambda X})\leq \mathrm{e}^{b\lambda+K\lambda^2}$. This implies especially all moments exist, $b=E(X)=E(C)(1-E(\sum_i T_i))^{-1}$ and also $\mathrm{Var}(X)\leq 2b^2+2K$. In fact

$$\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}E(X^n)=E(\mathrm{e}^{\lambda X})\leq \mathrm{e}^{b\lambda+K\lambda^2}=\left(\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}b^n\right)\left(\sum_{m=0}^{\infty}\frac{\lambda^{2m}}{m!}K^m\right).$$

For $\lambda=0$ we obtain equality, for the λ-term we compare the coefficients on both sides, i.e., $E(X)=b$, for the λ^2-term we obtain $E(X^2)\leq b^2+2K$.

6. Higher moments

In this section we show d_p-convergence, $1\leq p$, of a sequence $F, S(F), S^2(F), \ldots$. The map S is no longer a contraction in d_p, $p>2$. But a similar contractive behavior of S results from a moment condition $E(\sum_i |T_i|^p)<1$. As a warm up we state a weaker but more elegant result. An application is the algorithm 65 'Find' in Devroye [4].

Theorem 8. *Assume for some* $1\leq p$,

$$\left\|\sum_i |T_i|\right\|_p<1, \quad \|C\|_p<\infty.$$

Then $S: M_p\to M_p$ *has a unique fixed point, which* L_p*-norm is bounded by* $\|C\|_p(1-\|\sum_i |T_i|\|_p)^{-1}$. *The sequence* $F\in M_p$, $S(F)$, $S^2(F), \ldots$ *converges in* d_p*-metric exponentially fast to the unique fixed point.*

Proof. Argue by Jensen's inequality for $t := \sum_i |T_i|$, $c := \|\sum_i |T_i|\|_p$ with an obvious notation

$$\begin{aligned} E\Big(\sum_i |T_i X_i|\Big)^p &= E\bigg(E\bigg(\Big(\sum_i \frac{|T_i|}{t} t|X_i|\Big)^p \,\Big|\, T\bigg)\bigg) \\ &\leq E\bigg(E\Big(\sum_i |T_i| t^{p-1} |X_i|^p \,\Big|\, T\Big)\bigg) \\ &= E(|X_1|^p) E\bigg(\Big(\sum_i |T_i|\Big)^p\bigg) = c^p E(|X_1|^p). \end{aligned}$$

This provides, using $b_m^0 := d_p((S^m(F))^0, (S^{m-1}(F))^0)$, $1 \leq m$,

$$b_m^0 \leq c b_{m-1}^0$$

and an exponential convergence of b_m^0 to zero. Consequently $S: M_p^b \to M_p^b$, $b := E(C)(1 - E(\sum_i T_i))^{-1}$ (recall $M_p^b = \{F \in M_p \mid \int x \, dF(x) = b\}$) is a contraction in d_p-metric.

Next show

$$d_p(S(F), S(G)) \leq d_p((S(F))^0, (S(G))^0) + \Big\|\sum_i T_i\Big\|_p |E(F) - E(G)|.$$

Therefore b_m converges exponentially fast to zero. The results are now standard and we skip the details. □

We present now the main result, stronger than the above for $2 \leq p$. In order to show the basic structure we assume first $E(C) = 0$ and work in M_p^0.

Theorem 9. *Assume for some $2 \leq p \in \mathbb{N}$,*

$$E\Big(\sum_i T_i^2\Big) < 1, \quad E\Big(\sum_i |T_i|^p\Big) < 1, \quad \Big\|\sum_i T_i^2\Big\|_{p/2} < \infty,$$

$$E(C) = 0, \quad \|C\|_p < \infty.$$

Then for every $F \in M_p^0$ the sequences $S^m(F)$, $m \in \mathbb{N}$, converges exponentially fast to the unique fixed point of $S: M_p^0 \to M_p^0$ in every d_q-metric, $1 \leq q \leq p$.

Proof. The assumptions imply for $q \in [2, p]$,

$$E\Big(\sum_i |T_i|^q\Big) \leq \max\Big\{E\Big(\sum_i T_i^2\Big), E\Big(\sum_i |T_i|^p\Big)\Big\} < 1.$$

This follows, e.g., by the inequality

$$a^q \leq a^2\Big(1 - \frac{q-2}{p-2}\Big) + a^p \frac{q-2}{p-2},$$

valid for every $a \geq 0$. (Or use Hölder inequalities.)

