Gerold Alsmeyer

Random Recursive Equations and Their Distributional Fixed Points

January 23, 2012

Springer

Preface

Use the template *preface.tex* together with the Springer document class SVMono (monograph-type books) or SVMult (edited books) to style your preface in the Springer layout.

A preface is a book's preliminary statement, usually written by the *author or editor* of a work, which states its origin, scope, purpose, plan, and intended audience, and which sometimes includes afterthoughts and acknowledgments of assistance.

When written by a person other than the author, it is called a foreword. The preface or foreword is distinct from the introduction, which deals with the subject of the work.

Customarily *acknowledgments* are included as last part of the preface.

Place(s), *Firstname Surname*
month year *Firstname Surname*

Acknowledgements

Use the template *acknow.tex* together with the Springer document class SVMono (monograph-type books) or SVMult (edited books) if you prefer to set your acknowledgement section as a separate chapter instead of including it as last part of your preface.

Contents

Acronyms

Use the template *acronym.tex* together with the Springer document class SVMono (monograph-type books) or SVMult (edited books) to style your list(s) of abbreviations or symbols in the Springer layout.

Lists of abbreviations, symbols and the like are easily formatted with the help of the Springer-enhanced `description` environment.

cdf	(cumulative) distribution function
chf	characteristic function
CLT	central limit theorem
dRi	directly Riemann integrable
FT	Fourier transform
GWP	Galton-Watson process
gf	generating function
iff	if, and only if
iid	independent and identically distributed
i.o.	infinitely often
LT	Laplace transform
MC	Markov chain
mgf	moment generating function
pdf	probability density function
RDE	random difference equation
RP	renewal process
RTP	recursive tree process
RW	random walk
SFPE	stochastic fixed point equation
SLLN	strong law of large numbers
SRP	standard renewal process = zero-delayed renewal process
SRW	standard random walk = zero-delayed random walk
ui	uniformly integrable
WBP	weighted branching process
WLLN	weak law of large numbers

Symbols

$Bern(\theta)$	Bernoulli distribution with parameter $\theta \in (0,1)$
$\beta(a,b)$	Beta distribution with parameters $a,b \in \mathbb{R}_>$
$\beta^*(a,b)$	Beta distribution of the second kind with parameters $a,b \in \mathbb{R}_>$
$Bin(n,\theta)$	Binomial distribution with parameters $n \in \mathbb{N}$ and $\theta \in (0,1)$
δ_a	Dirac distribution in a
$Exp(\theta)$	Exponential distribution with parameter $\theta \in \mathbb{R}_>$
$\Gamma(\alpha,\beta)$	Gamma distribution with parameters $\alpha,\beta \in \mathbb{R}_>$
$Geom(\theta)$	Geometric distribution with parameter $\theta \in (0,1)$
$NBin(n,\theta)$	Negative binomial distribution with parameters $n \in \mathbb{N}$ and $\theta \in \mathbb{R}_>$
$Normal(\mu,\sigma^2)$	Normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_>$
$Cauchy(a,b)$	Cauchy distribution with parameters $a \in \mathbb{R}$ and $b \in \mathbb{R}_>$
$Poisson(\theta)$	Poisson distribution with parameter $\theta \in \mathbb{R}_>$
$\mathscr{S}(\alpha,b)$	Symmetric stable law with index $\alpha \in (0,2]$ and scaling parameter $b \in \mathbb{R}_>$
$\mathscr{S}_+(\alpha,b)$	One-sided stable law with index $\alpha \in (0,1]$ and scaling parameter $b \in \mathbb{R}_>$
$Unif(a,b)$	Uniform distribution on $[a,b]$, $a < b$
$Unif\{x_1,...,x_n\}$	Discrete uniform distribution on the set $\{x_1,...,x_n\}$

Chapter 1
Introduction

Probabilists are often facing the task to determine the asymptotic behavior of a given stochastic sequence $(X_n)_{n\ge 0}$, more precisely, to prove its convergence (in a suitable sense) to a limiting variable X_∞, as $n\to\infty$, and to find or at least provide information about the distribution (law) of X_∞, denoted as $\mathscr{L}(X_\infty)$. Of course, there is no universal approach to accomplish this task, but in situations where $(X_n)_{n\ge 0}$ exhibits some kind of recursive structure, expressed in form of a so-called *random recursive equation*, one is naturally prompted to take advantage of this fact in one way or another. Often, one is led to a *distributional equation* for the limit variable X_∞ of the form

$$\mathscr{L}(X_\infty) \;=\; \mathscr{L}\left(\Psi(X_\infty(1),X_\infty(2),...)\right) \tag{1.1}$$

where $X_\infty(1),X_\infty(2),...$ are independent copies of X_∞ and Ψ denotes a random function independent of these variables. (1.1) constitutes the general form of a so-called *stochastic fixed-point equation (SFPE)*, also called *recursive distributional equation* by ALDOUS & BANDYOPADHYAY in [1]. The distribution of X_∞ is then called a *solution to the SFPE* (1.1).

To provide an introduction of a collection of interesting equations of this kind, some of them related to very classical problems in probability theory, and of the methods needed for their analysis is the main goal of this course. The present chapter is devoted to an informal discussion of a selection of examples that will help the reader to gain a first impression of what is lying ahead.

1.1 A true classic: the central limit problem

Every student with some basic knowledge in theoretical probability knows that, given a sequence of iid real-valued random variables $X,X_1,X_2,...$ with mean 0 and variance 1, the associated sequence of standardized partial sums

$$S_n^* \;:=\; \frac{X_1+...+X_n}{n^{1/2}},\quad n\ge 1$$

converges in distribution to a standard normal random variable Z as $n \to \infty$. This is the classic version of the *central limit theorem (CLT)* and most efficiently proved by making use of characteristic functions (Fourier transforms). Namely, let ϕ denote the chf of X and note that a second order Taylor expansion of ϕ at 0 gives

$$\phi(t) \;=\; 1 - \frac{t^2}{2} + o(t^2) \quad \text{as } t \to 0.$$

Since S_n^* has chf $\psi_n(t) = \phi(n^{-1/2}t)^n$, we now infer

$$\lim_{n\to\infty} \psi_n(t) \;=\; \lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right)\right)^n \;=\; e^{-t^2/2}$$

for all $t \in \mathbb{R}$ and thus the asserted convergence by Lévy's continuity theorem combined with the fact that $e^{-t^2/2}$ is the chf of the standard normal distribution.

Having solved the central limit problem for good in the classical setup of iid random variables, the reader may wonder so far about its connection with random recursive equations. Let us therefore narrow our perspective by assuming that the weak convergence of $\mathscr{L}(S_n^*)$ to a limit law Q with mean 0 and unit variance has already been settled. Then the problem reduces to giving an argument that shows that Q must be the standard normal distribution. To this end, we make the observation that

$$S_{2n}^* \;=\; \frac{S_n^* + S_{n,n}^*}{2^{1/2}}, \tag{1.2}$$

where $S_{n,n}^* := n^{-1/2}(X_{n+1} + \ldots + X_{2n})$ for $n \geq 1$. Since $S_{n,n}^*$ is an independent copy of S_n^*, it follows that

$$\left(S_n^*, S_{n,n}^*\right) \;\stackrel{d}{\to}\; (Z, Z') \quad \text{as } n \to \infty$$

for two independent random variables Z, Z' with common distribution Q and then from (1.2) that, by the continuous mapping theorem,

$$Z \;\stackrel{d}{=}\; \frac{Z + Z'}{2^{1/2}}, \tag{1.3}$$

where $\stackrel{d}{=}$ means equality in distribution. In terms of the chf φ, say, of Z, this equation becomes

$$\varphi(t) \;=\; \varphi\left(\frac{t}{2^{1/2}}\right)^2, \quad t \in \mathbb{R}, \tag{1.4}$$

which via iteration leads to

$$\varphi(t) \;=\; \lim_{n\to\infty} \varphi\left(\frac{t}{2^{n/2}}\right)^{2^n} \;=\; \lim_{n\to\infty} \left(1 - \frac{t^2}{2^{n+1}} + o\left(\frac{t^2}{2^{n+1}}\right)\right)^{2^n} \;=\; e^{-t^2/2},$$

for all $t \in \mathbb{R}$, when noting that φ satisfies the same Taylor expansion as ϕ given above. Hence we have proved that Q is the standard normal law.

The random recursive equation (1.2) that has worked for us here may also be written as

$$S_{2n}^* \;=\; \Psi(S_n^*, S_{n,n}^*) \quad \text{with} \quad \Psi(x,y) := \frac{x+y}{2^{1/2}}.$$

Although stated in terms of random variables, it should be noticed that its recursive property is rather in terms of distributions: The distribution of S_{2n}^* is expressed as a functional of the distribution of S_n^* (recalling that $S_{n,n}^*$ is an independent copy of this variable). Then, by taking the limit $n \to \infty$ and using the continuity of Ψ, the limiting distribution has been identified as a solution to the SFPE (1.3), viz.

$$Z \stackrel{d}{=} \Psi(Z, Z') \;=\; \frac{Z+Z'}{2^{1/2}}.$$

Under the proviso that Z (or Q) has mean 0 and unit variance, we have shown that the standard normal distribution forms the unique solution to (1.3). We note in passing that, by a simple scaling argument, $Normal(0,\sigma^2)$, the normal distribution with mean 0 and variance $\sigma^2 > 0$, is found to be the unique solution to the very same SFPE within the class of distributions with mean 0 and variance σ^2.

Generalizing in another direction, fix any $N \ge 2$ and positive integers $k_1, ..., k_N$ satisfying $k_1 + ... + k_N = N$. Define $s_1 := k_1, s_2 := k_1 + k_2, ..., s_N := k_1 + ... + k_N$ and then

$$S_n^*(j) \;:=\; \frac{1}{(k_j n)^{1/2}} \sum_{m=s_{j-1}n+1}^{s_j n} X_m$$

for $j = 1, ..., N$. The latter random variables are clearly independent with $\mathscr{L}(S_n^*(j)) = \mathscr{L}(S_{k_j n}^*)$ for each $j = 1, ..., N$. Moreover, the random recursive equation

$$S_{Nn}^* \;=\; \sum_{j=1}^{N} T_j S_n^*(j) \tag{1.5}$$

with $T_j := (k_j/N)^{1/2}$ for $j = 1, ..., N$ holds true. Hence, the distribution of S_{Nn}^* is a functional of the distributions of $S_{k_1 n}^*, ..., S_{k_N N}^*$. By another appeal to the continuous mapping theorem, we obtain upon passing to the limit $n \to \infty$ that (under the same proviso as before)

$$Z \stackrel{d}{=} \sum_{j=1}^{N} T_j Z_j \tag{1.6}$$

where $Z_1, ..., Z_N$ are independent copies of Z. Equivalently,

$$\varphi(t) \;=\; \prod_{j=1}^{N} \varphi(T_j t), \quad t \in \mathbb{R}$$

holds for the chf φ of Z, and a similar argument as before may be employed to conclude that the standard normal law forms the unique solution to (1.6) within the class of distributions with mean 0 and unit variance. We close this section with the following natural question:

> *Under which conditions on* $(N, T_1, ..., T_N)$*, the parameters of the SFPE* (1.6)*, does the previous uniqueness statement remain valid?*

The restriction imposed by our construction is that N is finite and that $T_1^2, ..., T_N^2$ are positive rationals summing to unity. The very last property is clearly necessary, for (1.6) in combination with $\mathbb{V}\mathrm{ar} Z = 1$ entails

$$1 \;=\; \mathbb{V}\mathrm{ar} Z \;=\; \sum_{j=1}^{N} \mathbb{V}\mathrm{ar}\,(T_j Z_j) \;=\; \sum_{j=1}^{N} T_j^2.$$

That no further restriction on $(N, T_1, ..., T_N)$ is needed will be shown in ??????? in a more general framework. This means that N may even be infinite and $T_1, T_2, ..., T_N$ any real numbers such that $\sum_{j=1}^{N} T_j^2 = 1$.

Problems

Problem 1.1. For any $\alpha \in (0,2]$ and $b > 0$, the function $\phi(t) = \exp(-b|t|^{\alpha})$ is the chf of a (symmetric) distribution $\mathscr{S}(\alpha, b)$ on $\mathbb{R}$, called *symmetric stable law with index* α *and scaling parameter* b. Note that $\mathscr{S}(2,b) = \mathit{Normal}(0,2b)$ and $\mathscr{S}(1,b) = \mathit{Cauchy}(b)$, the symmetric Cauchy distribution with $\lambda\!\!\lambda$-density $\frac{1}{\pi}\frac{b}{b^2+x^2}$. Prove that $\mathscr{S}(\alpha, b)$ forms a solution to the SFPE

$$X \;\stackrel{d}{=}\; \frac{X_1 + ... + X_n}{n^{1/\alpha}} \tag{1.7}$$

for any $n \geq 2$, where $X_1, ..., X_n$ are independent copies of X.

Problem 1.2. Prove the following assertions for any $b > 0$:

(a) The function $\mathbb{R}_{\geq} \ni t \mapsto \varphi(t) = \exp(-bt^{\alpha})$ is the LT of a distribution $\mathscr{S}_{+}(\alpha, b)$ on $\mathbb{R}_{\geq}$, called *one-sided stable law with index* α *and scaling parameter* b, iff $\alpha \in (0,1]$.

(b) $\mathscr{S}_{+}(\alpha, b)$ forms a nonnegative solution to the SFPE (1.7).

Problem 1.3. Let $N \in \mathbb{N} \cup \{\infty\}$ and $T_1, ..., T_N \geq 0$. Find conditions on $N, T_1, ..., T_N$ such that $\mathscr{S}(\alpha, b)$ and $\mathscr{S}_{+}(\alpha, b)$ are solutions to the SFPE (1.6).

1.2 A prominent queuing example: the Lindley equation

In a single-server queuing system, the *Lindley equation* for the waiting time of a customer before receiving service provides another well-known example of a random recursive equation. To set up the model, suppose that an initially idle server is facing (beginning at time 0) arrivals of customers at random epochs $0 \le T_0 < T_1 < T_2 < ...$ with service requests of (temporal) size $B_0, B_1, ...$ Customers who find the server busy join a queue and are served in the order they have arrived (first in, first out). Typical performance measures are quantities like workload, queue length or waiting times of customers in the system. They may be studied over time (transient analysis) or in the long run (steady state analysis). Here we will focus on the time a customer spends in the queue (if there is one) before receiving service and will do so for the so-called *G/G/1-queue* specified by the following assumptions [☞ also [6]]:

(G/G/1-1) The sequence of arrival epochs $(T_n)_{n\ge 0}$ has iid positive increments $A_1, A_2, ...$ with finite mean λ and thus forms a renewal process with finite drift.
(G/G/1-2) The service times $B_0, B_1, ...$ are iid with finite positive mean μ.
(G/G/1-3) The sequences $(T_n)_{n\ge 0}$ and $(B_n)_{n\ge 0}$ are independent.
(G/G/1-4) There is one server and a waiting room of infinite capacity.
(G/G/1-5) The queue discipline is FIFO ("first in, first out").

The *Kendall notation* "G/G/1", which may be expanded by further symbols when referring to more complex systems, has the following meaning: The first letter refers to the arrival pattern, the second one to the service pattern, and the number in the third position to the number of servers (or counters). The letter "G" stands for "general" and is sometimes replaced with "GI" for "general independent". It means that both, interarrival times and service times are each iid with a general distribution.

Let W_n denote the quantity in question, that is, the waiting time of the n^{th} arriving customer before receiving service and notice that $W_0 = 0$, for the server is supposed to be idle before T_0. In order to derive Lindley's equation for W_n $(n \ge 1)$, we point out the following: Either $W_n = 0$, which happens if the n^{th} customer arrives after his predecessor has already left the system, or W_n equals the time spent in the system by the predecessor, i.e. $W_{n-1} + B_{n-1}$, minus the time A_n that elapses between the arrival of that customer and his own arrival. The first case occurs if $T_n > T_{n-1} + W_{n-1} + B_{n-1}$ or, equivalently, $W_{n-1} + B_{n-1} - A_n < 0$, while the second one occurs if $W_{n-1} + B_{n-1} - A_n \ge 0$. Consequently, the Lindley equation [76] takes the form

$$W_n \;=\; (W_{n-1} + X_n)^+ \tag{1.8}$$

for each $n \ge 1$, where $X_n := B_{n-1} - A_n$. Put also $S_0 := 0$ and $S_n = X_1 + ... + X_n$ for $n \ge 1$. Then $(S_n)_{n\ge 0}$ forms an ordinary zero-delayed random walk with drift $\mu - \lambda$. It is now an easy exercise [☞ Problem 1.5] to prove via iteration that

$$W_n \;=\; \max_{0\le k\le n}(S_n - S_k) \;\stackrel{d}{=}\; \max_{0\le k\le n} S_k \tag{1.9}$$

for each $n \ge 0$ and then to deduce the following result about the asymptotic behavior of W_n.

Theorem 1.4. *Under the stated assumptions, the waiting time W_n converges in distribution to $W_\infty := \max_{k\ge 0} S_k$ iff $\mu < \lambda$. In this case*

$$W_\infty \overset{d}{=} (W_\infty + X)^+, \tag{1.10}$$

where X denotes a generic copy of $X_1, X_2, \ldots$ independent of W_∞. Furthermore,

$\lim_{n\to\infty} W_n = \infty$ *a.s. if* $\mu > \lambda$,

and

$0 = \liminf_{n\to\infty} W_n < \limsup_{n\to\infty} W_n = \infty$ *a.s. if* $\mu = \lambda$.

Proof. Problem 1.5 □

It is intuitively obvious (and indeed true) that asymptotic stability of waiting times, i.e. distributional convergence of the W_n, is equivalent to the asymptotic stability of the whole system in the sense that other relevant functionals like queue length or workload approach a distributional limit as well. Adopting a naive standpoint by simply ignoring random fluctuations of system behavior, we expect this to be true iff the mean service time is smaller than the mean time between two arriving customers, for then the server works faster on average than the input rate. The previous result tells us that naive thinking does indeed lead to the correct answer.

Further dwelling on the stable situation, thus assuming $\mu < \lambda$, it is natural to strive for further information on the distribution of W_∞, which in general cannot be determined explicitly [☞ Problem 1.7 for an exception]. For this purpose, the queuing background no longer matters so that we may just assume to be given a general nonnegative sequence $(W_n)_{n\ge 0}$, called *Lindley process*, of the recursive form (1.8) with iid random variables $X_1, X_2, \ldots$ with negative mean. The reader is asked in Problem 1.6 to show that then W_n always converges in distribution to $W_\infty = \max_{k\ge 0} S_k$, regardless of the distribution of W_0. This implies that the SFPE (1.10) determines the distribution G, say, of W_∞ uniquely. *Implicit renewal theory*, to be developed in Chapter 4, will enable us to determine the asymptotic behavior of the tail probabilities $\mathbb{P}(W > t)$ as $t \to \infty$ with the help of (1.10). At this point we finally note that the latter may be stated in terms of $G(t) = \mathbb{P}(W \le t)$ as

$$G(t) = \int_{(-\infty,t]} G(t-x)\, \mathbb{P}(X \in dx), \quad t \ge 0, \tag{1.11}$$

called *Lindley's integral equation*.

Problems

Problem 1.5. Given a G/G/1-queue as described above, prove that W_n satisfies (1.9) and then Theorem 1.4.

Problem 1.6. Given a sequence of iid real-valued random variables $X, X_1, X_2, \ldots$ with associated SRW $(S_n)_{n\ge 0}$, consider the Lindley process

$$W_n \;=\; (W_{n-1}+X_n)^+, \quad n\ge 1$$

with arbitrary initial value $W_0 \ge 0$ independent of $X_1, X_2, \ldots$ Prove the following assertions:

(a) For each $n\ge 1$, $W_n \stackrel{d}{=} M_{n-1}\vee(W_0+S_n)$, where $M_n = \max_{0\le k\le n} S_k$ for $n\ge 0$.

(b) If $\mathbb{E}X<0$, then $W_n \stackrel{d}{\to} W_\infty = \max_{k\ge 0} S_k$.

(c) If $\mathbb{E}X<0$, then $\mathscr{L}(W_\infty)$ forms the unique solution to the SFPE (1.10) in the class of distributions on $\mathbb{R}_\ge$.

Problem 1.7. In the previous problem, suppose that X is integer-valued with negative mean and $\mathscr{L}(X^+) = Bern(p)$ for some $p>0$. Prove that W_∞ has a geometric distribution. [Hint: Consider the *strictly descending ladder epochs* $(\sigma_n^<)_{n\ge 0}$, recursively defined by $\sigma_0^< := 0$ and

$$\sigma_n^< \;:=\; \inf\left\{k>\sigma_{n-1}^< : S_k < S_{\sigma_{n-1}^<}\right\}$$

for $n\ge 1$, where $\inf\emptyset := \infty$. Then write W_∞ in terms of the associated *ladder heights* $S_{\sigma_n^<}\mathbf{1}_{\{\sigma_n^<<\infty\}}$ and use that, given $\sigma_n^< < \infty$, the random vectors

$$\left(\sigma_k^< - \sigma_{k-1}^<, S_{\sigma_k^<} - S_{\sigma_{k-1}^<}\right), \quad k=1,\ldots,n,$$

are conditionally iid [☞ Subsec. 2.2.1 in [2] for further information].]

Problem 1.8. Here is a version of the *continuous mapping theorem* that will frequently be used hereafter:
Let $\theta_1, \theta_2, \ldots$ be iid $\mathbb{R}^d$-valued $(d\ge 1)$ random variables with generic copy θ and independent of X_0. Suppose further that $X_n = \psi(X_{n-1},\theta_n)$ for all $n\ge 1$ and a continuous function $\psi:\mathbb{R}^{d+1}\to\mathbb{R}$. Prove that, if X_n converges in distribution to X_∞, then

$$\psi(X_{n-1},\theta_n) \stackrel{d}{\to} \psi(X_\infty,\theta) \quad\text{and}\quad X_\infty \stackrel{d}{=} \psi(X_\infty,\theta),$$

where X_∞ and θ are independent.

1.3 A rich pool of examples: branching processes

Consider a population starting from one ancestor (generation 0) in which individuals of the same generation produce offspring independently and also independent of the current generation size. The offspring distribution, denoted as $(p_n)_{n\ge 0}$, is supposed to be the same for all population members and to have finite mean m. Under these assumptions, the generation size process $(Z_n)_{n\ge 0}$, thus $Z_0 = 0$, forms a so-called *(simple) Galton-Watson (branching) process (GWP)* and satisfies the random recursive equation

$$Z_n \;=\; \sum_{k=1}^{Z_{n-1}} X_{n,k}, \quad n \ge 1, \tag{1.12}$$

where $\{X_{n,k} : n,k \ge 1\}$ forms a family of iid integer-valued random variables with common distribution $(p_n)_{n\ge 0}$. Here Z_n denotes the size of the n^{th} generation and $X_{n,k}$ the number off children of the k^{th} member of this generation (under an arbitrary labeling of these members). To exclude the trivial case $Z_0 = Z_1 = ... = 1$, we make the standing assumption $p_1 < 1$.

A classical result, known as the *extinction-explosion principle*, states that the population either dies out ($Z_n = 0$ eventually) or explodes ($Z_n \to \infty$), i.e.

$$\mathbb{P}(\{Z_n = 0 \text{ eventually}\} \cup \{Z_n \to \infty\}) = 1.$$

Moreover, extinction occurs almost surely if $m < 1$ (subcritical case) or $m = 1$ (critical case), while $q := \mathbb{P}(Z_n = 0 \text{ eventually}) < 1$ if $m > 1$ (supercritical case). Defining the offspring gf $f(s) := \sum_{n\ge 0} p_n s^n$ for $s \in [0,1]$, q equals the minimal fixed point of f in $[0,1]$.

It is easily verified that the normalized sequence $W_n = m^{-n} Z_n$, $n \ge 0$, constitutes a nonnegative mean one martingale which therefore converges to a nonnegative limit W with $\mathbb{E}W \le 1$ by the martingale convergence theorem [☞ Problem 1.9]. If $m \le 1$, then clearly $W = 0$ a.s. holds true, but if $m > 1$ we may hope for $W > 0$ a.s. on the survival event $\{Z_n \to \infty\}$ giving that Z_n grows like a random constant times m^n on that event as $n \to \infty$. A famous result by KESTEN & STIGUM [70] states that this holds true iff

$$\mathbb{E}Z_1 \log Z_1 \;=\; \sum_{n\ge 1} p_n\, n \log n \;<\; \infty \tag{ZlogZ}$$

which we will assume hereafter. Then $(W_n)_{n\ge 0}$ is ui and thus $\mathbb{E}|W_n - W| \to 0$, in particular $\mathbb{E}W = \mathbb{E}W_0 = 1$.

What can be said about the distribution of W? The following argument shows that once again its distribution satisfies a SFPE. First notice that, besides (1.12), we further have

$$Z_n \;=\; \sum_{j=1}^{Z_1} Z_{n-1}(j), \quad n \ge 1, \tag{1.13}$$

where $(Z_n(j))_{n\ge 0}$ denotes the generation size process of the subpopulation stemming from the j^{th} individual in the first generation of the whole population. In fact, we can define $(Z_n(j))_{n\ge 0}$ for any $j \ge 1$ in such a way that these processes are independent copies of $(Z_n)_{n\ge 0}$ and also independent of Z_1. Then, defining $W_n(j)$ in an obvious manner, we infer

$$W_n \;=\; \frac{1}{m} \sum_{j=1}^{Z_1} W_{n-1}(j), \quad n \ge 1$$

and then, by letting $n \to \infty$, that

$$W \;=\; \frac{1}{m}\sum_{j=1}^{Z_1} W(j) \quad \text{a.s.}, \tag{1.14}$$

where $W(j)$ denotes the almost sure limit of the martingale $(W_n(j))_{n\geq 0}$. By what has been pointed out before, the $W(j)$ are independent copies of W and independent of Z_1 so that (1.14) does indeed constitute an SFPE for $\mathscr{L}(W)$. In terms of the LT $\varphi(t) := \mathbb{E}e^{-tW}$ of W, it may be restated as

$$\varphi(t) \;=\; f\circ\varphi\left(\frac{t}{m}\right), \quad t\geq 0, \tag{1.15}$$

as one can readily verify. With the help of this equation, one can further show (under (ZlogZ)) that φ is the unique solution with right derivative $\varphi'(0+) = -\mathbb{E}W = -1$ at 0 [☞ Problem 1.10]. Since distributions are determined by their LT's we hence conclude that $\mathscr{L}(W)$ is the unique solution to (1.14).

There are many other functionals in connection with GWP's that can be described by a random recursive equation. Here we confine ourselves to two further examples in the case when $m \leq 1$ in which almost certain extinction occurs. First, consider the *total population size process*

$$Y_n \;:=\; \sum_{k=0}^{n} Z_k, \quad n\geq 0$$

which satisfies

$$Y_n \;=\; 1+\sum_{j=1}^{Z_1} Y_{n-1}(j), \quad n\geq 1, \tag{1.16}$$

where $(Y_n(j))_{n\geq 0}$ denotes the total population size process associated with the GWP $(Z_n(j))_{n\geq 0}$ defined above. Plainly, Y_n increases to an a.s. finite limit Y_∞ which, by (1.16), satisfies the SFPE

$$Y_\infty \;=\; 1+\sum_{j=1}^{Z_1} Y_\infty(j). \tag{1.17}$$

Problem 1.11 shows that this equation characterizes the distribution of Y_∞ uniquely. It is not obvious at all but has been shown by DWASS [40] that $\mathscr{L}(Y_\infty)$ can be obtained explicitly, namely

$$\mathbb{P}(Y_\infty = j) \;=\; \frac{1}{j}\, p_{j,j-1}, \quad j\geq 1,$$

where $p_{ij} := \mathbb{P}(Z_1 = j|Z_0 = i)$ for $i, j \geq 0$. The proof is based on a clever analysis of the random recursive equation (1.16) in terms of gf's.

As a second example, still assuming $m \leq 1$, we mention the *extinction time* of $(Z_n)_{n\geq 0}$, viz.

$$T \;:=\; \inf\{n\geq 1 : Z_n = 0\}.$$

If $T(j)$ denotes the corresponding random variable for the GWP $(Z_n(j))_{n\geq 0}$ for each $j \geq 1$, then the following SFPE follows immediately:

$$T \;=\; 1+\bigvee_{j=1}^{Z_1} T(j) \tag{1.18}$$

with the convention that $\bigvee_{j=1}^{0} x_j := 0$.

Problems

Problem 1.9. Given a GWP $(Z_n)_{n\geq 0}$ with one ancestor and finite offspring mean m, prove that $W_n = m^{-n} Z_n$ for $n \geq 0$ forms a nonnegative martingale.

Problem 1.10. Prove (1.15) and then, assuming (ZlogZ), that φ is the unique solution with right derivative at 0 satisfying $|\varphi'(0+)| = 1$.

Problem 1.11. Suppose $m \leq 1$ and let φ denote the LT of the final total population size Y_∞. Prove that φ satisfies the functional equation $\varphi(t) = e^{-t} f \circ \varphi(t)$ equivalent to (1.17) and that it forms the unique solution in the class of LT's of distributions. [Hint: Use the convexity of f.]

1.4 The sorting algorithm `Quicksort`

`Quicksort`, first introduced by HOARE [65, 66], is probably the nowadays most commonly used, so called *divide-and-conquer algorithm* to sort a list of n real numbers and serves as the standard sorting algorithm in UNIX-systems. Based on the general idea to successively divide a given task into subtasks of the same kind but smaller dimension, it forms a random recursive algorithm that may be briefly described as follows: Given n distinct reals $a_1,...,a_n$, which are to be sorted in increasing order, the first step is to create two sublists by first choosing an element x from the list, called *pivot*, and then to put all a_k smaller than x in the first sublist and all a_k bigger than x in the second sublist. The same procedure is then applied to the two sublists and all further on created ones as long as these contain at least two elements. Hence, the algorithm terminates when all sublists consist of only one element which are then merged to yield $a_1,...,a_n$ in increasing order. The way the pivots are chosen throughout the performance of the algorithm may be deterministic or at random, e.g. by picking any element of a given sublist with equal probability. Notice that the particular values of $a_1,...,a_n$ do not matter for the algorithm so that we may assume w.l.o.g. that $(a_1,...a_n)$ is a permutation of the numbers $1,...n$. When picking such a permutation at random, the number of key comparisons needed by `Quicksort` to output the ordered sample becomes a random variable X_n, and our goal hereafter is to study the distribution of X_n. But before proceeding we give an example first.

Example 1.12. In order to illustrate how `Quicksort` may perform on a given sample, we have depicted a permutation of the numbers $1,2,\ldots,12$. The table below shows that the algorithm needs four rounds to output the ordered sample. Each round consists of the further subdivisions of the currently given sublists with more than one element with respect to previously chosen pivots (shown in boldface). The final column of the table displays how many key comparisons are needed to complete the round.

List to be sorted	6 3 9 2 5 12 8 1 10 4 11 7	# key comparisons
Round 1	3 2 5 1 4 **6** 9 12 8 10 11 7	11
Round 2	2 1 **3** 5 4 • 8 7 **9** 12 10 11	9
Round 3	1 **2** • 4 **5** • 7 **8** • 10 11 **12**	5
Round 4	1 • • 4 • • 7 • • **10** 11 •	1

The reader may wonder about the necessity of Round 4 because the sample appears to be in correct order already after Round 3. The simple explanation is that after Round 3 we still have one sublist of length ≥ 2, namely $(10,11)$ which in the final round is assessed to be in correct order by choosing 10 as the pivot and making the one necessary comparison with the other element 11 [☞ also Figure 1.1 below].

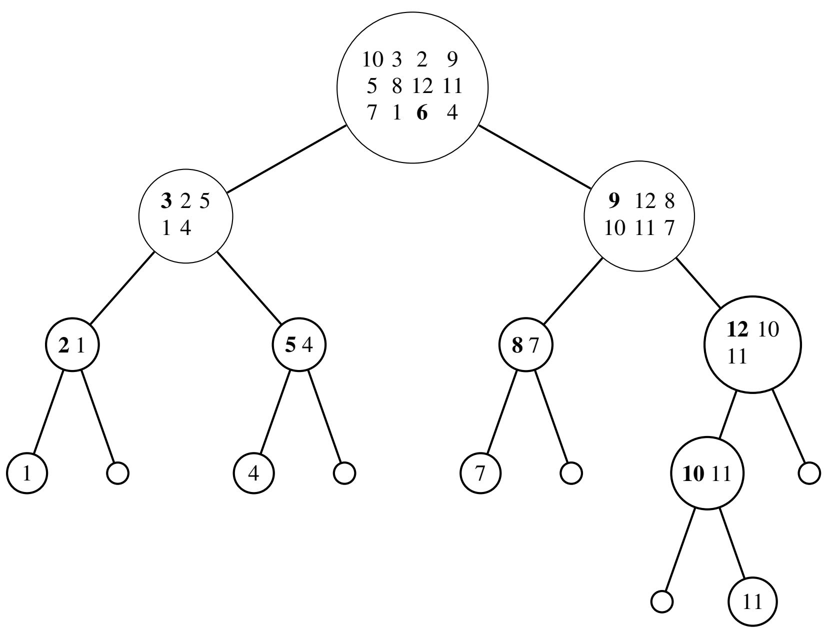


Fig. 1.1 Example 1.12: Left and right nodes of the tree are representing the respective sublists as created in the successive rounds by comparison with the pivot (shown in boldface) in the previous node.

As already announced, our performance analysis of `Quicksort` will be based on the number of key comparisons X_n needed to sort a random permutation of length

n. It seems plausible that this number is essentially proportional to the performance time and therefore the appropriate quantity to analyze.

To provide a rigorous model for X_n, let

$$\Omega_n := \{\pi \in \{1,...,n\}^n : \pi_i \neq \pi_j \text{ for } i \neq j\}$$

be the permutation group of $1,...,n$, here the set of possible inputs, and $\mathbb{P}_n$ the (discrete) uniform distribution on Ω_n. The discrete random variable $X_n : \Omega_n \to \mathbb{N}_0$ then maps any π on the number of key comparisons needed by `Quicksort` to sort π in increasing order where, for simplicity, we assume that pivots are always chosen as first elements in the appearing sublists[1]. Consequently, $Z_n(\pi) := \pi_1$ denotes the pivot in the input list and has a uniform distribution on $\{1,...,n\}$. It also gives the rank of this element in the list. The derivation of results about the distribution of X_n will be heavily based on the recursive structure of `Quicksort` which we are now going to make formally more explicit. Denote by L_n, R_n the rank tuples of the left and right sublist, respectively, created in the first round via key comparison with Z_n. Observe that these lists have lengths $Z_n - 1$ and $n - Z_n$, respectively, so that $L_n(\pi) \in \Omega_{Z_n(\pi)-1}$ and $R_n(\pi) \in \Omega_{n-Z_n(\pi)}$ for any $\pi \in \Omega_n$. After these settings the crucial random recursive equation for X_n takes the form

$$X_n = X_{Z_n-1} \circ L_n + X_{n-Z_n} \circ R_n + n - 1 \tag{1.19}$$

for any $n \geq 1$, where $X_0(\emptyset) := 0$. It follows by a combinatorial argument that, given $Z_n = i$, L_n and R_n are conditionally independent and uniformly distributed on Ω_{i-1} and Ω_{n-i}, respectively [☞ Problem 1.14]. Setting $\mathbb{P}_0(X_0 \in \cdot) := \delta_0$, it hence follows that

$$\begin{aligned}
\mathbb{P}_n(X_n \in \cdot) &= \sum_{i=1}^{n} \mathbb{P}_n(Z_n = i)\, \mathbb{P}_n(X_n \in \cdot | Z_n = i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_n(X_{Z_n-1} \circ L_n + X_{n-Z_n} \circ R_n + n - 1 \in \cdot | Z_n = i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_n(X_{Z_n-1} \circ L_n \in \cdot | Z_n = i) * \mathbb{P}_n(X_{n-Z_n} \circ R_n \in \cdot | Z_n = i) * \delta_{n-1} \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_{i-1}(X_{i-1} \in \cdot) * \mathbb{P}_{n-i}(X_{n-i} \in \cdot) * \delta_{n-1}
\end{aligned}$$

for each $n \geq 1$. From now on, we assume that *all* X_n, Z_n, $n \geq 1$, are defined on just one sufficiently large probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ which further carries independent random variables $X_0', X_0'', X_1', X_1'', ...$, which are also independent of $(X_n, Z_n)_{n\geq 1}$, such that

$$X_0' = X_0'' := 0 \quad \text{and} \quad X_n \stackrel{d}{=} X_n' \stackrel{d}{=} X_n'' \quad \text{for } n \geq 1.$$

Then equation (1.19) provides us with the distributional relation

---

[1] this version is sometimes referred to as *vanilla* `Quicksort`

$$X_n \stackrel{d}{=} X'_{Z_n-1} + X''_{n-Z_n} + n-1 \tag{1.20}$$

for all $n \geq 1$.

Our ultimate goal to be accomplished later [☞ Section 5.6] is to show that a suitable normalization of X_n converges in distribution to some X_∞ and to characterize $\mathscr{L}(X_\infty)$ as the solution to a certain SFPE. At this point, we must contend ourselves with some preliminary considerations towards this result due to RÖSLER [96] including a heuristic derivation of the SFPE.

In order to gain an idea about a suitable normalization of X_n, we first take a look at its expectation. Let $H_n := \sum_{k=1}^n \frac{1}{k}$ be the n^{th} harmonic sum and

$$\gamma := \lim_{n\to\infty} (H_n - \log n) = 0.5772...$$

denote Euler's constant.