We show the theorem plus the condition

$$d_\alpha^\alpha(S^m(F), S^{m-1}(F)) \leq \text{const} \cdot \mathrm{e}^{-\lambda m}$$

for some $\lambda > 0$ and some constant const independent of m for all $\alpha = 2, 3, \ldots, n$, by induction on n.

For $n = 2$ the theorem is true. Now assume the theorem and the additional condition for $n < p$ and we will show the validity for $n+1$.

Let $F \in M_{n+1}^0$ be fixed and U_i, $i \in \mathbb{N}$, be independent uniformly distributed r.v. Notice that $F^{-1}(U_i)$, F^{-1} the (left-continuous inverse of F) has distribution function F. Define

$$Y_{i,m} = (S^m(F))^{-1}(U_i) - (S^{m-1}(F))^{-1}(U_i), \quad m = 1, 2, \ldots.$$

Notice $E(Y_{i,m}) = 0$ and $d_q(S^m(F), S^m(G)) = \| Y_{i,m} \|_q$ for $1 \leq q \leq n+1$. Estimate for $m \geq 1$,

$$\begin{aligned}
&d_{n+1}^{n+1}(S^{m+1}(F), S^m(F)) \\
&\quad \leq E\left(\sum_i T_i Y_{i,m}\right)^{n+1} \\
&\quad = E\left(\sum_i T_i^{n+1} Y_{i,m}^{n+1}\right) \\
&\quad\quad + E\left(\sum_{l=2}^{n+1} \sum_{i_1 < \cdots < i_l \in \mathbb{N}} \sum_{\substack{1 \leq \alpha_1, \ldots, \alpha_l \leq n \\ \sum_{j=1}^l \alpha_j = n+1}} \frac{(n+1)!}{\alpha_1! \cdots \alpha_l!} T_{i_1}^{\alpha_1} Y_{i_1,m}^{\alpha_1} \cdots T_{i_l}^{\alpha_l} Y_{i_l,m}^{\alpha_l}\right) \\
&\quad = E\left(\sum_i T_i^{n+1}\right) d_{n+1}^{n+1}(S^m(F), S^{m-1}(F)) \\
&\quad\quad + \sum_{l=2}^{n} \sum_{i_1 < \cdots < i_l \in \mathbb{N}} \sum_{\substack{2 \leq \alpha_1, \ldots, \alpha_l \leq n \\ \sum_{j=1}^l \alpha_j = n+1}} \frac{(n+1)!}{\alpha_1! \cdots \alpha_l!} E\left(\prod_{j=1}^l T_{i_j}^{\alpha_j}\right) \\
&\qquad\qquad \times \prod_{k=1}^l d_{\alpha_k}^{\alpha_k}(S^m(F), S^{m-1}(F)).
\end{aligned}$$

Further estimate for given α as above using Hölder's inequality

$$\begin{aligned}
\left| \sum_{i \in \mathbb{N}^l} E\left(\prod_{j=1}^l T_{i_j}^{\alpha_j}\right) \right| &\leq E\left(\prod_{j=1}^l \left(\sum_{i \in \mathbb{N}} |T_i|^{\alpha_j}\right)\right) \\
&\leq \prod_j^l \left\| \sum_i |T_i^{\alpha_j}| \right\|_{(n+1)/\alpha_j} \\
&\leq \prod_j^l \left\| \sum_i T_i^2 \right\|_{(n+1)/2}^{\alpha_j/2} \\
&\leq \left\| \sum_i T_i^2 \right\|_{(n+1)/2}^{(n+1)/2} < \infty.
\end{aligned}$$

We obtain the recursive formula $m \geqslant 1$,

$$d_{n+1}^{n+1}(S^{m+1}(F), S^m(F)) \leqslant c d_{n+1}^{n+1}(S^m(F), S^{m-1}(F)) + \text{const} \cdot e^{-\lambda m},$$

with $c := E(\sum_i |T_i|^{n+1}) < 1$ and λ, const independent of m and F. Further notice

$$d_{n+1}(S(F), F) \leqslant \|S(F)\|_{n+1} + \|F\|_{n+1} < \infty.$$

Therefore the sequence $F, S(F), S^2(F), \ldots$ converges exponentially fast to the fixed point of $S: M_{n+1}^0 \to M_{n+1}^0$ and also the induction hypothesis is satisfied for $n+1$.