Lemma 1.13. *For each $n \geq 1$,*

$$\mathbb{E}X_n = 2(n+1)H_n - 4n \tag{1.21}$$

holds true and, furthermore,

$$\mathbb{E}X_n = 2(n+1)\log n + (2\gamma - 4)n + 2\gamma + 1 + O\left(\frac{1}{n}\right) \tag{1.22}$$

as $n \to \infty$.

Proof. Taking expectations in (1.20), we obtain

$$\begin{aligned}
\mathbb{E}X_n &= n-1+\sum_{j=1}^n \mathbb{P}(Z_n=j)\left(\mathbb{E}X_{j-1}+\mathbb{E}X_{n-j}\right) \\
&= n-1+\frac{1}{n}\sum_{j=1}^n \left(\mathbb{E}X_{j-1}+\mathbb{E}X_{n-j}\right) \\
&= n-1+\frac{2}{n}\sum_{j=1}^{n-1} \mathbb{E}X_j
\end{aligned}$$

and then upon division by $n+1$ and a straightforward calculation that

$$\frac{\mathbb{E}X_n}{n+1} = \frac{\mathbb{E}X_{n-1}}{n} + \frac{2(n-1)}{n(n+1)} \tag{1.23}$$

for all $n \geq 1$. We leave it to the reader [☞ Problem 1.15] to verify this recursion and to derive (1.21) from it. The asymptotic expansion (1.22) then follows directly when using that $H_n = \log n + \gamma + (2n)^{-1} + O(n^{-2})$ as $n \to \infty$. □

The reader is asked to show in Problem 1.16 that $\mathbb{V}\mathrm{ar}\, X_n \simeq \sigma^2 n^2$ as $n \to \infty$, where $\sigma^2 := 7 - \frac{2}{3}\pi^2$. In view of this fact it is now reasonable to study the asymptotic behavior of the normalization

$$\widehat{X}_n := \frac{X_n - \mathbb{E}X_n}{n}.$$

The contraction argument due to RÖSLER [96] that proves convergence in distribution of $\widehat{X}_n$ to a limit $\widehat{X}_\infty$ with mean 0 and variance σ^2 will be postponed to Section 5.6. Here we outline the argument that shows that $\mathscr{L}(X_\infty)$, called `Quicksort` *distribution*, may again be characterized by an SFPE.

The argument embarks on the distributional equation (1.20), which after normalization becomes

$$\widehat{X}_n \stackrel{d}{=} \frac{Z_n - 1}{n} \widehat{X}'_{Z_n - 1} + \frac{n - Z_n}{n} \widehat{X}''_{n - Z_n} + g_n(Z_n) \tag{1.24}$$

for $n \geq 2$, where $\widehat{X}_0 = \widehat{X}_1 := 0$ and $g_n : \{1, ..., n\} \to \mathbb{R}$ is defined by

$$g_n(k) := \frac{n-1}{n} + \frac{1}{n} \left(\mathbb{E}X_{k-1} + \mathbb{E}X_{n-k} - \mathbb{E}X_n\right). \tag{1.25}$$

Notice that the $\widehat{X}'_n, \widehat{X}''_n$, $n \geq 0$, continue to be independent of $(X_n, Z_n)_{n \geq 1}$. The reader can easily verify [☞ Problem 1.17] that $Z_n/n \stackrel{d}{\to} \mathit{Unif}(0,1)$, and we will prove in 5.6 that $0 \leq \sup_{n \geq 1, 1 \leq k \leq n} g_n(k) < \infty$ as well as

$$\lim_{n \to \infty} g_n(\lceil nt \rceil) = g(t) := 1 + 2t \log t + 2(1-t) \log(1-t)$$

for all $t \in (0,1)$ uniformly on compact subsets, where $\lceil x \rceil := \inf\{n \in \mathbb{Z} : x \leq n\}$. By combining these facts and $\widehat{X}_n \stackrel{d}{\to} \widehat{X}_\infty$, it can be deduced from (1.24) that $\mathscr{L}(\widehat{X}_\infty)$ solves the SFPE

$$\widehat{X}_\infty \stackrel{d}{=} U\widehat{X}'_\infty + (1-U)\widehat{X}''_\infty + g(U) \tag{1.26}$$

where $\widehat{X}'_\infty, \widehat{X}''_\infty$ and U are independent with $\widehat{X}'_\infty \stackrel{d}{=} \widehat{X}''_\infty \stackrel{d}{=} \widehat{X}_\infty$ and $U \stackrel{d}{=} \mathit{Unif}(0,1)$. This is the `Quicksort` equation, and we will also show in 5.6 that $\mathscr{L}(\widehat{X}_\infty)$ forms its unique solution within the class of zero mean distributions with finite variance.

Binary search trees. A binary search tree (BST) of size n is a labeled binary tree with n internal nodes generated from a permutation $\pi = (\pi_1, ..., \pi_n) \in \Omega_n$. One way to construct it is as follows: Start with π_1, store it in the root of the tree and attach to it two empty nodes, called external. Then π_2 is compared with π_1 and becomes the left descendant if $\pi_2 < \pi_1$, and the right descendant otherwise. Attach two empty nodes to the now internal node occupied by π_2. Proceed with any π_k in the same manner by moving it along internal nodes until an external one is reached where it is stored. At each internal node with value x, say, where $x \in \{\pi_1, ..., \pi_{k-1}\}$, move left if $\pi_k < x$ and right otherwise. Finish step k by attaching two external nodes to the node now occupied by π_k. After n steps all keys have been stored, giving a binary tree with n internal and $n+1$ external nodes. This is exemplified in Fig. 1.2 with the permutation from Example 1.12. As one can see, the same tree as in Fig. 1.1 is obtained when ignoring external nodes. In fact, an application of Quicksort *always* leads to the same result as the procedure just described when only storing the leading element of each sublist (the pivot) in the nodes.

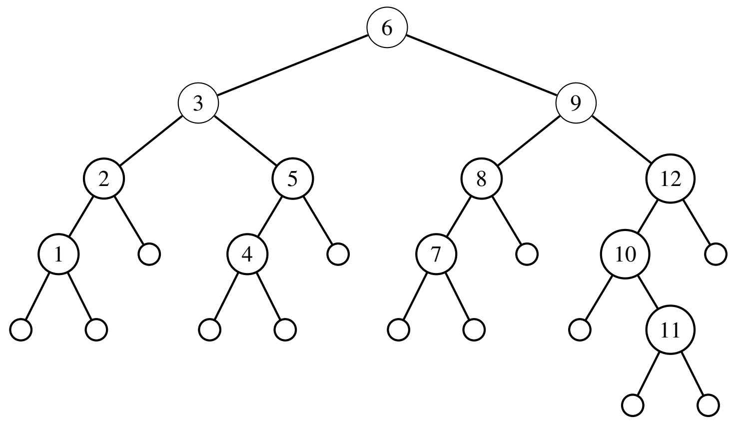


Fig. 1.2 The permutation $(6,3,9,2,5,12,8,1,10,4,11,7)$ from Example 1.12 stored in a binary search tree. External nodes are shown as empty circles.

Problems

Problem 1.14. Prove that, given $Z_n = i$, the rank tuples L_n and R_n are conditionally independent with a discrete uniform distribution on Ω_{i-1} and Ω_{n-i}, respectively.

Problem 1.15. Complete the proof of Lemma 1.13 by verifying (1.23) and then deriving (1.21) from it.

Problem 1.16. Prove that $\sigma_n^2 := \mathbb{V}\mathrm{ar} X_n$ satisfies

$$\sigma^2 \;:=\; \lim_{n\to\infty} \frac{\sigma_n^2}{n^2} \;=\; 7 - \frac{2}{3}\pi^2 \;=\; 0.4203... \tag{1.27}$$

by doing the following parts:

(a) Use (1.20) to show that

$$\sigma_n^2 \;=\; c_n - (n-1)^2 + \frac{2}{n}\sum_{k=1}^{n-1} \sigma_k^2 \tag{1.28}$$

for all $n \geq 1$, where $\mu_n := \mathbb{E}X_n$ and $c_n := \frac{1}{n}\sum_{k=1}^{n}(\mu_{k-1} + \mu_{n-k} - \mu_n)^2$.

(b) Use (1.28) to derive the recursion

$$\frac{d_n}{n+1} \;=\; \frac{d_{n-1}}{n} \;+\; \frac{2\left(c_{n-1}-(n-2)^2\right)}{n(n+1)} \tag{1.29}$$

for any $n \ge 2$, where $d_n := \sigma_n^2 + (n-1)^2 - c_n$ for $n \ge 1$. Note that $\sigma_1^2 = c_1 = d_1 = 0$.

(c) Use Lemma 1.13 to show that

$$\frac{c_n}{n^2} \;=\; \frac{4}{n}\sum_{k=1}^{n}\left(\frac{k}{n}\log\left(\frac{k}{n}\right)+\left(1-\frac{k}{n}\right)\log\left(1-\frac{k}{n}\right)\right)^2 \;+\; o(1)$$

as $n \to \infty$ and thereby

$$\lim_{n\to\infty}\frac{c_n}{n^2} \;=\; c \;:=\; \frac{10}{3}-\frac{2}{9}\pi^2. \tag{1.30}$$

(d) Finally, combine the previous parts to infer

$$\lim_{n\to\infty}\frac{\sigma_n^2}{n^2} \;=\; 3(c-1)$$

which is easily seen to be the same as (1.27).

(e) Those readers who want to work harder should prove the stronger assertion [stated by FILL & JANSON in [54]]

$$\frac{\sigma_n^2}{n^2} \;=\; 3(c-1)-\frac{2\log n}{n}+O\left(\frac{1}{n}\right)$$

as $n \to \infty$.

Problem 1.17. Prove that $Z_n/n \xrightarrow{d} \mathit{Unif}(0,1)$ if Z_n has a discrete uniform distribution on $\{1,\ldots,n\}$ for each $n \ge 1$.

Problem 1.18. Let D_n denote the *depth* or *height* of a random BST with n internal nodes, thus $D_0 = D_1 = 0$. Prove that

$$D_n \;\stackrel{d}{=}\; 1 + D''_{Z_n-1} \vee D''_{n-Z_n}$$

for each $n \ge 1$, where D'_k, D''_k, $k \ge 0$, and Z_n are independent random variables such that $\mathscr{L}(D_k) = \mathscr{L}(D'_k) = \mathscr{L}(D''_k)$ for each k and $\mathscr{L}(Z_n) = \mathit{Unif}(\{1,\ldots,n\})$.

1.5 Random difference equations and perpetuities

A *random difference equation (RDE)* is probably the simplest nontrivial example of a random recursive equation, the recursion being defined by a random affine linear function $\Psi(x) = Mx + Q$ for a pair (M,Q) of real-valued random variables. More precisely, let $(M_n,Q_n)_{n\geq 1}$ be a sequence of independent copies of (M,Q), X_0 a further random variable independent of this sequence, and define the sequence $(X_n)_{n\geq 0}$ recursively by

$$X_n \;:=\; M_n X_{n-1} + Q_n, \quad n \geq 1. \tag{1.31}$$

This is the general form of a (one-dimensional) RDE and has been used in many applications to model a quantity that is subject to an intrinsic random increase or decay, given by M_n for the time interval $(n-1,n]$, and an external random in- or output of size Q_n right before time n for any $n \geq 1$. Here are some special cases:

- The choice $M \equiv 1$ leads to the classical random walk (RW)

  $$X_n \;=\; X_0 + \sum_{k=1}^{n} Q_k, \quad n \geq 0$$

  with initial value (delay) X_0, which constitutes one of the most fundamental type of a random sequence.
- If $Q \equiv 0$, then we obtain its multiplicative counterpart

  $$X_n \;=\; X_0 \prod_{k=1}^{n} M_k, \quad n \geq 0,$$

  called *multiplicative RW*.
- If $M \equiv \alpha$ for some constant $\alpha \neq 0$, then

  $$X_n \;=\; \alpha X_{n-1} + Q_n \;= \ldots =\; \alpha^n X_0 + \sum_{k=1}^{n} \alpha^{n-k} Q_k, \quad n \geq 1$$

  is a so-called *autoregressive process of order 1*, usually abbreviated as *AR(1)*, and one of the simplest examples of a *linear times series*.
- As a particular case of an AR(1)-process consider the situation where $\alpha \in (0,1)$, $X_0 = 0$ and $Q_n = \alpha\xi_n$ with $\mathscr{L}(\xi_n) = Bern(p)$ for some $p \in (0,1)$. Then we have

  $$X_n \;=\; \sum_{k=1}^{n} \alpha^k \xi_{n+1-k} \;\stackrel{d}{=}\; \sum_{k=1}^{n} \alpha^k \xi_k \;=:\; \widehat{X}_n$$

  for each $n \geq 0$, and since $\widehat{X}_n$ increases a.s. to the limit

  $$\widehat{X}_\infty \;:=\; \sum_{n\geq 1} \alpha^n \xi_n,$$

  we infer that $X_n \stackrel{d}{\to} \widehat{X}_\infty$. The law of $\widehat{X}_\infty$ is called a *Bernoulli convolution* and has received considerable interest with regard to the question when it is nonsingular with respect to Lebesgue measure. The interested reader may consult the survey by PERES, SCHLAG & SOLOMYAK [94] and the references given there.

Returning to the general situation, we first note that an iteration of (1.31) leads to

$$\begin{aligned} X_n &= M_n X_{n-1} + Q_n \\ &= M_n M_{n-1} X_{n-2} + M_n Q_{n-1} + Q_n \\ &= M_n M_{n-1} M_{n-2} X_{n-3} + M_n M_{n-1} Q_{n-2} + M_n Q_{n-1} + Q_n \\ &\;\vdots \\ &= M_n M_{n-1} \cdot \ldots \cdot M_1 X_0 + \sum_{k=1}^{n} M_n \cdot \ldots \cdot M_{k+1} Q_k \end{aligned}$$

for each $n \geq 1$. Now use the independence assumptions and replace $(M_k, Q_k)_{1 \leq k \leq n}$ with the copy $(M_{n+1-k}, Q_{n+1-k})_{1 \leq k \leq n}$ to see that

$$X_n \stackrel{d}{=} \Pi_n X_0 + \sum_{k=1}^{n} \Pi_{k-1} Q_k \tag{1.32}$$

for any $n \geq 1$, where $(\Pi_n)_{n \geq 0}$ is the multiplicative RW associated with $(M_n)_{n \geq 1}$ and starting at $\Pi_0 = 1$.

We are interested in finding conditions that ensure the convergence in distribution of X_n, but confine ourselves at this point to some basic observations. By an appeal to the continuous mapping theorem [as stated in Problem 1.8], we infer from (1.31) that $X_n \stackrel{d}{\to} X_\infty$ implies the SFPE

$$X_\infty \stackrel{d}{=} M X_\infty + Q, \tag{1.33}$$

naturally the independence of (M, Q) and X_∞. Furthermore, by (1.32), it entails that $X_\infty \stackrel{d}{=} \widehat{X}_\infty$, where

$$\widehat{X}_\infty := \lim_{n \to \infty} \left(\Pi_n X_0 + \sum_{k=1}^{n} \Pi_{k-1} Q_k \right)$$

exists in the sense of distributional convergence.

It is natural to ask whether $\mathscr{L}(X_\infty)$ depends on the initial value X_0. Consider the bivariate RDE

$$(X_n, X_n') = (M_n X_{n-1} + Q_n, M_n X_{n-1}' + Q_n), \quad n \geq 1$$

with two distinct initial values X_0 and X_0'. Then

$$X_n - X_n' = M_n (X_{n-1} - X_{n-1}') = \ldots = \Pi_n (X_0 - X_0')$$

for each $n \geq 1$. Consequently, sufficient for $\mathscr{L}(X_\infty)$ to be independent of X_0 is that

$$\lim_{n \to \infty} \Pi_n = 0 \quad \text{a.s.} \tag{1.34}$$

$$\text{and}\quad \widehat{X}_\infty \;=\; \sum_{k\ge 1} \Pi_{k-1}Q_k \quad \text{exists a.s. in } \mathbb{R}. \tag{1.35}$$

The infinite series $\sum_{k\ge 1}\Pi_{k-1}Q_k$ is called *perpetuity* which is an actuarial notion for the present value of a infinite payment stream, here $Q_1,Q_2,...$, at times 1,2,... discounted by the random products $\Pi_1,\Pi_2,...$ The reader is asked in Problem 1.20 to show that (1.34) and (1.35) are valid if $\mathbb{E}\log|M|<0$, $\mathbb{E}|M|^\theta<\infty$ and $\mathbb{E}|Q|^\theta<\infty$ for some $\theta>0$. On the other hand, these conditions are far from being necessary. RDE's and perpetuities will be further discussed in Subsections 3.1.4 and 4.4.1.

Problems

Problem 1.19. Suppose that $M\ge 0$ a.s. and that $\mathbb{E}\log M$ exists, i.e. $\mathbb{E}\log^+ M<\infty$ or $\mathbb{E}\log^- M<\infty$. Prove that exactly one of the following cases occurs for the multiplicative RW $(\Pi_n)_{n\ge 0}$ and characterize them in terms of M.

$$\begin{aligned}
\Pi_n \;&=\; 1 \quad \text{a.s. for all } n\ge 0\\
\lim_{n\to\infty}\Pi_n \;&=\; \infty \quad \text{a.s.}\\
\lim_{n\to\infty}\Pi_n \;&=\; 0 \quad \text{a.s.}\\
0 \;=\; \liminf_{n\to\infty}\Pi_n \;<\; \limsup_{n\to\infty}\Pi_n \;&=\; \infty \quad \text{a.s.}
\end{aligned}$$

Problem 1.20. Assuming $\mathbb{E}\log|M|<0$, $\mathbb{E}|M|^\theta<\infty$ and $\mathbb{E}|Q|^\theta<\infty$ for some $\theta>0$, prove the following assertions:

(a) There exists $\kappa\in(0,\theta]$ such that $\mathbb{E}|M|^\kappa<1$. [Hint: Consider the function $s\mapsto\mathbb{E}|M|^s$ for $s\in[0,\theta]$.]

(b) $|\Pi_n|\to 0$ a.s. and

$$\left|\sum_{k\ge 1}\Pi_{k-1}Q_k\right| \;\le\; \sum_{k\ge 1}|\Pi_{k-1}Q_k| \;<\; \infty \quad \text{a.s.}$$

(c) The last assertion remains valid if $\mathbb{E}\log^+|Q|<\infty$ [use a Borel-Cantelli argument].

Problem 1.21. Given an RDE $X_n=M_nX_{n-1}+Q_n$ for $n\ge 1$, prove that, if X_n converges in distribution and $\mathbb{P}(Q=0)<1$, then $\mathbb{P}(M=0)=0$.

Problem 1.22. Assuming M and Q to be constants, find all solutions to the SFPE (1.33), i.e. $X\overset{d}{=}MX+Q$.

1.6 A nonlinear time series model

Motivated by its relevance for the modeling of financial data, BORKOVEC & KLÜPPELBERG [21] studied the limit distribution of the following nonlinear time series model, designed to allow conditional variances to depend on past information (*conditional heteroscedasticity*) and reflecting the observations of early empirical work by MANDELBROT [84] and FAMA [49] which had shown that “that large changes in equity returns and exchange rates, with high sampling frequency, tend to be followed by large changes settling down after some time to a more normal behavior” [☞ [21, p. 1220]]. This leads to models of the form

$$X_n \;=\; \sigma_n \varepsilon_n, \quad n \ge 1, \tag{1.36}$$

where the ε_n, called *innovations*, are iid symmetric random variables and the σ_n, called *volatilities*, describe the change of the (conditional) variance. If σ_n^2 is a linear function of the p prior squared observations, viz.

$$\sigma_n^2 \;=\; \beta + \sum_{k=1}^{p} \lambda_k X_{n-k}^2, \quad n \ge 1, \tag{1.37}$$

where $\beta, \lambda_p > 0$ and $\lambda_1, ..., \lambda_{p-1} \ge 0$, we are given an *autoregressive conditionally heteroscedastic (ARCH) model of order p* as introduced by ENGLE [45]. Here we focus on the simplest case $p = 1$ and note that a combination of (1.36) and (1.37) then leads to the random recursive equation

$$X_n \;=\; \left(\beta + \lambda X_{n-1}^2\right)^{1/2} \varepsilon_n, \quad n \ge 1, \tag{1.38}$$

for some $\beta, \lambda > 0$, naturally assuming that X_0 and $(\varepsilon_n)_{n\ge1}$ are independent. It may further be extended by adding an autoregressive term, viz.

$$X_n \;=\; \alpha X_{n-1} + \left(\beta + \lambda X_{n-1}^2\right)^{1/2} \varepsilon_n, \quad n \ge 1, \tag{1.39}$$

with $\alpha \in \mathbb{R}$, to give an *AR(1)-model with ARCH(1) errors*. This is the model actually studied in [21] and belongs to a larger class of autoregressive models with ARCH errors introduced by WEISS [113].

If X_n converges in distribution to a random variable X_∞, the latter may obviously again be described by an SFPE, namely

$$X_\infty \overset{d}{=} \alpha X_\infty + \left(\beta + \lambda X_\infty^2\right)^{1/2} \varepsilon \tag{1.40}$$

where ε is a copy of the ε_n and independent of X_∞. The intersting questions are, for which parameter triples (α, β, λ) convergence in distribution actually occurs, whether in this case the SFPE characterizes $\mathscr{L}(X_\infty)$, and what information the SFPE provides about the tail behavior of $\mathscr{L}(X_\infty)$.

We close this section with some observations of a more general kind, exemplified by the present model. Writing (1.39) in the form

$$X_n \;=\; \phi(X_{n-1},\varepsilon_n),\quad n\ge 1,$$

where $\phi(x,y)=\phi_{\alpha,\beta,\lambda}(x,y)=\alpha x+(\beta+\lambda x^2)^{1/2}y$, we immediately infer, by using the independence of X_{n-1},ε_n and the identical distribution of the innovations, that $(X_n)_{n\ge 0}$ forms a temporally homogeneous Markov chain (MC) with state space $\mathbb{R}$ and transition kernel

$$P(x,A) \;=\; \mathbb{P}((\alpha x+(\beta+\lambda x^2)^{1/2}\varepsilon\in A),\quad A\in\mathscr{B}(\mathbb{R}).$$

The continuity of $\phi(\cdot,y)$ for any $y\in\mathbb{R}$ further shows that $(X_n)_{n\ge 0}$ forms a *Feller chain*, defined by the property that

$$x\;\mapsto\; Pf(x)\;:=\;\int f(y)\,P(x,dy)\;=\;\mathbb{E}f(\alpha x+(\beta+\lambda x^2)^{1/2}\varepsilon)\;\in\;\mathscr{C}_b(\mathbb{R})$$

whenever $f\in\mathscr{C}_b(\mathbb{R})$. In other words, a *Feller kernel* P maps bounded continuous functions to bounded continuous functions. Next we point out that π forms a solution to the SFPE (1.40), i.e. $X\stackrel{d}{=}\phi(X,\varepsilon)$, iff it is a a stationary distribution for $(X_n)_{n\ge 0}$. The latter means that

$$\pi P\;:=\;\int P(x,\cdot)\,\pi(dx)\;=\;\pi$$

and therefore that $\mathscr{L}(X_{n-1})=\pi$ implies $\mathscr{L}(X_n)=\pi$. Thus, to determine all solutions to the SFPE means to find all stationary distributions of the MC $(X_n)_{n\ge 0}$. Here is a lemma that sometimes provides a simple tool to check the existence of a stationary distribution for a Feller chain on $\mathbb{R}$.

Lemma 1.23. *Let $(X_n)_{n\ge 0}$ be a Feller chain on $\mathbb{R}$.*

(a) If $X_n\stackrel{d}{\to}X_\infty$, then $\mathscr{L}(X_\infty)$ is a stationary distribution.
(b) If $(X_n)_{n\ge 0}$ is tight, then there exists a stationary distribution.

Proof. Problem 1.24 □

Problems

Problem 1.24. Prove Lemma 1.23. [Hint for part (b): Show that tightness implies that $(n^{-1}\sum_{k=1}^n\mu P^k)_{n\ge 1}$, contains a weakly convergent subsequence, where P^k denotes the k-step transition kernel of the chain and $\mu:=\mathbb{P}(X_0\in\cdot)$. Then verify that the weak limit is necessarily a stationary distribution.]

Problem 1.25. Given the random recursive equation (1.38) with $\lambda \in (0,1)$, $\mathbb{E}\varepsilon^2 = 1$ and $\mathbb{E}X_0^2 < \infty$, prove the following assertions:

(a) $(X_n)_{n\ge 0}$ is L^2-bounded, that is $\sup_{n\ge 0} \mathbb{E}X_n^2 < \infty$.
(b) $(X_n)_{n\ge 0}$ possesses a stationary distribution which is nondegenerate.

Problem 1.26. Consider the random recursive equation (1.39) with $\alpha \neq 0$, $\mathbb{E}|\varepsilon| < \infty$ (thus $\mathbb{E}\varepsilon = 0$) and $\mathbb{E}|X_0| < \infty$.

(a) Prove that $(\alpha^{-n}X_n)_{n\ge 0}$ is a martingale.
(b) Assuming $\mathbb{E}\varepsilon^2 < \infty$ and $\mathbb{E}X_0^2 < \infty$, find a necessary and sufficient condition on $(\alpha,\beta,\lambda) \in (\mathbb{R}\backslash\{0\}) \times \mathbb{R}_> \times \mathbb{R}_>$ for $(X_n)_{n\ge 0}$ to be L^2-bounded.

1.7 A noisy voter model on a directed tree

Let $\mathbb{T}_3(n) = \bigcup_{k=0}^n \{1,2,3\}^k$ be the rooted homogenous tree of order 3 and height n in *Ulam-Harris labeling* and $\mathbb{T}_3 = \mathbb{T}_3(\infty)$ its infinite height counterpart. This means that $\{1,2,3\}^0$ consists of the root $\varnothing$ and that each vertex $\mathsf{v} = (\mathsf{v}_1,...,\mathsf{v}_k) \in \{1,2,3\}^k$ at level k ($< n$ for $\mathbb{T}_3(n)$) has exactly 3 children, labeled $(\mathsf{v}_1,...,\mathsf{v}_k,i)$ for $i = 1,2,3$ [☞ Fig. 1.3 below]. Let us write $\mathsf{v}_1...\mathsf{v}_k$ as shorthand for $(\mathsf{v}_1,...,\mathsf{v}_k)$, $|\mathsf{v}|$ for the length of v, and uv for the concatenation of u and v. Note that u is the parent node of $\mathsf{u1},\mathsf{u2},\mathsf{u3}$.

For any fixed $n \ge 1$, let $\{X_n(\mathsf{v}) : \mathsf{v} \in \{1,2,3\}^n\}$ be a family of iid *Bern*(p)-variables ($p > 0$) and $\{\xi(\mathsf{v}) : \mathsf{v} \in \mathbb{T}_3(n-1)\}$ a second family of iid *Bern*(ε)-variables ($\varepsilon > 0$ small) independent of the former one. As in ALDOUS & BANDYOPADHYAY [1, Example 13], we now define recursively

$$X_n(\mathsf{u}) := \xi(\mathsf{u}) + \mathbf{1}_{\{X_n(\mathsf{u1})+X_n(\mathsf{u2})+X_n(\mathsf{u3})\ge 2\}} \bmod 2 \tag{1.41}$$

and $X_n := X_n(\varnothing)$. A possible interpretation, reflecting the title of this subsection, is the following: Each parent node adopts the majority opinion, which can be 0 or 1, of its children, except with a small probability ε adopting the opposite opinion.

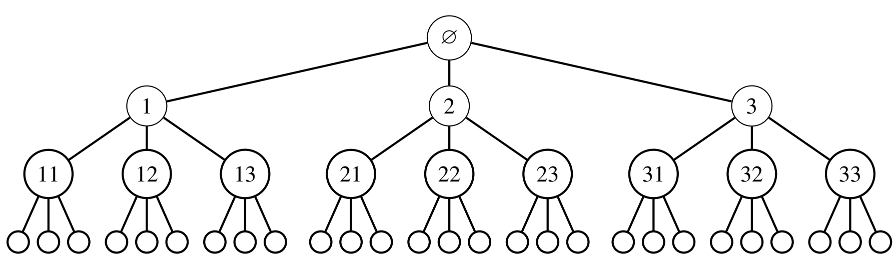


Fig. 1.3 The rooted homogenous tree $\mathbb{T}_3(3)$ with Ulam-Harris labeling

Following [1], we call the process $(X_n(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}_3(n)}$ a *recursive tree process (RTP) of depth n* and note that the recursion is bottom-up because the value of $X_n(\mathsf{v})$ is defined via the values of the corresponding variables of the children $\mathsf{v}i$ for $i = 1,2,3$. Hence, the terminal or output value is $X_n(\varnothing)$.

The reader is asked in Problem 1.27 to verify the basic recursive relation

$$X_n \stackrel{d}{=} \xi + \mathbf{1}_{\{X_{n-1}+X'_{n-1}+X''_{n-1}\geq 2\}} \bmod 2 \tag{1.42}$$

for $n \geq 1$, where $X_{n-1}, X'_{n-1}, X''_{n-1}$ are iid and independent of $\xi \stackrel{d}{=} Bern(\varepsilon)$, and $\mathscr{L}(X_0(\varnothing)) = Bern(p)$. This constitutes again a random recursive equation for the X_n, but only in terms of their distributions. In other words, we are given here a mapping that maps the distribution of X_{n-1} to the distribution of X_n. Now it is readily seen that, if $\mathscr{L}(X_{n-1}) = Bern(q)$, then $\mathscr{L}(X_n) = Bern(g(q))$, where

$$g(s) \;:=\; (1-\varepsilon)(s^3+3s^2(1-s)) + \varepsilon(1-s^3-3s^2(1-s)) \tag{1.43}$$

for $s \in [0,1]$. As Fig. 1.4 shows, the function g has three fixed points $p(\varepsilon), \frac{1}{2}, 1-p(\varepsilon)$ with $p(\varepsilon) \to 0$ as $\varepsilon \to 0$. Hence, if $\mathscr{L}(X_0) = Bern(q)$ with q being one of these fixed points, then $\mathscr{L}(X_n) = Bern(q)$ for all $n \geq 1$. The asymptotic behavior of X_n when $\mathscr{L}(X_0) = Bern(q)$ for $q \notin \{p(\varepsilon), \frac{1}{2}, 1-p(\varepsilon)\}$ is discussed in Problem 1.28.

Returning to the RTP $(X_n(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}_3(n)}$ defined above, it follows that the marginal distributions of all $X_n(\mathsf{v})$ are the same whenever $Bern(q)$ for $q \in \{p(\varepsilon), \frac{1}{2}, 1-p(\varepsilon)\}$ is chosen as the distribution of the variables at the bottom of the tree (level n). In this case the RTP is called *invariant*, and it may be extended to an invariant RTP $(X_n(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}_3}$ on the infinite tree $\mathbb{T}_3$ with the help of Kolmogorov's consistency theorem.

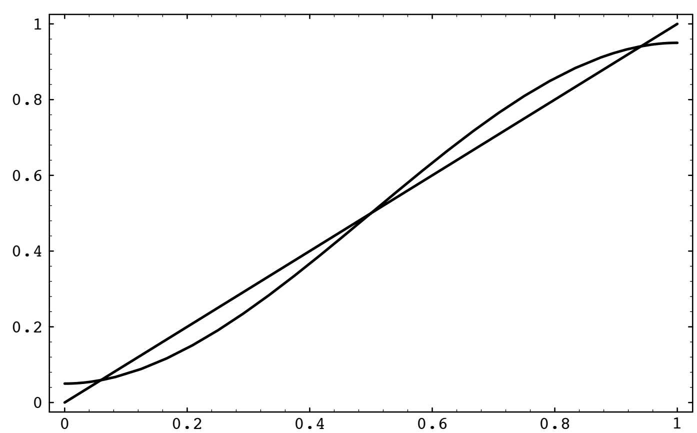


Fig. 1.4 The function $g(s) = (1-\varepsilon)(s^3+3s^2(1-s)) + \varepsilon(1-s^3-3s^2(1-s))$ with $\varepsilon = 0.05$.

Problems

Problem 1.27. Prove (1.42) under the stated assumptions.

Problem 1.28. Let $(X_n)_{n\ge 0}$ be a sequence of Bernoulli variables satisfying (1.42) with $\mathscr{L}(\xi)=Bern(\varepsilon)$ for any small ε and $\mathscr{L}(X_0)=Bern(q)$ for any $q\in[0,1]$. Let g be defined by (1.43) with fixed points $p(\varepsilon),\frac{1}{2},1-p(\varepsilon)$. Prove that

$$X_n \xrightarrow{d} X_\infty, \quad \text{where } \mathscr{L}(X_\infty) \;=\; \begin{cases} Bern(p(\varepsilon)), & \text{if } q<\frac{1}{2},\\ Bern(1/2), & \text{if } q=\frac{1}{2},\\ Bern(1-p(\varepsilon)), & \text{if } q>\frac{1}{2}.\end{cases}$$

Problem 1.29. Here is a simpler variation of the noisy voter model on the binary trees $\mathbb{T}_2(n)=\bigcup_{k=0}^n\{1,2\}^n$, $n\ge 1$: Consider an RTP $(X_n(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}_2(n)}$ of depth n with a family $\{X_n(\mathsf{v}):\mathsf{v}\in\{1,2\}\}^n$ of iid $Bern(p)$-variables $(0\le p\le 1)$. For any parental vertex $\mathsf{u}\in\mathbb{T}_2(n-1)$, define

$$X_n(\mathsf{u}) \;:=\; \xi(\mathsf{u})+X_n(\mathsf{u}\zeta(\mathsf{u})) \bmod 2,$$

where $\{(\xi(\mathsf{u}),\zeta(\mathsf{u})):\mathsf{u}\in\mathbb{T}_2(n-1)\}$ is independent of $\{X_n(\mathsf{v}):\mathsf{v}\in\{1,2\}\}^n$ and consisting of iid random vectors with common distribution $Bern(\varepsilon)\otimes\mathrm{Unif}(\{1,2\})$ for some $\varepsilon\in(0,1)$. This means that u adopts the opinion of the randomly chosen child $\mathsf{u}\zeta(\mathsf{u})$, except with probability ε adopting the opposite opinion. Put $X_n:=X_n(\varnothing)$ for $n\ge 0$, where $\mathscr{L}(X_0(\varnothing))=Bern(p)$, and prove:

(a) For all $n\ge 1$, $X_n \stackrel{d}{=} \xi+X_{n-1} \bmod 2$.
(b) For any $p\in[0,1]$, X_n converges in distribution to $Bern(1/2)$.
(c) The RTP's defined above are invariant iff $p=1/2$.

1.8 An excursion to hydrology: the Horton-Strahler number

The *Strahler number*[2] or *Horton-Strahler number* was first developed by two Americans, the ecologist and soil scientist HORTON [67] and the geoscientist STRAHLER [104, 105], as a measure in hydrology for stream size based on a hierarchy of tributaries[3] In this context, it is also referred to as the *Strahler stream order*. It further arises in the analysis of hierarchical biological structures (like biological trees) and of social networks. BENDER in his introductory book [14] on mathematical modeling has a nicely written section on stream networks which provides a little more background information.

[2] in German called *Fluss- oder Gewässerordnungszahl nach Strahler*

[3] defined as a river which flows into a parent river or lake instead of directly flowing into a sea or ocean.

In mathematics, the Strahler number is simply a numerical measure of the branching complexity of a finite (mathematical) tree and defined as follows (when using Ulam-Harris labeling as in the previous section): Starting at the bottom, all leaves (the sources in the hydrological context) get Strahler number 1. For any other vertex v, suppose it has children $\mathsf{v1},...,\mathsf{vk}$ with respective Strahler numbers $S(\mathsf{v1}),...,S(\mathsf{vk})$ having maximal value s, say. Then the Strahler number $S(\mathsf{v})$ at v is recursively defined as

- s if this value is attained uniquely among the $S(\mathsf{vi})$, $\mathsf{i}=1...,k$.
- $s+1$ if $S(\mathsf{vi})=S(\mathsf{vj})$ for at least two distinct $\mathsf{i},\mathsf{j}\in\{1,...,k\}$.

Finally, the Strahler number or index of the tree is defined as $S(\varnothing)$.

In the river network context, the trees are typically binary and the numbers are assigned to the edges leaving a node upwards rather than the node itself [☞ Fig. 1.5]. Of course, the nodes represent the points where two streams come together. When two streams of the same order k meet, they form a stream of order $k+1$, whereas if one of them has a lower order it is viewed as subordinate to the higher stream, the order of which thus remains unchanged. The index of a stream or river may range from 1 (a stream with no tributaries) to 12 (the most powerful river, the Amazon, at its mouth). The Ohio River is of order eight and the Mississippi River is of order 10. 80% of the streams and rivers on the planet are first or second order [☞ http://en.wikipedia.org/wiki/Strahler_number].