Finally notice $d_q \leqslant d_r$ for $1 \leqslant q \leqslant r$. □

We remove now the condition $E(C) = 0$.

Theorem 10. *Assume for some* $2 \leqslant p \in \mathbb{N}$,

$$E\left(\sum_i T_i^2\right) < 1, \quad E\left(\sum_i |T_i|^p\right) < 1, \quad \left\|\sum_i T_i^2\right\|_{p/2} < \infty,$$

$$\left|E\left(\sum_i T_i\right)\right| < 1, \quad \left\|\sum_i T_i\right\|_p < \infty, \quad \|C\|_p < \infty.$$

Then for every $F \in M_p$ *the sequence* $S^m(F)$, $m \in \mathbb{N}$, *converges exponentially fast to the unique fixed point of* $S: M_p \to M_p$ *in* d_p*-metric.*

Proof. Again we use an induction similar as above. For $F \in M_{n+1}$ define

$$b_m^0 := d_{n+1}((S^m(F))^0, (S^{m-1}(F))^0).$$

As in the previous theorem establish

$$b_m^0 \leqslant c b_{m-1}^0 + \text{const} \cdot e^{-\lambda m},$$

for some $\lambda > 0$ and a constant independent of m. Therefore b_m^0converges exponentially fast to zero. Then use for distribution functions F, G and $q = n+1$,

$$d_q(S(F), S(G)) \leqslant d_q((S(F))^0, (S(G))^0) + \left\|\sum_i T_i\right\|_q |E(F) - E(G)|.$$

Therefore the sequence $F, S(F), S^2(F), \ldots$ converges exponentially fast in d_{n+1}-metric to the fixed point of $S: M_2 \to M_2$. □

Remark. If we replace $|E(\sum_i T_i)| < 1$ by $E(\sum_i T_i) = 1$ and $E(C) = 0$ in the above theorem then the conclusion of the theorem still holds. See also the remark in Section 3 for the uniqueness of the fixed point.

The last theorem is stronger than the previous two. In fact notice, using the l_p-norm for a sequence $l_p(T)=(\sum_i |T_i|^p)^{1/p}$,

$$E\left(\sum_i T_i^2\right)=\|l_2(T)\|_2^2\le\|l_1(T)\|_2^2\le\|l_1(T)\|_p^2<1,$$

$$E\left(\sum_i |T_i|^p\right)=\|l_p(T)\|_p^p\le\|l_1(T)\|_p^p<1,$$

$$\left\|\sum_i T_i^2\right\|_{p/2}=\|l_2(T)\|_p^2\le\|l_1(T)\|_p^2<1,$$

$$\left|E\left(\sum_i T_i\right)\right|\le\|l_1(T)\|_1\le\|l_1(T)\|_p<1,$$

$$\left\|\sum_i T_i\right\|_p\le\|l_1(T)\|_p<1.$$

7. Quicksort, branching processes and random Cantor sets

In this section we provide three examples.

Example 11. *Quicksort.* The algorithm Quicksort (Hoare [5]) is probably the most used sorting algorithm. Given a list of n (different) numbers, select by random a number (every number has the same probability), and make a list L_1 of numbers smaller and L_2 of numbers larger than the chosen one. For this we need $n-1$ comparisons of real numbers. Then proceed with the lists L_1 and L_2 in the same way, if they have more than one element. Otherwise quit. Finally we end up with an ordered list of the given numbers. This procedure is called Quicksort. (This algorithm 64 is a random algorithm. There are also deterministic versions around, e.g., algorithm 271 'Quickersort' or algorithm 301 'Qsort'.)

Primarily we are interested in the time to run this procedure. Notice that the time is a r.v. Mainly the time is used for comparisons, or at least proportional to the number of comparisons. Let U_n be the total number of comparisons to sort the n numbers into their natural order. The r.v. U_n, $n\in\mathbb{N}$, satisfy the recursive structure

$$L(U_n)=L(U_{Z_n-1}+\bar{U}_{n-Z_n}+n-1).$$

$U_0\equiv 0$, U_i, $\bar{U}_i:\Omega\to\mathbb{N}$, $Z_n:\Omega\to\{1,2,\dots,n\}$, $0\le i\le n$, are independent r.v. for any fixed n. Further $L(U_i)=L(\bar{U}_i)$ and Z_n has a uniform distribution.