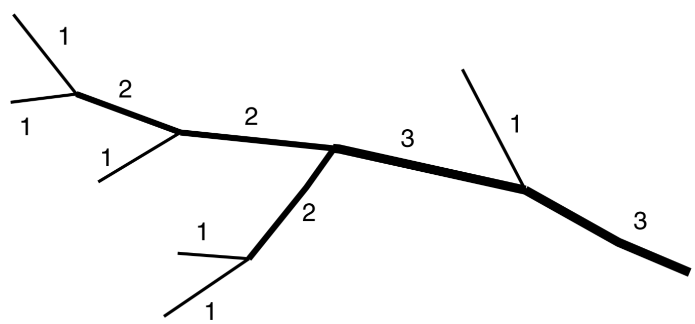


Fig. 1.5 U.S. Corps of Engineer diagram showing the Strahler stream order.
[license by http://creativecommons.org/licenses/by-sa/3.0/deed.en]

Now let us return to the mathematical framework. Given a finite tree τ, the above definition of the $(S(\mathsf{v}))_{\mathsf{v}\in\tau}$ provides us with another example of a RTP of finite depth which becomes stochastic as soon as τ is chosen by some random mechanism. For instance τ may be the realization of a Galton-Watson tree up to some finite generation. In this case, with $(Z_n)_{n\ge0}$ denoting the associated GWP, one can easily derive the following random recursive equation for the Strahler index S_n of the Galton-Watson tree up to generation n:

$$S_n \;=\; \mathbf{1}_{\{Z_1=0\}}+\mathbf{1}_{\{Z_1\ge1\}}\Big(\max_{1\le k\le Z_1} S_{n-1}(k)+\mathbf{1}_{\{N_n>1\}}\Big), \tag{1.44}$$

where $S_{n-1}(k)$ denotes the Strahler index of the subtree rooted at the k^{th} member of the first generation and $N_n := |\{1 \le k \le Z_1 : S_{n-1}(k) = \max_{1\le i\le Z_1} S_{n-1}(i)\}|$. In this formulation, only $S_{n-1}(1),...,S_{n-1}(Z_1)$ are specified and, given Z_1, conditionally i.i.d. with the same distribution as S_{n-1}. However, we can also define an infinite sequence $(S_{n-1}(k))_{k\ge 1}$ of independent copies of S_{n-1} which are unconditionally independent of Z_1. This does not affect the validity of (1.44). Since N_n is then obviously a measurable function of $Z_1, S_{n-1}(1), S_{n-1}(2), ...$, we see that (1.44) fits into the general form

$$S_n \;=\; \Psi(Z_1, S_{n-1}(1), S_{n-1}(2), ...)$$

for some measurable function Ψ (not depending on n).

As another example, one may consider $(S(v))_{v\in\tau}$ when τ is drawn at random from the set of binary trees with n nodes. This was done by DEVROYE & KRUSCZEWSKI [32] who showed that, if $S_n := S(\varnothing)$, then

$$\begin{aligned} \mathbb{E}S_n \;&=\; \log_4 n + O(1) \quad \text{as } n\to\infty \\ \text{and}\quad \mathbb{P}(|S_n - \log_4 n| \ge x) \;&\le\; c4^{-x} \end{aligned}$$

for all $x > 0$, $n \ge 1$ and some $c > 0$. Therefore, the distribution of S_n exhibits

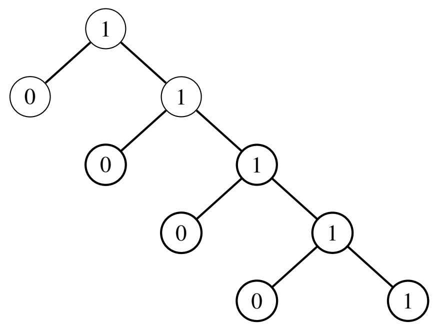


Fig. 1.6 The binary tree with 5 internal nodes having minimal Strahler number 1. Due to its shape when including external nodes (those with numbers 0) it is sometimes called *"gourmand de la vigne"*.

very sharp concentration about its mean which is approximately equal to $\log_4 n$. In connection with this result it is worthwhile to point out that the binary trees with extremal Strahler numbers are

- the single-stranded tree with n nodes and Strahler number 1 [☞ Fig. 1.6],
- the complete tree with k levels, $n = 2^k - 1$ nodes and Strahler number $S_n = k = \log_2(n+1)$ [☞ Fig. 1.7].

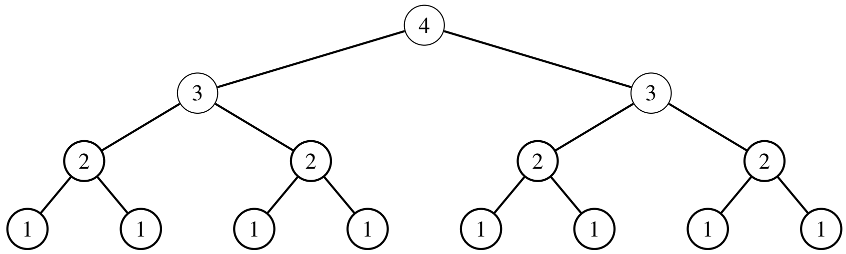


Fig. 1.7 The binary tree with $2^4 - 1 = 15$ internal nodes and maximal Strahler number 4.

Chapter 2
Renewal theory

Terminology. An additive sequence $(S_n)_{n\geq 0}$ of real-valued random variables with increments $X_1, X_2, \ldots$ is called

random walk (RW)	if $X_1, X_2, \ldots$ are iid and independent of S_0;
renewal process (RP)	if it is a RW such that $S_0, X_1, X_2, \ldots$ are nonnegative and $\mathbb{P}(X_1 > 0) > 0$;
standard random walk (SRW)	if it is a RW with $S_0 = 0$. It is also called *zero-delayed RW*;
standard renewal process (SRP)	if it is a RP with $S_0 = 0$. It is also called *zero-delayed RP*.

Given a RW $(S_n)_{n\geq 0}$, we use X for a generic copy of its increments. The initial variable S_0 is also called *delay*, the mean of X, if it exists, the *drift* of $(S_n)_{n\geq 0}$. Finally, we are given a *standard model* $(\Omega, \mathfrak{A}, (S_n)_{n\geq 0}, (\mathbb{P}_\lambda)_{\lambda\in\mathscr{P}(\mathbb{R})})$ if $(S_n)_{n\geq 0}$, defined on $(\Omega, \mathfrak{A})$, constitutes a RW under each $\mathbb{P}_\lambda$ with the same increment distribution F, say, and $\mathbb{P}_\lambda(S_0 \in \cdot) = \lambda$, hence $\mathbb{P}_\lambda(S_n \in \cdot) = \lambda * F^{*n}$ for each $n \in \mathbb{N}$, where F^{*n} denotes n-fold convolution of F.

2.1 An introduction and first results

Let us begin with a short description of the classical renewal problem: Suppose we are given an infinite supply of light bulbs which are used one at a time until they fail. Their lifetimes are denoted as $X_1, X_2, \ldots$ and assumed to be iid random variables with positive mean μ. If the first light bulb is installed at time $S_0 := 0$, then

$$S_n := \sum_{k=1}^{n} X_k \quad \text{for } n \geq 1$$

denotes the time at which the n^{th} bulb fails and is replaced with a new one. In other words, each S_n marks a renewal epoch. Due to this interpretation, a sequence

$(S_n)_{n\ge 0}$ with iid nonnegative increments having positive mean is called *renewal process (RP)*. Let $N(t)$ denote the number of renewals up to time t, that is

$$N(t) \;:=\; \sup\{n\ge 0 : S_n\le t\} \quad \text{for } t\ge 0. \tag{2.1}$$

An equivalent definition is

$$N(t) \;:=\; \sum_{n\ge 1}\mathbf{1}_{[0,t]}(S_n)$$

and has the advantage that it immediately extends to general measurable subsets A of $\mathbb{R}_\ge$ by putting

$$N(A) \;:=\; \sum_{n\ge 1}\mathbf{1}_A(S_n) \;=\; \sum_{n\ge 1}\delta_{S_n}(A). \tag{2.2}$$

We see that N is in fact a *random counting measure*, also called *point process*, on $(\mathbb{R}_\ge,\mathscr{B}(\mathbb{R}_\ge))$. By further defining its *intensity measure*

$$\mathbb{U}(A) \;:=\; \mathbb{E}N(A) \;=\; \sum_{n\ge 1}\mathbb{P}(S_n\in A), \quad A\in\mathscr{B}(\mathbb{R}_\ge), \tag{2.3}$$

we arrive at the so-called *renewal measure* of $(S_n)_{n\ge 1}$ which measures the expected number of renewals in a set and is one of the central objects in renewal theory. Its “distribution function”

$$[0,\infty) \;\ni\; t \;\mapsto\; \mathbb{U}(t) \;:=\; \mathbb{U}([0,t]) \;=\; \sum_{n\ge 1}\mathbb{P}(S_n\le t) \tag{2.4}$$

is called *renewal function* of $(S_n)_{n\ge 1}$ and naturally of particular interest.

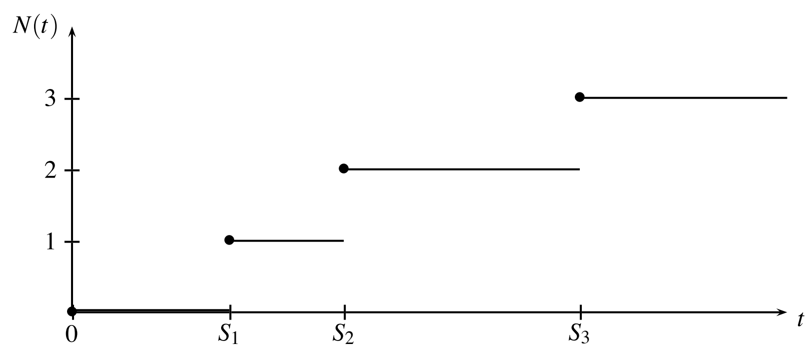


Fig. 2.1 The renewal counting process $(N(t))_{t\ge 0}$ with renewal epochs $S_1,S_2,\ldots$

Natural questions to be asked now are ...

(Q1) Is the number of renewals up to time t, denoted as $N(t)$, almost surely finite for all $t>0$? And what about its expectation $\mathbb{E}N(t)$?

(Q2) What is the asymptotic behavior of $t^{-1}N(t)$ and its expectation as $t \to \infty$, that is the long run average (expected) number of renewals per unit of time?
(Q3) What can be said about the long run behavior of $\mathbb{E}(N(t+h)-N(t))$ for any fixed $h > 0$?

... with partial answers provided by the following lemma. Given a distribution F, let F^{*n} denote its n-fold convolution for $n \in \mathbb{N}$ and $F^{*0} := \delta_0$.

Lemma 2.1. *Let $(S_n)_{n\ge 0}$ be a RP with $S_0 = 0$, increment distribution F, drift $\mu = \mathbb{E}S_1 \in (0,\infty]$ and renewal measure $\mathbb{U} = \sum_{n\ge 1}\mathbb{P}(S_n \in \cdot)$. Then the following assertions hold true:*

(a) $N(t) < \infty$ *a.s. for all* $t \ge 0$.
(b) $\mathbb{P}(N(t) = n) = F^{*n}(t) - F^{*(n+1)}(t)$ *for all* $n \in \mathbb{N}_0$ *and* $t \ge 0$.
(c) $\mathbb{U} = \sum_{n\ge 1} F^{*n}$, *in particular* $\mathbb{U}(t) = \sum_{n\ge 1} F^{*n}(t)$ *for any* $t \ge 0$.
(d) $\mathbb{E}e^{aN(t)} < \infty$ *for all* $t \ge 0$ *and some* $a > 0$.
(e) $t^{-1}N(t) \to \mu^{-1}$ *a.s. with the usual convention that* $\infty^{-1} := 0$.
(f) **[Elementary Renewal Theorem]** $\lim_{t\to\infty} t^{-1}\mathbb{U}(t) = \mu^{-1}$.

Proof. (a) follows immediately from $S_n \to \infty$ a.s.

(b) follows when noting that

$$\{N(t) = n\} = \{S_n \le t < S_{n+1}\} = \{S_n \le t\}\backslash\{S_{n+1} \le t\}$$

for all $n \in \mathbb{N}_0$ and $t \ge 0$.

(c) Here it suffices to note that $\mathscr{L}(S_n) = F^{*n}$ for all $n \in \mathbb{N}_0$.

(d) Since $\mu = \mathbb{E}S_1 > 0$, there exists $b > 0$ such that $F(b) < 1$. Consider the RP $(S'_n)_{n\ge 0}$ with increments given by $X'_n := b\mathbf{1}_{\{X_n > b\}}$ for $n \in \mathbb{N}$ and renewal counting process $(N'(t))_{t\ge 0}$. Then $S'_n \le S_n$ for all $n \in \mathbb{N}_0$ implies $N(t) \le N'(t)$ for all $t \ge 0$. Now observe that, for $n \in \mathbb{N}$ and $0 < t < b$,

$$\mathbb{P}(N'(t) > n) \;=\; \mathbb{P}(X'_1 = \ldots = X'_n = 0) \;=\; F(b)^n$$

implying $\mathbb{E}e^{aN(t)} \le \mathbb{E}e^{aN'(t)} < \infty$ for any $a < -\log F(b)$ as one easily see. We leave it as an exercise [☞ Problem 2.3] to extend the last assertion to all $t \ge b$.

(e) Since $N(t) \to \infty$ a.s., the SLLN implies $N(t)^{-1}S_{N(t)} \to \mu$ a.s. By combining this with the obvious inequality $S_{N(t)} \le t < S_{N(t)+1}$ [use (2.1)] we find

$$\frac{S_{N(t)}}{N(t)} \;\le\; \frac{t}{N(t)} \;\le\; \frac{N(t)+1}{N(t)}\cdot\frac{S_{N(t)+1}}{N(t)+1}$$

and then obtain $t^{-1}N(t) \to \mu$ a.s. by letting t tend to ∞ in this inequality.

(f) Use $\mathbb{U}(t) = \mathbb{E}N(t)$, (e) and Fatou's lemma to infer

$$\liminf_{t\to\infty}\frac{\mathbb{U}(t)}{t} \;\ge\; \mathbb{E}\left(\liminf_{t\to\infty}\frac{N(t)}{t}\right) \;\ge\; \frac{1}{\mu}.$$

Towards a reverse estimate, notice that

$$N(t)+1=\tau(t):=\inf\{n\ge 0: S_n>t\}$$

and thus $\mathbb{U}(t)+1=\mathbb{E}\tau(t)$ for all $t\ge 0$. If $X_1,X_2,\ldots$ are bounded by some $c>0$, in particular giving $\mu<\infty$, then we obtain with the help of Wald's identity [☞ Prop. 2.53]

$$\mathbb{E}\tau(t) \;=\; \frac{\mathbb{E}S_{\tau(t)}}{\mu} \;=\; \frac{t+\mathbb{E}(S_{\tau(t)}-t)}{\mu} \;\le\; \frac{t+c}{\mu}$$

and thereby

$$\limsup_{t\to\infty}\frac{\mathbb{U}(t)}{t} \;\le\; \limsup_{t\to\infty}\frac{\mathbb{E}\tau(t)}{t} \;\le\; \frac{1}{\mu}$$

as required. Otherwise, consider the RP $(S_{c,n})_{n\ge 0}$ with generic increment $X\wedge c$, drift $\mu_c:=\mathbb{E}(X\wedge c)>0$ and renewal measure $\mathbb{U}_c$. Plainly, $\mathbb{U}(t)\le\mathbb{U}_c(t)$ for all $t\ge 0$, whence

$$\limsup_{t\to\infty}\frac{\mathbb{U}(t)}{t} \;\le\; \limsup_{t\to\infty}\frac{\mathbb{U}_c(t)}{t} \;\le\; \frac{1}{\mu_c}$$

for any $c>0$. Finally, use $\lim_{c\to\infty}\mu_c=\mu$ to arrive at

$$\limsup_{t\to\infty}\frac{\mathbb{U}(t)}{t} \;\le\; \frac{1}{\mu}$$

which completes the proof of the lemma. □

Remark 2.2. The reader is asked in Problem 2.4 below to verify that all assertions of the previous lemma except for (b) remain valid if $(S_n)_{n\ge 0}$ has arbitrary delay distribution $F_0:=\mathbb{P}(S_0\in\cdot)$. As for part (b), it must be modified as

$$\mathbb{P}(N(t)=n) \;=\; F_0*F^{*(n-1)}(t)-F_0*F^{*n}(t)$$

for $n\in\mathbb{N}$ and $t\ge 0$, and $\mathbb{P}(N(t)=0)=1-F_0(t)$.

Problems

Problem 2.3. Let $(S_n)_{n\ge 0}$ be a RP with $S_0=0$, increment distribution $Bern(p)$ for some $p\in(0,1)$ and renewal counting process $(N(t))_{t\ge 0}$. Prove the following assertions:

(a) $\mathscr{L}(N(n))=NBin(n+1,p)$ for each $n\in\mathbb{N}_0$.
(b) For any $t\ge 0$, $\mathbb{E}e^{aN(t)}<\infty$ iff $a<-\log(1-p)$.

Problem 2.4. Prove Lemma 2.1, with part (b) modified in the form stated in Rem. 2.2, for a general delayed RP with delay distribution F_0.

2.2 An important special case: exponential lifetimes

A case of particular interest occurs when the RP $(S_n)_{n\geq 0}$ has exponential increments, i.e. $F = Exp(1/\mu)$ for some $\mu > 0$. Then S_n has a Gamma distribution with parameters n and $1/\mu$, i.e. $F^{*n} = \Gamma(n, 1/\mu)$, the $\lambda\!\!\lambda$-density of which equals

$$f_n(x) \;=\; \frac{x^{n-1}}{\mu^n(n-1)!}\, e^{-x/\mu}\, \mathbf{1}_{(0,\infty)}(x)$$

for each $n \in \mathbb{N}$. Since $\mathbb{U} = \sum_{n\geq 1} F^{*n}$, we find that its λ-density u, called *renewal density*, equals

$$u(x) \;=\; \sum_{n\geq 1} f_n(x) \;=\; e^{-x/\mu} \sum_{n\geq 1} \frac{x^{n-1}}{\mu^n(n-1)!} \;=\; \frac{1}{\mu}$$

for all $x > 0$, hence $\mathbb{U} = \mu^{-1}\lambda\!\!\lambda^+$, where $\lambda\!\!\lambda^+ := \lambda\!\!\lambda(\cdot \cap \mathbb{R}_>)$. Equivalently, the expected number of renewals in an interval $[t, t+h] \subset \mathbb{R}_\geq$ of length $h > 0$ *always* equals $\mu^{-1}h$. The reason lurking behind this phenomenon is of course the lack of memory property of the exponential distribution. Here is a heuristic argument: Suppose we start observing the RP at a time $t > 0$ and reset our clock to 0. Then renewals occur at $S_{\tau(t)} - t, S_{\tau(t)+1} - t, \ldots$ with interrenewal times $X_{\tau(t)+1}, X_{\tau(t)+2}, \ldots$ after the *delay* $R(t) := S_{\tau(t)} - t$. Proposition **??** will show that $R(t)$ and $X_{\tau(t)+1}, X_{\tau(t)+2}, \ldots$ are independent and the latter sequence further iid with $\mathscr{L}(X_{\tau(t)+1}) = Exp(1/\mu)$. Consequently, we will see the same arrival pattern as someone who starts observing the system at time 0 if $\mathscr{L}(R(t)) = Exp(1/\mu)$ as well. But this is indeed intuitively clear by the lack of memory property and may also formally be proved fairly easily [☞ Problem 2.6].

Turning to the associated renewal counting process $(N(t))_{t\geq 0}$, the previous considerations entail that $\mathscr{L}(N(t+h) - N(t)) = \mathscr{L}(N(h))$ for any $t \geq 0$ and $h > 0$ which means that $(N(t))_{t\geq 0}$ *has stationary increments*. They further provide some evidence (though not a proof) that the numbers of renewals in $[0,t]$ and $[t, t+h]$ are independent. In fact, one can more generally show that, for any choice $0 = t_0 < t_1 < \ldots < t_n < \infty$ and $n \in \mathbb{N}$, the random variables $N(t_k) - N(t_{k-1})$, $k = 1, \ldots, n$, are independent which means that $(N(t))_{t\geq 0}$ *has independent increments*. It remains to find the distribution of $N(t)$ for any $t > 0$. By Lemma 2.1(b), it follows that $p_n(t) := \mathbb{P}(N(t) = n)$ satisfies

$$p_n(t) \;=\; F^{*n}(t) - F^{*(n+1)}(t)$$

for all $t > 0$ and $n \in \mathbb{N}_0$. If $n = 0$, this yields

$$p_0(t) \;=\; 1 - F(t) \;=\; e^{-t/\mu}, \quad t > 0.$$

For $n \geq 1$, $p_n(\cdot)$ is differentiable with

$$p_n'(t) \;=\; f_n(t) - f_{n+1}(t) \;=\; e^{-t/\mu}\left(\frac{t^{n-1}}{\mu^n(n-1)!} - \frac{t^n}{\mu^{n+1}n!}\right), \quad t > 0,$$

and $p_n(0) = 0$. Consequently,

$$p_n(t) \;=\; e^{-t/\mu}\,\frac{(t/\mu)^n}{n!}, \quad t > 0,$$

and we have arrived at the following result.

Theorem 2.5. *If $(S_n)_{n\ge 0}$ is a SRP having exponential increments with parameter $1/\mu$, then the associated renewal counting process $(N(t))_{t\ge 0}$ forms a* ***homogeneous Poisson process with intensity (rate)*** *$1/\mu$, that is:*

(PP1) $N(0) = 0$.

(PP2) $(N(t))_{t\ge 0}$ *has independent increments, i.e.,*

$$N(t_1), N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$$

are independent random variables for each choice of $n \in \mathbb{N}$ and $0 < t_1 < t_2 < \ldots < t_n < \infty$.

(PP3) $(N(t))_{t\ge 0}$ *has stationary increments, i.e., $N(s+t) - N(s) \stackrel{d}{=} N(t)$ for all $s,t \ge 0$.*

(PP4) $N(t) \stackrel{d}{=} Poisson(t/\mu)$ *for each $t \ge 0$.*

If $\mu = 1$, then $(N(t))_{t\ge 0}$ is also called ***standard Poisson process****.*

Poisson processes have many nice properties some of which are stated in the Problem section below.

Problems

Problem 2.6. Let $(S_n)_{n\ge 0}$ be a RP with $S_0 = 0$, increment distribution $Exp(1/\mu)$ for some $\mu > 0$ and renewal measure $\mathbb{U} = \sum_{n\ge 1} \mathbb{P}(S_n \in \cdot)$. Let also $R(t) = S_{\tau(t)} - t$ for $t \ge 0$. Prove the following assertions:

(a) $\mathbb{P}(R(t) > s) = e^{-(s+t)/\mu} + \int_{[0,t]} e^{-(t+s-x)/\mu}\,\mathbb{U}(dx)$ for all $s > 0$.

(b) $\mathscr{L}(R(t)) = Exp(1/\mu)$ for all $t \ge 0$. [Use (a).]

Problem 2.7. Let $(N(t))_{t\ge 0}$ be a homogeneous Poisson process with intensity θ. Find the conditional distribution of $N(s)$ given $N(t) = n$ for any $0 < s < t$ and $n \in \mathbb{N}_0$.

Problem 2.8 (Superposition of Poisson processes). Given two independent homogeneous Poisson processes $(N_1(t))_{t\ge 0}$ and $(N_2(t))_{t\ge 0}$ with intensities θ_1 and θ_2, respectively, prove that the *superposition* $N(t) := N_1(t) + N_2(t)$ for $t \geq 0$ forms a homogeneous Poisson process with intensity $\theta_1 + \theta_2$.

Problem 2.9 (Thinning of Poisson processes). Given a homogeneous Poisson process $(N(t))_{t\geq 0}$ with associated SRP $(S_n)_{n\geq 0}$ of arrival epochs, let $(\xi_n)_{n\geq 1}$ be an independent sequence of iid *Bern(p)* variables for some $p \in (0,1)$. Let $(N_1(t))_{t\geq 0}$ be the *thinning* or *p-thinning* of $(N(t))_{t\geq 0}$, defined by

$$N_1(t) \;:=\; \sum_{n\geq 1} \xi_n \delta_{S_n}([0,t]), \quad t \geq 0,$$

and put $N_2(t) = N(t) - N_1(t)$ for $t \geq 0$.

Problem 2.10 (Conditional equidistribution of points). Let $(N(t))_{t\geq 0}$ be a homogeneous Poisson process with intensity θ and associated SRP $(S_n)_{n\geq 0}$. Let further $(U_n)_{n\geq 1}$ be a sequence of iid *Unif*$(0,1)$ variables. Prove that

$$\mathscr{L}((S_1,\dots,S_n)|N(t)=n) \;=\; \mathscr{L}((tU_{(1)},\dots,tU_{(n)}))$$

for all $t > 0$ and $n \in \mathbb{N}$, where $(U_{(1)},\dots,U_{(n)})$ denotes the increasing order statistic of the random vector $(U_1,\dots,U_n)$. This means that, given $N(t) = n$, a sample of $S_1,\dots,S_n$ may be generated by throwing n points uniformly at random into the interval $[0,t]$.

2.3 Lattice-type

A more profound analysis of the renewal measure $\mathbb{U}$ of a SRP $(S_n)_{n\geq 0}$ must take into account the fact that, if X takes values only in a closed discrete subgroup of $\mathbb{R}$, thus in $\mathbb{G}_d := d\mathbb{Z}$ for some $d > 0$, then $\mathbb{U}$ puts only mass on this subgroup as well and consequently looks very different from Lebesgue measure encountered in the previous section. The subsequent definitions provide the appropriate specifications of the *lattice-type* of a distribution F on $\mathbb{R}$ and of a RW $(S_n)_{n\geq 0}$.

Definition 2.11. For a distribution F on $\mathbb{R}$, its *lattice-span* $d(F)$ is defined as

$$d(F) \;:=\; \sup\{d \in [0,\infty] : F(\mathbb{G}_d) = 1\}.$$

Let $\{F_x : x \in \mathbb{R}\}$ denote the translation family associated with F, i.e., $F_x(B) := F(x+B)$ for all Borel subsets B of $\mathbb{R}$. Then F is called

– *nonarithmetic*, if $d(F) = 0$ and thus $F(\mathbb{G}_d) < 1$ for all $d > 0$.
– *completely nonarithmetic*, if $d(F_x) = 0$ for all $x \in \mathbb{R}$.
– *d-arithmetic*, if $d \in \mathbb{R}_>$ and $d(F) = d$.
– *completely d-arithmetic*, if $d \in \mathbb{R}_>$ and $d(F_x) = d$ for all $x \in \mathbb{G}_d$.

If X denotes any random variable with distribution F, thus $\mathscr{L}(X-x)=F_x$ for each $x\in\mathbb{R}$, then the previous attributes are also used for X, and we also write $d(X)$ instead of $d(F)$ and call it the lattice-span of X.

For our convenience, a nonarithmetic distribution is sometimes referred to as *0-arithmetic* hereafter, for example in the lemma below. A random variable X is nonarithmetic iff it is not a.s. taking values only in a lattice $\mathbb{G}_d$, and it is completely nonarithmetic if this is not either the case for any shifted lattice $x+\mathbb{G}_d$, i.e. any affine closed subgroup of $\mathbb{R}$. As an example of a nonarithmetic, but not completely nonarithmetic random variable we mention $X=\pi+Y$ with a standard Poisson variable Y. Then $d(X-\pi)=d(Y)=1$. If $X=\frac{1}{2}+Y$, then $d(X)=\frac{1}{2}$ and $d(X-\frac{1}{2})=1$. In this case, X is $\frac{1}{2}$-arithmetic, but not completely $\frac{1}{2}$-arithmetic. The following simple lemma provides the essential property of a completely d-arithmetic random variable $(d\ge 0)$.

Lemma 2.12. *Let X,Y be two iid random variables with lattice-span $d\ge 0$. Then $d\le d(X-Y)$ with equality holding iff X is completely d-arithmetic.*

Proof. Let F denote the distribution of X,Y. The inequality $d\le d(X-Y)$ is trivial, and since $(X+z)-(Y+z)=X-Y$, we also have $d(X+z)\le d(X-Y)$ for all $z\in\mathbb{R}$. Suppose X is *not* completely d-arithmetic. Then $d(X+z)>d$ for some $z\in\mathbb{G}_d$ and hence also $c:=d(X-Y)>d$. Conversely, if the last inequality holds true, then

$$1 \;=\; \mathbb{P}(X-Y\in\mathbb{G}_c) \;=\; \int_{\mathbb{G}_d}\mathbb{P}(X-y\in\mathbb{G}_c)\,F(dy)$$

implies

$$\mathbb{P}(X-y\in\mathbb{G}_c) \;=\; 1 \quad \text{for all } F\text{-almost all } y\in\mathbb{G}_d$$

and thus $d(X-y)\ge c>d$ for F-almost all $y\in\mathbb{G}_d$. Therefore, X cannot be completely d arithmetic. □

Definition 2.13. A RW $(S_n)_{n\ge 0}$ with increments $X_1,X_2,\ldots$ is called

– *(completely) nonarithmetic* if X_1 is (completely) nonarithmetic.
– *(completely) d-arithmetic* if $d>0$, $\mathbb{P}(S_0\in\mathbb{G}_d)=1$, and X_1 is (completely) d-arithmetic.

Furthermore, the lattice-span of X_1 is also called the lattice-span of $(S_n)_{n\ge 0}$ in any of these cases.

The additional condition on the delay in the d-arithmetic case, which may be restated as $d(S_0)=kd$ for some $k\in\mathbb{N}\cup\{\infty\}$, is needed to ensure that $(S_n)_{n\ge 0}$ is really concentrated on the lattice $\mathbb{G}_d$. The unconsidered case where $(S_n)_{n\ge0}$ has d-arithmetic increments but non- or c-arithmetic delay for some $c \notin \mathbb{G}_d \cup \{\infty\}$ will not play any role in our subsequent analysis.

2.4 Uniform local boundedness and stationary delay distribution

Given a RP $(S_n)_{n\ge0}$ with renewal measure $\mathbb{U} = \sum_{n\ge0} \mathbb{P}(S_n \in \cdot)$ and renewal counting measure $N = \sum_{n\ge0} \delta_{S_n}$, we now turn to question (Q3) about the asymptotic behavior of $\mathbb{U}((t,t+h]) = \mathbb{E}(N(t+h) - N(t))$ for any fixed $h > 0$. Notice that, unlike in the previous sections, summation in the definitions of $\mathbb{U}$ and N now ranges over $n \ge 0$. Denoting by λ and F the distribution of S_0 and X, we thus have

$$\mathbb{U} \;=\; \sum_{n\ge0} \lambda * F^{*n} \;=\; \lambda * \sum_{n\ge0} F^{*n} \;=\; \lambda * \mathbb{U}_0, \tag{2.5}$$

where $\mathbb{U}_0$ is the renewal measure of the SRP $(S_n - S_0)_{n\ge0}$. Assuming a standard model, (2.5) may in fact also be stated as $\mathbb{U}_\lambda = \lambda * \mathbb{U}_0$ for any $\lambda \in \mathscr{P}(\mathbb{R}_\ge)$ if $\mathbb{U}_\lambda$ denotes the renewal measure under $\mathbb{P}_\lambda$ and $\mathbb{U}_x$ is used for $\mathbb{U}_{\delta_x}$.

2.4.1 Uniform local boundedness

The first step towards our main results in the next sections is the following lemma which particularly shows *uniform local boundedness* of $\mathbb{U}_\lambda$, defined by

$$\sup_{t\in\mathbb{R}} \mathbb{U}([t,t+h]) \;<\; \infty \quad \text{for all } h > 0.$$

Lemma 2.14. *Let $(S_n)_{n\ge0}$ be a RP in a standard model. Then*

$$\sup_{t\in\mathbb{R}} \mathbb{P}_\lambda(N([t,t+h]) \ge n) \;\le\; \mathbb{P}_0(N(h) \ge n) \tag{2.6}$$

for all $h > 0$, $n \in \mathbb{N}_0$ and $\lambda \in \mathscr{P}(\mathbb{R}_\ge)$. In particular,

$$\sup_{t\in\mathbb{R}} \mathbb{U}_\lambda([t,t+h]) \;\le\; \mathbb{U}_0(h) \tag{2.7}$$

and $\{N([t,t+h]) : t \in \mathbb{R}\}$ is uniformly integrable under each $\mathbb{P}_\lambda$ for all $h > 0$.

Proof. If (2.6) holds true, then the uniform integrability of $\{N([t,t+h]) : t \in \mathbb{R}\}$ is a direct consequence, while (2.7) follows by summation over n. So (2.6) is the only assertion to be proved. Fix $t \in \mathbb{R}$, $h > 0$, and define $\tau := \inf\{n \ge 0 : S_n \in [t,t+h]\}$.

Then

$$N([t,t+h]) \;=\; \begin{cases} \sum_{k\ge 0}\mathbf{1}_{[t,t+h]}(S_{\tau+k}), & \text{if } \tau<\infty,\\ 0, & \text{otherwise.}\end{cases}$$

The desired estimate now follows from

$$\begin{aligned}
\mathbb{P}_\lambda(N([t,t+h])\ge n) \;&=\; \mathbb{P}_\lambda\left(\tau<\infty,\sum_{k\ge 0}\mathbf{1}_{[t,t+h]}(S_{\tau+k})\ge n\right)\\
&=\; \sum_{j\ge 0}\mathbb{P}_\lambda\left(\tau=j,\sum_{k\ge 0}\mathbf{1}_{[t,t+h]}(S_{j+k})\ge n\right)\\
&\le\; \sum_{j\ge 0}\mathbb{P}_\lambda\left(\tau=j,\sum_{k\ge 0}\mathbf{1}_{[0,h]}(S_{j+k}-S_j)\ge n\right)\\
&=\; \sum_{j\ge 0}\mathbb{P}_\lambda(\tau=j)\,\mathbb{P}_0(N(h)\ge n)\\
&=\; \mathbb{P}_\lambda(\tau<\infty)\,\mathbb{P}_0(N(h)\ge n)
\end{aligned}$$

for all $n\in\mathbb{N}$ and $\lambda\in\mathscr{P}(\mathbb{R}_\ge)$. □

2.4.2 *Finite mean case: the stationary delay distribution*

As already explained in the previous section, the behavior of $\mathbb{U}((t,t+h])$ is expected to be different depending on whether the underlying RP $(S_n)_{n\ge 0}$ is arithmetic or not. We make the standing assumption hereafter that $(S_n)_{n\ge 0}$ has either lattice-span $d=0$ or $d=1$. The latter is no restriction in the arithmetic case because one may otherwise switch to the RP $(d^{-1}S_n)_{n\ge 0}$. Recall that $\mathbb{G}_d=d\mathbb{Z}$ for $d>0$ and put also $\mathbb{G}_0=\mathbb{R}$ as well as $\mathbb{G}_{d,\alpha}:=\mathbb{G}_d\cap\mathbb{R}_\alpha$ for $\alpha\in\{\ge,>\}$. Let $\lambda\!\!\lambda_0$ denote Lebesgue measure, thus $\lambda\!\!\lambda_0=\lambda\!\!\lambda$, and $\lambda\!\!\lambda_1$ counting measure on $\mathbb{Z}$. Since $\mathbb{U}$ is concentrated on $\mathbb{Z}$ in the 1-arithmetic case, it is clear that convergence of $\mathbb{U}((t,t+h])$ in this case can generally take place only as $t\to\infty$ through $\mathbb{Z}$. This should be kept in mind for the following discussion.

Intuitively, the asymptotic behavior of $\mathbb{U}((t,t+h])$ should not depend on where the RP started, that is, on the delay distribution. In a standard model, this means that the limit of $\mathbb{U}_\lambda((t,t+h])$, if it exists, should be independent of $\lambda\in\mathscr{P}(\mathbb{G}_d)$. If we can find a delay distribution ν such that $\mathbb{U}_\nu$ may be computed explicitly, in particular $\mathbb{U}_\nu((t,t+h])$ for any $h>0$ and $t\to\infty$ through $\mathbb{G}_d$, then we may hope for being able to provide a coupling argument that shows $|\mathbb{U}_\lambda((t,t+h])-\mathbb{U}_\nu((t,t+h])|\to 0$ for any $\lambda\in\mathscr{P}(\mathbb{G}_{d,\ge})$ and thus confirm the afore-mentioned intuition. For a quick assessment of what the limit of $g(h)=\lim_{\mathbb{G}_d\ni t\to\infty}\mathbb{U}_\lambda((t,t+h])$ for any $h>0$ looks like, observe that

$$g(h_1+h_2) \;=\; \lim_{\mathbb{G}_d\ni t\to\infty}\mathbb{U}_\lambda((t,t+h_1]) \;+\; \lim_{\mathbb{G}_d\ni t\to\infty}\mathbb{U}_\lambda((t+h_1,t+h_1+h_2])$$

$$= \; g(h_1) + g(h_2) \quad \text{for all positive } h_1, h_2 \in \mathbb{G}_d$$

which shows that g must be linear on $\mathbb{G}_{d,\geq}$. In combination with the elementary renewal theorem, this entails that $g(h) = h/\mu$ for all $h \in \mathbb{G}_{d,>}$, thus $g \equiv 0$ if $\mu = \infty$.