From this relation it is possible (e.g., Hoare [5]) to derive the expected number of comparisons, $E(U_n)$ is of the order $2n\ln n$ (see also Knuth [6] and Sedgewick [9] for more details). In fact for any sorting algorithm the expected number of comparisons is at least $n\log_2 n$ by the information theoretic lower bound. Therefore Quicksort seems to be reasonably good.

In the worst case however U_n can be as large as $\frac{1}{2}n^2$ and in the best case as good as $n \log_2 n$ asymptotically. Therefore we are interested in estimates that the number of comparisons is close to the expected number.

The main competitor for Quicksort is Heapsort. Heapsort has a nice upper bound of the worst case of the order $4n \ln n$. The disadvantage is as simulations show that Heapsort is in average slower than Quicksort.

From the probabilistic point of view the best we can ask for are probability estimates for a poor performance of Quicksort (in the sense that U_n is worse than the upper bound for comparisons for Heapsort). For these we like to calculate moments or even exponential moments of U_n for large n. Here is a way to do that.

Define $Y_n := (U_n - E(U_n))/n$. Then we obtain the recursive formula

$$L(Y_n) = L\left(Y_{Z_n-1}\frac{Z_n-1}{n} + \bar{Y}_{n-Z_n}\frac{n-Z_n}{n} + C_n(Z_n)\right),$$

$$C_n(i) := n^{-1}(n-1+E(U_{i-1})+E(U_{n-i})-E(U_n)),$$

$n \in \mathbb{N}$, $i = 1, \ldots, n$. Define $C:[0,1]\to\mathbb{R}$ by

$$C(x) := 2x \ln x + 2(1-x)\ln(1-x)+1,$$

with the continuous extension.

For large n the r.v. Z_n/n can be approximated by a uniformly distributed r.v. τ on $[0,1]$. Some calculation shows $C_n(n\cdot) \approx C(\cdot)$. If we now assume that Y_n converges in distribution to some r.v. Y as n tends to infinity then we expect the relation

$$L(Y) = L(Y\tau + \bar{Y}(1-\tau) + C(\tau)),$$

$L(Y) = L(\bar{Y})$, $Y, \bar{Y}, \tau : \Omega \to \mathbb{R}$ independent, τ uniformly distributed on $[0,1]$. The distribution of Y is a fixed point of a map S as used in the paper. Take

$$T_1 = \tau, \quad T_2 = 1-\tau, \quad 0 = T_3 = T_4 = \cdots, \quad C = C(\tau).$$

Obviously

$$\sum_i T_i \equiv 1, \quad E\left(\sum_i |T_i|^p\right) < 1, \quad E(C(\tau)) = 0, \quad \|C(\tau)\|_p < \infty,$$

$1 < p < \infty$. Therefore we may apply our results including the remarks. The map S has a (unique up to translation) fixed point in d_p-metric, p any value in $(1,\infty)$. This fixed point has also exponential moments of every order by Section 5.

In a last step we have to show d_p-convergence of Y_n to Y and the convergence of the exponential moments. This can be done. Details are in Rösler [8].

A consequence are probability estimates for poor performance of Quicksort by Chebychev's inequality using exponential moments

$$P(U_n \geq 4n \ln n) \leq \text{const}(\lambda) n^{-\lambda}$$

for any $\lambda \in \mathbb{R}$. This shows that the exceptional set is rather small and Quicksort is a reliable algorithm (as the computer scientists believed all the time).