Suppose now we are given a RP $(S_n)_{n\geq 0}$ in a standard model with *finite* drift μ and increment distribution F. The first thing to note is that $\mathbb{U}_\lambda = \lambda * \mathbb{U}_0$ satisfies the convolution equation

$$\mathbb{U}_\lambda \; = \; \lambda + F * \mathbb{U}_\lambda \quad \text{for any } \lambda \in \mathscr{P}(\mathbb{R}_\geq)$$

which in terms of the renewal function becomes a so-called *renewal equation* to be studied in more detail in Section 2.7, namely

$$\mathbb{U}_\lambda(t) \; = \; \lambda(t) \; + \; \int_{[0,t]} F(t-x) \, \mathbb{U}_\lambda(dx) \quad \text{for any } \lambda \in \mathscr{P}(\mathbb{R}_\geq) \tag{2.8}$$

The goal is to find a λ such that $\mathbb{U}_\lambda(t) = \mu^{-1}t$ for all $t \in \mathbb{R}_\geq$ (thus $\mathbb{U}_\lambda = \mu^{-1}\lambda\!\!\lambda_0^+$) and we will now do so by simply plugging the result into (2.8) and solving for $\lambda(t)$. Then, with $\overline{F} := 1 - F$,

$$\begin{aligned} \lambda(t) \; &= \; \frac{t}{\mu} \; - \; \frac{1}{\mu}\int_0^t F(t-x)\, dx \\ &= \; \frac{1}{\mu}\int_0^t \overline{F}(t-x)\, dx \; = \; \frac{1}{\mu}\int_0^t \overline{F}(x)\, dx \quad \text{for all } t \geq 0. \end{aligned}$$

We thus see that there is only one λ, now called F^s, that gives the desired property of $\mathbb{U}_\lambda$, viz.

$$F^s(t) \; := \; \frac{1}{\mu}\int_0^t \overline{F}(x)\, dx \; = \; \frac{1}{\mu}\int_0^t \mathbb{P}(X > x)\, dx \quad \text{for all } t \geq 0,$$

which is continuous and requires that μ is finite. To all those who prematurely lean back now let is be said that this is not yet the end of the story because there are questions still open, viz. “Is this really the answer we have been looking for if the RP is arithmetic?” and “What about the infinite mean case?”

If $(S_n)_{n\geq 0}$ is 1-arithmetic a continuous delay distribution appears to be inappropriate because it gives a continuous renewal measure. In fact, the stationary delay distribution F^s must now rather be concentrated on $\mathbb{N}$, but only give $\mathbb{U}_{F^s}(t) = \mu^{-1}t$ for $t \in \mathbb{N}_0$. By pursuing the same argument as above, but for $t \in \mathbb{N}_0$ only, one finds [☞ Problem 2.18] that F^s must satisfy

$$F^s(n) \; = \; \frac{1}{\mu}\sum_{k=0}^{n-1} \overline{F}(k) \; = \; \frac{1}{\mu}\sum_{k=1}^{n} \mathbb{P}(X \geq k) \quad \text{for all } n \in \mathbb{N}$$

as the unique solution among all distributions concentrated on $\mathbb{N}$. We summarize our findings as follows.

Proposition 2.15. *Let $(S_n)_{n\geq 0}$ be a RP in a standard model with finite drift μ and lattice-span $d \in \{0,1\}$. Define its* ***stationary delay distribution*** *F^s on $\mathbb{R}_>$ by*

$$F^s(t) \ := \ \begin{cases} \frac{1}{\mu}\int_0^t \mathbb{P}(X > x)\,dx, & \text{if } d = 0,\\ \frac{1}{\mu}\sum_{k=1}^{n(t)} \mathbb{P}(X \geq k), & \text{if } d = 1 \end{cases} \tag{2.9}$$

for $t \in \mathbb{R}_>$, where $n(t) := \lfloor t \rfloor = \sup\{n \in \mathbb{Z} : n \leq t\}$. Then $\mathbb{U}_{F^s} = \mu^{-1}\lambda\!\!\lambda_d^+$.

Now observe that the integral equation (2.8) remains valid if λ is any locally finite measure on $\mathbb{R}_\geq$ and $\mathbb{U}_\lambda$ is still defined as $\lambda * \mathbb{U}_0$. This follows because (2.8) is linear in λ. Hence, if we drop the normalization μ^{-1} in the definition of F^s, we obtain without further ado the following extension of the previous proposition.

Corollary 2.16. *Let $(S_n)_{n\geq 0}$ be a RP in a standard model with lattice-span $d \in \{0,1\}$. Define the locally finite measure ξ on $\mathbb{R}_>$ by*

$$\xi(t) \ := \ \begin{cases} \int_0^t \mathbb{P}(X > x)\,dx, & \text{if } d = 0,\\ \sum_{k=1}^{n(t)} \mathbb{P}(X \geq k), & \text{if } d = 1 \end{cases} \tag{2.10}$$

for $t \in \mathbb{R}_>$ and $n(t)$ as in Prop. 2.15. Then $\mathbb{U}_\xi = \lambda\!\!\lambda_d^+$.

2.4.3 Infinite mean case: restricting to finite horizons

There is no stationary delay distribution if $(S_n)_{n\geq 0}$ has infinite mean μ, but Cor. 2.16 helps us to provide a family of delay distributions for which stationarity still yields when restricting to finite horizons, that is to time sets $[0,a]$ for $a \in \mathbb{R}_>$. As a further ingredient we need the observation that the renewal epochs in $[0,a]$ of $(S_n)_{n\geq 0}$ and $(S_{a,n})_{n\geq 0}$, where $S_{a,n} := S_0 + \sum_{k=1}^n (X_k \wedge a)$, are the same. As a trivial consequence they also have the same renewal measure on $[0,a]$, whatever the delay distribution is. But by choosing the latter appropriately, we also have a domination result on (a,∞) as the next result shows.

Proposition 2.17. *Let $(S_n)_{n\geq 0}$ be a RP in a standard model with drift $\mu = \infty$ and lattice-span $d \in \{0,1\}$. With ξ given by* (2.10) *and for $a > 0$, define distributions F_a^s on $\mathbb{R}_>$ by*

$$F_a^s(t) \ := \ \frac{\xi(t\wedge a)}{\xi(a)} \quad \text{for } t \in \mathbb{R}_>. \tag{2.11}$$

Then, for all $a \in \mathbb{R}_>$, $\mathbb{U}_{F_a^s} \le \xi(a)^{-1}\lambda\!\!\lambda_d^+$ with equality holding on $[0,a]$.

Proof. Noting that F_a^s can be written as $F_a^s = \xi(a)^{-1}\xi - \lambda_a$, where $\lambda_a \in \mathscr{P}(\mathbb{R}_>)$ is given by

$$\lambda_a(t) \;:=\; \frac{\xi(t)-\xi(a\wedge t)}{\xi(a)} \;=\; \mathbf{1}_{(a,\infty)}(t)\frac{\xi(t)-\xi(a)}{\xi(a)} \quad \text{for all } t \in \mathbb{R}_>,$$

we infer with the help of Cor. 2.16 that

$$\mathbb{U}_{F_a^s} \;=\; \xi(a)^{-1}\mathbb{U}_\xi - \lambda_a * \mathbb{U}_0 \;\le\; \xi(a)^{-1}\mathbb{U}_\xi \;=\; \xi(a)^{-1}\lambda\!\!\lambda_d \quad \text{on } \mathbb{R}_>$$

as claimed. □

Problems

Problem 2.18. Given a 1-arithmetic RP $(S_n)_{n\ge 0}$ in a standard model with finite drift μ and increment distribution F, prove that F^s as defined in (2.9) for $d=1$ is the unique distribution on $\mathbb{N}$ such that $\mathbb{U}_{F^s} = \mu^{-1}\lambda\!\!\lambda_1^+$.

Problem 2.19. Under the assumptions of Prop. 2.15, let μ_p and μ_p^s for $p>0$ denote the p^{th} moment of F and F^s, respectively. Prove that

$$\mu_p^s \;:=\; \int t^p\, F^s(dt) \;=\; \begin{cases} \dfrac{\mu_{p+1}}{(p+1)\mu}, & \text{if } d=0,\\[2ex] \dfrac{1}{\mu}\,\mathbb{E}\left(\displaystyle\sum_{n=1}^X n^p\right), & \text{if } d=1. \end{cases} \tag{2.12}$$

and in the 1-arithmetic case furthermore

$$\frac{\mu_{p+1}}{(p+1)\mu} \;\le\; \frac{1}{\mu}\,\mathbb{E}\left(\sum_{n=1}^X n^p\right) \;\le\; \frac{\mu_{p+1}}{(p+1)\mu} + \frac{\mathbb{E}(X+1)^p}{\mu}. \tag{2.13}$$

Hence, $\mu_p^s < \infty$ iff $\mu_{p+1} < \infty$. Note also that $\mu^s = \mu_1^s$ satisfies

$$\mu^s \;=\; \frac{\mathbb{E}X(X+d)}{2\mu} \;=\; \frac{\mu_2}{2\mu} + \frac{d}{2} \;=\; \frac{\sigma^2+\mu^2}{2\mu} + \frac{d}{2}, \tag{2.14}$$

where $\sigma^2 := \mathbb{V}\mathrm{ar}X$.

2.5 Blackwell's renewal theorem

Blackwell's renewal theorem first obtained by ERDÖS, FELLER & POLLARD [46] for arithmetic RP's and by BLACKWELL [20] for nonarithmetic ones, may be rightfully called the mother of all deeper results in renewal theory. Not only it provides an answer to question (Q3) stated in the first section of this chapter on the expected number of renewals in a finite remote interval, but is also the simpler, yet equivalent version of the *key renewal theorem* discussed in the next section that allows us to determine the asymptotic behavior of many interesting quantities in applied stochastic models.

The following notation is introduced so as to facilitate a unified formulation of subsequent results for the arithmetic and the nonarithmetic case. For $d \in \{0,1\}$, define

$$d\text{-}\lim_{t\to\infty} f(t) \ := \ \begin{cases} \lim_{t\to\infty} f(t), & \text{if } d=0, \\ \lim_{n\to\infty} f(n), & \text{if } d=1. \end{cases}$$

Recall that $\lambda\!\!\lambda_0$ denotes Lebesgue measure on $\mathbb{G}_0 = \mathbb{R}$, while $\lambda\!\!\lambda_1$ is counting measure on $\mathbb{G}_1 = \mathbb{Z}$.

Theorem 2.20. [Blackwell's renewal theorem] *Let $(S_n)_{n\ge 0}$ be a RP in a standard model with lattice-span $d \in \{0,1\}$ and positive drift μ. Then*

$$d\text{-}\lim_{t\to\infty} \mathbb{U}_\lambda([t,t+h]) \ = \ \mu^{-1}\lambda\!\!\lambda_d([0,h]) \tag{2.15}$$

for all $h \ge 0$ and $\lambda \in \mathscr{P}(\mathbb{G}_{d,\ge})$, where $\mu^{-1} := 0$ if $\mu = \infty$.

The result, which actually extends to RW's with positive drift as will be seen later, has been proved by many authors and using various methods. The interested reader is referred to the monography [2, Ch. 3] for a detailed historical account. Here we will employ a coupling argument which to some extent forms a blend of the proofs given by LINDVALL [77], ATHREYA, MCDONALD & NEY [7], THORISSON [108] and finally by LINDVALL & ROGERS [78], all based on coupling as well. The proof is split into several steps given in separate subsections.

2.5.1 *First step of the proof : shaking off technicalities*

1st reduction: $S_0 = 0$.
It is no loss of generality to prove (2.15) for zero-delayed RP's only. Indeed, if S_0 has distribution $\lambda \in \mathscr{P}(\mathbb{G}_{d,\ge})$, then

$$\mathbb{U}_\lambda([t,t+h]) \ = \ \int \mathbb{U}_0([t-x,t-x+h])\,\lambda(dx)$$

together with $\sup_{t\in\mathbb{R}} \mathbb{U}_0([t,t+h]) \le \mathbb{U}_0([-h,h]) < \infty$ [☞ Lemma 2.14] implies by an appeal to the dominated convergence theorem that (2.15) is valid for $\mathbb{U}_\lambda$ if so for $\mathbb{U}_0$.

2nd reduction: $(S_n)_{n\ge 0}$ is completely d-arithmetic ($d \in \{0,1\}$).
The second reduction that will be useful hereafter is to assume that the increment distribution is completely d-arithmetic so that, by Lemma 2.12, its symmetrization has the the same lattice-span.

Lemma 2.21. *Let $(S_n)_{n\ge 0}$ be a SRP with lattice-span $d \in \{0,1\}$ and renewal measure $\mathbb{U}$. Let $(\rho_n)_{n\ge 0}$ a SRP independent of $(S_n)_{n\ge 0}$ and with geometric increments, viz. $\mathbb{P}(\rho_1 = n) = (1-\theta)^{n-1}\theta$ for some $\theta \in (0,1)$ and $n \in \mathbb{N}$. Then $(S_{\rho_n})_{n\ge 0}$ is a completely d-arithmetic SRP with renewal measure $\mathbb{U}^{(\rho)}$ satisfying $\mathbb{U}^{(\rho)} = (1-\theta)\delta_0 + \theta\,\mathbb{U}$.*

Proof. First of all, let $(I_n)_{n\ge 1}$ be a sequence of iid Bernoulli variables with parameter θ independent of $(S_n)_{n\ge 0}$. Each I_n may be interpreted as the outcome of a coin tossing performed at time n. Let $(J_n)_{n\ge 0}$ be the SRP associated with $(I_n)_{n\ge 1}$ and let $(\rho_n)_{n\ge 0}$ be the sequence of copy sums associated with $\rho = \rho_1 := \inf\{n \ge 1 : I_n = 1\}$. Then $(\rho_n)_{n\ge 0}$ satisfies the assumptions of the lemma, and one can easily verify [☞ Problem 2.24] that $(S_{\rho_n})_{n\ge 0}$ forms a SRP. Next observe that, for each $A \in \mathscr{B}(\mathbb{R})$,

$$\mathbb{U}^{(\rho)}(A) - \delta_0(A) \;=\; \mathbb{E}\left(\sum_{n\ge 1} I_n \mathbf{1}_A(S_n)\right) \;=\; \mathbb{E}I_1\left(\mathbb{U}(A) - \delta_0(A)\right)$$

which proves the relation between $\mathbb{U}^{(\rho)}$ and $\mathbb{U}$, for $\mathbb{E}I_1 = \theta$.

It remains to show that $(S_{\rho_n})_{n\ge 0}$ is completely d-arithmetic. Let $(S'_n, \rho'_n)_{n\ge 0}$ be an independent copy of $(S_n,\rho_n)_{n\ge 0}$ and put $\rho' := \rho'_1$. By Lemma 2.12, it suffices to show that the symmetrization $S_\rho - S'_{\rho'}$ is d-arithmetic. Since $c := d(S_\rho - S'_{\rho'}) \ge d$, we must only consider the case $c > 0$ and then show that $\mathbb{P}(X_1 \in c\mathbb{Z}) = 1$. But

$$1 \;=\; \mathbb{P}(S_\rho - S'_{\rho'} \in c\mathbb{Z}) \;=\; \sum_{m,n\ge 1} \theta^2 (1-\theta)^{m+n-2}\,\mathbb{P}(S_m - S'_n \in c\mathbb{Z})$$

clearly implies $\mathbb{P}(S_m - S'_n \in c\mathbb{Z}) = 1$ for all $m,n \in \mathbb{N}$. Hence

$$0 < \mathbb{P}(S_1 - S'_1 = 0) = \mathbb{P}(S_2 - S'_1 \in c\mathbb{Z}, S_1 - S'_1 = 0) = \mathbb{P}(X_1 \in c\mathbb{Z})\,\mathbb{P}(S_1 - S'_1 = 0)$$

giving $\mathbb{P}(X_1 \in c\mathbb{Z}) = 1$ as asserted. □

2.5.2 Setting up the stage: the coupling model

Based on the previous considerations, we now assume that $(S_n)_{n\ge 0}$ is a zero-delayed completely d-arithmetic RP with drift μ. As usual, the increment distribution is denoted by F and a generic copy of the increments by X. The starting point of the coupling construction is to consider this sequence together with a second one $(S_n')_{n\ge 0}$ such that the following conditions are satisfied:

(C1) $(S_n,S_n')_{n\ge 0}$ is a bivariate RP with iid increments (X_n,X_n'), $n\ge 1$.

(C2) $(S_n'-S_0')_{n\ge 0}\stackrel{d}{=}(S_n)_{n\ge 0}$ and thus $X'\stackrel{d}{=}X$.

(C3) $S_0'\stackrel{d}{=}F^s$ if $\mu<\infty$, and $S_0'\stackrel{d}{=}F_a^s$ for some $a>0$ if $\mu=\infty$.

Here F^s and F_a^s denote the stationary delay distribution and its truncated variant defined in (2.9) and (2.11), respectively. By the results in Section 2.5, the renewal measure $\mathbb{U}'$ of $(S_n')_{n\ge 0}$ satisfies $\mathbb{U}'([t,t+h])=\mu^{-1}\lambda\!\!\lambda_d([0,h])$ for all $t,h\in\mathbb{R}_>$ if $\mu<\infty$, and $\mathbb{U}'([t,t+h])\le\xi(a)^{-1}\lambda\!\!\lambda_d([0,h])$ for all $t,h\in\mathbb{R}_>$ if $\mu=\infty$ where $\xi(a)$ tends to ∞ as $a\to\infty$. Hence $\mathbb{U}'$ satisfies (2.15) in the finite mean case and does so approximately for sufficiently large a if $\mu=\infty$. The idea is now to construct a third RP $(S_n'')_{n\ge 0}$ from the given two which is a copy of $(S_n')_{n\ge 0}$ and such that S_n'' is equal or at least almost equal to S_n for all $n\ge T$, T an a.s. finite stopping time for $(S_n,S_n')_{n\ge 0}$, called *coupling time*. This entails that the *coupling process* $(S_n'')_{n\ge 0}$ has renewal measure $\mathbb{U}'$ while simultaneously being close to $\mathbb{U}$ on remote intervals because with high probability such intervals contain only renewal epochs S_n'' for $n\ge T$.

Having outlined the path towards the asserted result we must now complete the specification of the above bivariate model so as to facilitate a successful coupling. But the only unspecified component of the model is the joint distribution of (X,X') for which the following two alternatives will be considered:

(C4a) X and X' are independent or, equivalently, $(S_n)_{n\ge 0}$ and $(S_n')_{n\ge 0}$ are independent.

(C4b) $X'=Y\,\mathbf{1}_{[0,b]}(|X-Y|)+X\,\mathbf{1}_{(b,\infty)}(|X-Y|)$, where Y is an independent copy of X an b is chosen so large that $G_b:=\mathbb{P}(X-Y\in\cdot\,||X-Y|\le b)$ is d-arithmetic (and thus nontrivial).

The existence of b with $d(G_b)=d$ follows from the fact that $G:=\mathbb{P}(X-Y\in\cdot)$ is d-arithmetic together with $G_b\stackrel{w}{\to}G$.

Condition (C4a) is clearly simpler than (C4b) and will serve our needs in the finite mean case in which the symmetrization X_1-X_1' is integrable with mean zero and also d-arithmetic. Hence we infer from Thm. 2.22 below that $(S_n-S_n')_{n\ge 0}$ is (topologically) recurrent on $\mathbb{G}_d$.