The next two examples have a lot in common, although they look quite different. Let $I = \emptyset \cup \bigcup_{n\in\mathbb{N}} \mathbb{N}^n$ be the index set. Let $(C(\sigma), T(\sigma))$, $T(\sigma): \Omega \to \mathbb{R}^{\mathbb{N}}$, $C(\sigma): \Omega \to \mathbb{R}$, $\sigma \in I$, be independent identically distributed random variables. Define $l: \Omega \times I \to \mathbb{R}$, $S: \Omega \times \mathbb{N} \to \mathbb{R}$ by $l_\emptyset := 1 =: S_0$,

$$l_{\sigma i} := l_\sigma T_i(\sigma), \quad i \in \mathbb{N},\ \sigma \in I,$$

$$S_n := \sum_{\sigma\in\mathbb{N}^n} l_\sigma, \quad n \in \mathbb{N}.$$

The notation σi and $i\sigma$, $i \in \mathbb{N}$, $\sigma \in I$, has the obvious meaning. If S_n is well defined (in both examples all r.v. are positive) then S_n satisfies the recursive relation

$$L(S_{n+1}) = L\Bigl(\sum_{\sigma\in\mathbb{N}^n} \sum_{i\in\mathbb{N}} l_{i\sigma}\Bigr) = L\Bigl(\sum_{i\in\mathbb{N}} T_i(\emptyset) X_i\Bigr),$$

with X_i, $i \in \mathbb{N}$, r.v. independent of everything before and distributed as S_n. Notice $L(l_{\sigma i}) = L(l_{i\sigma}) = L(T_i(\emptyset) l_\sigma)$.

In both examples we will take $C = 0$ and $T. \geq 0$.

Example 12. *Weighted branching processes.* Let us recall some facts from branching processes. Start a branching process with one individuum. This individuum has a random number of children, the offspring. Each child, member of this first generation, has independent of the others again a random number of children. The distribution of the offspring is the same for all individuals. In general each member of the nth generation has an offspring independent of the others and with the same distribution. In branching processes we are interested in the number of individuals in the nth generation.

We generalize this concept. Think of dividing a mother plant into some smaller pieces. Then let these pieces grow for some time. Divide each of these descendants again and let the pieces grow. We keep track of the sum of all sizes of the individuals in the nth generation.

In our set-up each member of the nth generation has again a number of children of a certain size depending on the size of the parent. The partitioning and the growing procedure are random and indendent for each individual. The partitioning procedure is identical, but on a different scale (selfsimilar). The growing consists of a multiplication of the size with a random factor.

A special case of this set-up are branching processes. Assume here additionally $T(\sigma): \Omega \to \{0, 1\}^{\mathbb{N}}$, $P(\sum_i T_i < \infty) = 1$. All individuals have the size 1 or 0. The process S_n counts the number of individuals of the size 1 at generation n. This is a branching process with offspring distribution $p_k := P(\sum_i T_i(\emptyset) = k)$, $k = 0, 1, \ldots$. The value p_k is the probability of k descendants of one individual.

Recall a few elementary results in branching processes. For simplicity we exclude the trivial case of exactly one offspring ($p_1 \neq 1$). Branching processes with an expected number m ($m := \sum_k kp_k$) of offspring of one individual less than or equal to 1 will die out almost everywhere, i.e., $\lim_{n\to\infty} S_n = 0$ a.e. In case of $1 < m < \infty$ the extinction probability $q = P(\lim_n S_n = 0)$ is the largest fixed point in $[0, 1)$ of the generating function $f(s) := \sum_k s^k p_k$. Further S_n/m^n is a martingale converging a.e. (and in L_2 if $\sum_k k^2 p_k < \infty$) to some non-degenerate random variable W,

$$S_n/m^n \to W.$$

Using generating functions $g(s) := E(s^W)$, $s \in [0, 1]$, we obtain the well known formula

$$g(s) = E(E(s^{\sum_i T_i W_i/m} \mid T)) = E(g^{\sum_i T_i}(s^{1/m})) = f(g(s^{1/m})).$$

For our more general set-up assume additionally $T. \geq 0$ and $C = 0$ for we have no immigration. Notice that S is always well defined in a.e. sense. Define $m := \sum_i T_i$, $p_k := P(\sum_i 1_{T_i > 0} = k)$, $k = 0, 1, \ldots, \infty$. Exclude the trivial case of always one offspring, i.e., $p_1 = 1$. Then the crucial assumptions for our theorems are

$$\left\| \sum_i T_i \right\|_2 < \infty, \quad E\left(\sum_i T_i/m \right) = 1, \quad E\left(\sum_i T_i^2/m^2 \right) < 1.$$

The first one is an existence assumption (and could be weakened). The second serves as a definition of m. The third is the main assumption. Notice this is satisfied for, e.g., if $T_i \leq 1$ or for ordinary branching processes $T_i \in \{0, 1\}$. Notice for branching processes $E(\sum_i T_i^2/m^2) = 1/m$.

S_n/m^n is a nonnegative square integrable martingale. This converges a.e. and in d_2-metric (Theorem 10 offers a d_p version) to some non-degenerate limit r.v. W. The distribution function of W is the (unique by given first moment 1 and finite variance) fixed point of $S: M_2 \to M_2$. This fixed point $L(W)$ satisfies the well known relation

$$L(W) = L\left(\sum_i \frac{T_i}{m} W_i \right),$$

$L(W_i) = L(W)$, $i \in \mathbb{N}$. From this equation one can calculate the second (or higher) moments.

The corresponding formula for the generating function g is

$$g(s) = E\left(\prod_i g(s^{T_i/m}) \right).$$

The structure is not as easy as before. From this we obtain, e.g., the extinction probability

$$P(W = 0) = \sum_{k=0,1,\ldots,\infty} p_k P^k(W = 0).$$

Like for ordinary branching processes $P(W = 0) = q$ appears as a fixed point of some generating function $f(s) := \sum_{k=0,1,\ldots,\infty} s^k p_k$. This equation has the solution $q = 1$ and at most one more in $[0, 1)$. If $p_0 = 0$ then $q = 0$ is the other. If $p_0 > 0$ and $p_\infty = 0$, $\sum_{k=0,1,\ldots} k p_k \geq 1$ then there is no other solution q. Otherwise there is exactly one solution $q \in [0, 1)$. Finally argue $P(W = 0)$ is the smallest solution of $f(q) = q$ in $[0, 1]$ by $E(W) = 1$.

Example 13. *Hausdorff dimension of random Cantor sets.* Recall the construction of a Cantor set. Let $0<a<b<1$ be fixed constants. Divide $J=[0,1]$ into $J_1=[0,a]$, $J_2=[b,1]$ and $J_{-1}=(a,b)$. Discard J_{-1}. Repeat this procedure. In a next step divide J_1 and J_2 again with the same proportions into $J_{ij}\subset J_i$, $i\in\{1,2\}$, $j\in\{-1,1,2\}$. $J_{i,-1}$ is discarded.

Continuing this procedure we obtain a family J_σ, $\sigma\in\{1,2\}^n$, $n\in\mathbb{N}$, of sets. The Cantor set is defined by

$$B:=\bigcap_n \bigcup_{\sigma\in\{1,2\}^n} J_\sigma.$$

Recall the Hausdorff dimension $\mathscr{H}(B)$ of a set B in a metric space is

$$\mathscr{H}(B):=\lim_{\varepsilon\to 0}\inf\left\{\beta \;\middle|\; \sum_{i\in\mathbb{N}}(\operatorname{diam} U_i)^\beta<\infty,\ B\subset\bigcup_{i\in\mathbb{N}} U_i,\ \operatorname{diam} U_i\leqslant\varepsilon,\ i\in\mathbb{N}\right\}.$$

Here diam denotes the diameter of the balls U_i. The Hausdorff dimension of the standard Cantor set is known to be the $\beta\in(0,1)$ with $a^\beta+b^\beta=1$, ($\beta=\ln 2/\ln 3$).

Suppose a and b are randomly chosen or the partition is random in some sense. What is the Hausdorff dimension of the now random Cantor set? This is the general setting of Mauldin and Williams [7].

By some random construction we divide a compact perfect set $J\subset[0,1]^d$ into a countable number of disjoint sets J_i, $i=-1,1,2,\ldots$. The J_i, $i=1,2,\ldots$, are assumed to be compact and perfect. J_{-1} will always be discarded. Starting with $[0,1]^d$ and repeating this procedure, we obtain a family J_σ, $\sigma\in\mathbb{N}^n$, $n\in\mathbb{N}$, of compact perfect sets. Then

$$B:=\bigcap_n \bigcup_{\sigma\in\mathbb{N}^n} J_\sigma$$

is a random Cantor set.

Define the r.v. $T(\sigma):\Omega\to\mathbb{R}^{\mathbb{N}}$, $\sigma\in\bigcup_n\mathbb{N}^n$, by

$$T(\sigma)=(T_i(\sigma))_{i=1,2,\ldots}=\left(\frac{m(J_{\sigma i})}{m(J_\sigma)}\right)_{i=1,2,\ldots}$$

Here m is some measure on $[0,1]^d$, or, as in this set-up, the diameter.

The only assumption on the random construction we shall make is all $T(\sigma)$ are independent identically distributed r.v. We exclude trivial cases of one or less successors $P(\sum_i 1_{T_i>0}>1)>0$.

Then the function $\gamma\to\phi(\gamma):=E(\sum_i T_i^\gamma)$ is nonincreasing and right-continuous in γ, strictly decreasing and continuous on $\{\gamma\mid\phi(\gamma)<\infty\}$. Define $\beta:=\inf\{\gamma\mid\phi(\gamma)\leqslant 1\}$. Then $0<\beta\leqslant 1$ and $\phi(\beta)\leqslant 1$. Mauldin and Williams showed that this β is a.e. the Hausdorff dimension of the (non-empty) random Cantor set. This result includes the standard Cantor set as a special case. That the Hausdorff dimension is smaller than or equal to β follows in their proof by a moment estimate on

$$S_n:=\sum_{\sigma\in\mathbb{N}^n} m^\gamma(J_\sigma)=\sum_{i_1,\ldots,i_n}(T_{i_1}T_{i_2}(i_1)\cdots T_{i_n}(i_1i_2\cdots i_{n-1}))^\gamma.$$

We shall show that this is a consequence of our theorems.

Rewriting S_n it is easy to see the relation

$$L(S_n) = L\left(\sum_{i\in\mathbb{N}} T_i^\gamma X_i\right),$$

where the X_i, $i \in \mathbb{N}$, are independent r.v. with distribution $L(S_{n-1})$. This is the recursive structure for our set-up. The T^γ plays the role of the T, the r.v. C is identical 0. We may apply our results. Notice $L(S_n) = S^n$ (point mass at 1) and also $0 \leqslant T.(\cdot) \leqslant 1$.

If $\phi(\gamma) > 1$ then there is no fixed point and no limit distribution of S_n with finite first moment. Therefore assume $\phi(\gamma) \leqslant 1$ and show the sufficient conditions of Theorem 4 including the remark. In fact

$$\phi(2\gamma) = E\left(\sum_i T_i^{2\gamma}\right) < \phi(\gamma) \leqslant 1.$$

Therefore the distribution of S_n converges in d_2-metric to some fixed point G. In Section 5 we showed even the convergence of higher or exponential moments of every order under suitable conditions.

If $\phi(\gamma) < 1$ then G is the point distribution in 0. The process S_n is a nonnegative supermartingale and converges a.e. to 0.

If $\phi(\gamma) = 1$ then $E(S_n)$ is constant $1 = E(G) = E(S_n)$. The process S_n is a nonnegative martingale and converges a.e. to some r.v. W. The distribution function is specified by

$$L(W) = L\left(\sum_i T_i^\gamma W_i\right),$$

W_i, $i \in \mathbb{N}$, independent identical $L(W)$ distributed r.v. $P(W = 0)$ can be calculated as in the example on branching processes.

References

[1] K.B. Athreya and P. Ney, Branching Processes (Springer, Berlin, 1972).

[2] P.J. Bickel and D.A. Freedman, Some asymptotic theory for the bootstrap, Ann. Probab. 9 (1981) 1196–1217.

[3] M. Denker and U. Rösler, A moment estimate for rank statistics, J. Statist. Plann. Inference 12 (1985) 269–284.

[4] L. Devroye, Exponential bounds for the running time of a selection algorithm, J. Comput. Syst. Sci. 29 (1984) 1–7.

[5] C.A.R. Hoare, Algorithm 64: Quicksort, Comm. ACM 4(7) (1961) 321.

[6] D.E. Knuth, The Art of Computer Programming, Vol. 3: Sorting and Searching (Addison-Wesley, Reading MA, 1973).

[7] R.D. Mauldin and S.C. Williams, Random recursive construction: Asymptotic geometric and topological properties, Trans. Amer. Math. Soc. 295 (1986) 325–346.

[8] U. Rösler, A limit theorem for 'Quicksort', Inform. Théor. Appl. / Theor. Inform. Appl. 25 (1991) 85–100.

[9] R. Sedgewick, The analysis of Quicksort programs, Acta Inform. 7 (1977) 327–355.