On the other hand, if $\mu=\infty$, the difference of two independent X,X' fails to be integrable, while under (C4b) we have $X-X'=(X-Y)\mathbf{1}_{[-b,b]}(X-Y)$ which is again symmetric with mean zero and d-arithmetic by choice of b. Once again we hence infer the recurrence of the symmetric RW $(S_n-S_n')_{n\ge 0}$ on $\mathbb{G}_d$.

We close this subsection with the recurrence theorem for centered RW's needed here to guarantee successful coupling. The proof is omitted because it cannot be given shortly and is of no importance for our purposes. It may be found e.g. in [2, Ch. 2].

Theorem 2.22. *Any SRW $(S_n)_{n\ge 0}$ with lattice-span $d\in\{0,1\}$ and drift zero is (topologically) recurrent on $\mathbb{G}_d$, that is*

$$\mathbb{P}(|S_n-x|<\varepsilon \text{ infinitely often})=1$$

for any $x\in\mathbb{G}_d$ and $\varepsilon>0$.

2.5.3 Getting to the point: the coupling process

In the following suppose that (C1–3) and (C4a) are valid if $\mu<\infty$, while (C4a) is replaced with (C4b) if $\mu=\infty$. Fix any $\varepsilon>0$ if F is nonarithmetic, while $\varepsilon=0$ if F has lattice-span $d>0$. Since $(S_n-S_n')_{n\ge 0}$ is recurrent on $\mathbb{G}_d$ (recall that the delay distribution of S_0' is also concentrated on $\mathbb{G}_d$) we infer the a.s. finiteness of the *ε-coupling time*

$$T\ :=\ \inf\{n\ge 0:|S_n-S_n'|\le\varepsilon\}$$

and define the *coupling process* $(S_n'')_{n\ge 0}$ by

$$S_n''\ :=\ \begin{cases}S_n', & \text{if } n\le T,\\ S_n-(S_T-S_T'), & \text{if } n\ge T\end{cases}\quad \text{for } n\in\mathbb{N}_0, \tag{2.16}$$

which may also be stated as

$$S_n''\ :=\ \begin{cases}S_n', & \text{if } n\le T,\\ S_T'+\sum_{k=T+1}^n X_k, & \text{if } n> T\end{cases}\quad \text{for } n\in\mathbb{N}_0. \tag{2.17}$$

The subsequent lemma accounts for the intrinsic properties of this construction.

Lemma 2.23. *Under the stated assumptions, the following assertions hold true for the coupling process $(S_n'')_{n\ge 0}$:*

(a) $(S_n'')_{n\ge 0}\stackrel{d}{=}(S_n')_{n\ge 0}$.
(b) $|S_n''-S_n|\le\varepsilon$ *for all* $n\ge T$.

Proof. We only need to show (a) because (b) is obvious from the definition of the coupling process and the coupling time. Since T is a stopping time for the bivariate

RP $(S_n,S'_n)_{n\ge 0}$, Problem 2.25 shows that $X_{T+1},X_{T+2},...$ are iid with the same distribution as X and further independent of $T,(S_n,S'_n)_{0\le n\le T}$. But this easily seen to imply assertion (a), namely

$$\begin{aligned}
&\mathbb{P}(S''_0\in B_0,X''_j\in B_j \text{ for } 1\le j\le n)\\
&= \sum_{k=0}^{n}\mathbb{P}(T=k,S'_0\in B_0,X'_j\in B_j \text{ for } 1\le j\le k)\,\mathbb{P}(X_j\in B_j \text{ for } k<j\le n)\\
&\quad +\mathbb{P}(T>n,S'_0\in B_0,X'_j\in B_j \text{ for } 1\le j\le n)\\
&= \sum_{k=0}^{n}\mathbb{P}(T=k,S'_0\in B_0,X'_j\in B_j \text{ for } 1\le j\le k)\,\mathbb{P}(X'_j\in B_j \text{ for } k<j\le n)\\
&\quad +\mathbb{P}(T>n,S'_0\in B_0,X'_j\in B_j \text{ for } 1\le j\le n)\\
&= \mathbb{P}(S'_0\in B_0,X'_j\in B_j \text{ for } 1\le j\le n)
\end{aligned}$$

for all $n\in\mathbb{N}$ and $B_1,...,B_n\in\mathscr{B}(\mathbb{R}_\ge)$. □

Before moving on to the finishing argument, let us note that a coupling with a.s. finite 0-coupling time is called *exact coupling*, while we refer to an *ε-coupling* otherwise.

2.5.4 The final touch

As usual, let $N(I)$ denote the number of renewals S_n in I, and let $N''(I)$ be the corresponding variable for the coupling process $(S''_n)_{n\ge 0}$. Define further $N_k(I) := \sum_{j=0}^{k}\mathbf{1}_I(S_n)$ and $N''_k(I)$ in a similar manner. Fix any $h>0$, $\varepsilon\in(0,h/2)$, and put $I:=[0,h]$, $I_\varepsilon:=[\varepsilon,h-\varepsilon]$, and $I^\varepsilon:=[-\varepsilon,h+\varepsilon]$. The following proof of (2.15) focusses on the slightly more difficult nonarithmetic case, i.e. $d=0$ hereafter. We first treat the case $\mu<\infty$.

A. The finite mean case. By Lemma 2.23(a), $(S''_n)_{n\ge 0}$ has renewal measure $\mathbb{U}'$ which in turn equals $\mu^{-1}\lambda\!\!\lambda_0^+$ by our model assumption (C3). It follows from the coupling construction that

$$\{S''_n\in t+I_\varepsilon\}\ \subset\ \{S_n\in t+I\}\ \subset\ \{S''_n\in t+I^\varepsilon\}$$

for all $t\in\mathbb{R}_\ge$ and $n\ge T$. Consequently,

$$N''(t+I_\varepsilon)-N_T(t+I)\ \le\ N(t+I)\ \le\ N''(t+I^\varepsilon)+N_T(t+I)$$

and therefore, by taking expectations,

$$\mathbb{U}'(t+I_\varepsilon)-\mathbb{E}N_T(t+I)\ \le\ \mathbb{U}(t+I)\ \le\ \mathbb{U}'(t+I^\varepsilon)+\mathbb{E}N_T(t+I) \tag{2.18}$$

for all $t \in \mathbb{R}_{\geq}$. But $\mathbb{U}'(t+I_\varepsilon) = \mu^{-1}(h-2\varepsilon)$ and $\mathbb{U}'(t+I^\varepsilon) = \mu^{-1}(h+2\varepsilon)$ for all $t > \varepsilon$. Moreover, the uniform integrability of $\{N(t+I) : t \in \mathbb{R}\}$ [☞ Lemma 2.14] in combination with $N_T(t+I) \leq N(t+I)$ and $\lim_{t\to\infty} N_T(t+I) = 0$ a.s. entails

$$\lim_{t\to\infty} \mathbb{E}N_T(t+I) = 0.$$

Therefore, upon letting t tend to infinity in (2.18), we finally arrive at

$$\frac{h-2\varepsilon}{\mu} \;\leq\; \liminf_{t\to\infty} \mathbb{U}(t+I) \;\leq\; \limsup_{t\to\infty} \mathbb{U}(t+I) \;\leq\; \frac{h+2\varepsilon}{\mu}.$$

As ε can be made arbitrarily small, we have proved (2.15).

B. The infinite mean case. Here we have $\mathbb{U}' \leq \xi(a)^{-1}\lambda\!\!\!\lambda_0^+$ where a may be chosen so large that $\xi(a)^{-1} \leq \varepsilon$. Since validity of (2.18) remains unaffected by the drift assumption, we infer by just using the upper bound

$$\limsup_{t\to\infty} \mathbb{U}(t+I) \;\leq\; \xi(a)^{-1}(h+2\varepsilon) \;\leq\; \varepsilon(h+2\varepsilon)$$

and thus again the assertion, for ε can be made arbitrarily small. This completes our coupling proof of Blackwell's theorem. □

Problems

Problem 2.24. Let $(S_n)_{n\geq 0}$ and $(\rho_n)_{n\geq 0}$ be two independent SRP's such that $\rho_1, \rho_2, \ldots$ take values in $\mathbb{N}_0$.

(a) Prove that $(S_{\rho_n})_{n\geq 0}$ forms a SRP as well.
(b) Find the distribution of S_{ρ_1} if $(S_n)_{n\geq 0}$ has exponential increments and $\mathscr{L}(\rho_1) = \mathit{Geom}(\theta)$ for some $\theta \in (0,1)$.

Problem 2.25. Let $(S_n)_{n\geq 0}$ be a RW adapted to a filtration $(\mathscr{F}_n)_{n\geq 0}$ such that $\mathscr{F}_n$ is independent of $(X_k)_{k>n}$ for each $n \in \mathbb{N}_0$. Let T be an a.s. finite stopping time with respect to $(\mathscr{F}_n)_{n\geq 0}$.

(a) Prove that X_{T+1} is independent of $\mathscr{F}_T$ and $X_{T+1} \stackrel{d}{=} X_1$.
(b) Use (a) and an induction to infer that $(X_{T+n})_{n\geq 1}$ is a sequence of iid random variables independent of $\mathscr{F}_T$.

2.6 The key renewal theorem

Given a RP $(S_n)_{n\geq 0}$ in a standard model with drift μ and lattice-span d, the simple observation

$$\mathbb{U}_\lambda([t-h,t]) \;=\; \int \mathbf{1}_{[0,h]}(t-x)\,\mathbb{U}_\lambda(dx) \;=\; \mathbf{1}_{[0,h]} * \mathbb{U}_\lambda(t)$$

for all $t \in \mathbb{R}$, $h \in \mathbb{R}_>$ and $\lambda \in \mathscr{P}(\mathbb{R}_\geq)$ shows that the nontrivial part of Blackwell's renewal theorem may also be stated as

$$d\text{-}\lim_{t\to\infty} \mathbf{1}_{[0,h]} * \mathbb{U}_\lambda(t) \;=\; \frac{1}{\mu}\int \mathbf{1}_{[0,h]}\, d\lambda\!\!\lambda_d \tag{2.19}$$

for all $h \in \mathbb{R}_>$ and $\lambda \in \mathscr{P}(\mathbb{G}_{d,\geq})$, in other words, as a limiting result for convolutions of indicators of compact intervals with the renewal measure. This raises the question, supported further by numerous applications [☞ e.g. [2, Ch. 1]], to which class $\mathscr{R}$ of functions $g : \mathbb{R} \to \mathbb{R}$ an extension of (2.19) in the sense that

$$d\text{-}\lim_{t\to\infty} g * \mathbb{U}_\lambda(t) \;=\; \frac{1}{\mu}\int g\, d\lambda\!\!\lambda_d \quad \text{for all } g \in \mathscr{R} \tag{2.20}$$

is possible. Obviously, all finite linear combinations of indicators of compact intervals are elements of $\mathscr{R}$. By taking monotone limits of such step functions, one can further easily verify that $\mathscr{R}$ contains any g that vanishes outside a compact interval I and is Riemann integrable on I. On the other hand, in view of applications a restriction to functions with compact support appears to be undesirable and calls for appropriate conditions on g that are not too difficult to check in concrete examples. In the nonarithmetic case one would naturally hope for $\lambda\!\!\lambda_0$-integrability as being a sufficient condition, but unfortunately this is not generally true. The next subsection specifies the notion of *direct Riemann integrability*, first introduced and thus named by Feller [51], and provides also a discussion of necessary and sufficient conditions for this property to hold. Assertion (2.20) for functions g of this kind, called *key renewal theorem*, is proved in Subsection 2.6.2.

2.6.1 Direct Riemann integrability

Definition 2.26. Let g be a real-valued function on $\mathbb{R}$ and define, for $\delta > 0$ and $n \in \mathbb{Z}$,

$$\begin{aligned} I_{n,\delta} &:= (\delta n, \delta(n+1)], \\ m_{n,\delta} &:= \inf\{g(x) : x \in I_{n,\delta}\}, \quad M_{n,\delta} := \sup\{g(x) : x \in I_{n,\delta}\} \\ \underline{\sigma}(\delta) &:= \delta \sum_{n\in\mathbb{Z}} m_{n,\delta} \quad \text{and} \quad \overline{\sigma}(\delta) := \delta \sum_{n\in\mathbb{Z}} M_{n,\delta}. \end{aligned}$$

The function g is called *directly Riemann integrable (dRi)* if $\underline{\sigma}(\delta)$ and $\overline{\sigma}(\delta)$ are both absolutely convergent for all $\delta > 0$ and

$$\lim_{\delta\to 0}(\overline{\sigma}(\delta)-\underline{\sigma}(\delta)) \;=\; 0.$$

The definition reduces to ordinary Riemann integrability if the domain of g is only a compact interval instead of the whole line. In the case where $\int_{-\infty}^{\infty} g(x)\,dx$ may be defined as the limit of such ordinary Riemann integrals $\int_{-a}^{b} g(x)\,dx$ with a,b tending to infinity, the function g is called *improperly Riemann integrable*. An approximation of g by upper and lower step functions having integrals converging to a common value is then still only taken over compact intervals which are made bigger and bigger. However, in the above definition such an approximation is required to be possible *directly* over the whole line and therefore of a more restrictive type than improper Riemann integrability.

The following lemma, partly taken from [6, Prop. V.4.1], collects a whole bunch of necessary and sufficient criteria for direct Riemann integrability.

Proposition 2.27. *Let g be an arbitrary real-valued function on $\mathbb{R}$. Then the following two conditions are necessary for direct Riemann integrability:*

(dRi-1) *g is bounded and $\lambda\!\!\lambda_0$-a.e. continuous.*
(dRi-2) *g is $\lambda\!\!\lambda_d$-integrable for all $d \ge 0$.*

Conversely, any of the following conditions is sufficient for g to be dRi:

(dRi-3) *For some $\delta > 0$, $\underline{\sigma}(\delta)$ and $\overline{\sigma}(\delta)$ are absolutely convergent, and g satisfies (dRi-1).*
(dRi-4) *g has compact support and satisfies (dRi-1).*
(dRi-5) *g satisfies (dRi-1) and $f \le g \le h$ for dRi functions f,h.*
(dRi-6) *g vanishes on $\mathbb{R}_<$, is nonincreasing on $\mathbb{R}_\ge$ and $\lambda\!\!\lambda_0$-integrable.*
(dRi-7) *$g = g_1 - g_2$ for nondecreasing functions g_1, g_2 and $f \le g \le h$ for dRi functions f,h.*
(dRi-8) *g^+ and g^- are dRi.*

Proof. (a) Suppose that g is dRi. Then the absolute convergence of $\underline{\sigma}(1)$ and $\overline{\sigma}(1)$ ensures that g is bounded, for

$$\sup_{x\in\mathbb{R}}|g(x)| \;\le\; \sup_{n\in\mathbb{Z}}(|m_n^1|+|M_n^1|) \;<\; \infty.$$

That g must also be $\lambda\!\!\lambda_0$-a.e. continuous is a standard fact from Lebesgue integration theory but may also be quickly assessed as follows: If g fails to have this property then, with $g_*(x) := \liminf_{y\to x} g(y)$ and $g^*(x) := \limsup_{y\to x} g(y)$, we have

$$\alpha \;:=\; \lambda\!\!\lambda_0(\{g^* \ge g_* + \varepsilon\}) > 0 \quad \text{for some } \varepsilon > 0.$$

As $m_{n,\delta} \le g_*(x) \le g^*(x) \le M_{n,\delta}$ for all $x \in (n\delta,(n+1)\delta)$, $n \in \mathbb{Z}$ and $\delta > 0$, it follows that

$$\overline{\sigma}(\delta) - \underline{\sigma}(\delta) \ \ge \ \int (g^*(x) - g_*(x)) \, \lambda\!\!\lambda_0(dx) \ \ge \ \varepsilon\alpha \quad \text{for all } \delta > 0$$

which contradicts direct Riemann integrability. We have thus proved necessity of (dRi-1).

As for (dRi-2), it suffices to note that, with

$$\underline{\phi}(\delta) \ := \ \delta \sum_{n\in\mathbb{Z}} |m_{n,\delta}| \quad \text{and} \quad \delta\overline{\phi}(\delta) \ := \ \sum_{n\in\mathbb{Z}} |M_{n,\delta}|,$$

we have $\int |g(x)| \, \lambda\!\!\lambda_0(dx) \le \underline{\phi}(1) + \overline{\phi}(1)$ and $\int |g(x)| \, \lambda\!\!\lambda_d(dx) \le \underline{\phi}(d) + \overline{\phi}(d)$ for each $d > 0$.

(b) Turning to the sufficient criteria, put

$$g_\delta \ := \ \sum_{n\in\mathbb{Z}} m_{n,\delta} \mathbf{1}_{I_{n,\delta}} \quad \text{and} \quad g^\delta \ := \ \sum_{n\in\mathbb{Z}} M_{n,\delta} \mathbf{1}_{I_{n,\delta}} \quad \text{for } \delta > 0. \tag{2.21}$$

If (dRi-3) holds true, then $g_\delta \uparrow g$ and $g^\delta \downarrow g$ $\lambda\!\!\lambda_0$-a.e. as $\delta \downarrow 0$ by the $\lambda\!\!\lambda_0$-a.e. continuity of g. Hence the monotone convergence theorem implies (using $-\infty < \underline{\sigma}(\delta) \le \overline{\sigma}(\delta) < \infty$)

$$\underline{\sigma}(\delta) \ = \ \int g_\delta \, d\lambda\!\!\lambda_0 \uparrow \int g \, d\lambda\!\!\lambda_0 \quad \text{and} \quad \overline{\sigma}(\delta) \ = \ \int g^\delta \, d\lambda\!\!\lambda_0 \uparrow \int g \, d\lambda\!\!\lambda_0$$

proving that g is dRi.

Since each of (dRi-4) and (dRi-5) implies (dRi-3), there is nothing to prove under these conditions.

Assuming (dRi-6), the monotonicity of g on $\mathbb{R}_\ge$ gives

$$M_{n,\delta} = g(n\delta+) \quad \text{and} \quad m_{n,\delta} = g((n+1)\delta) \ge M_{n,\delta} \quad \text{for all } n \in \mathbb{N}_0, \ \delta > 0.$$

Consequently,

$$\begin{aligned} 0 \ \le \ \underline{\sigma}(\delta) \ &\le \ \int_0^\infty g(x) \, dx \ \le \ \overline{\sigma}(\delta) \\ &\le \ \delta g(0) + \underline{\sigma}(\delta) \ \le \ \int_0^\infty g(x) \, dx \ + \ \delta g(0) \ < \ \infty \end{aligned}$$

and therefore $\overline{\sigma}(\delta) - \underline{\sigma}(\delta) \le \delta g(0) \to 0$ as $\delta \to 0$.

Assuming (dRi-7) the monotonicity of g_1 and g_2 ensures that g has at most countably many discontinuities and is thus $\lambda\!\!\lambda_0$-a.e. continuous. g is also bounded because $f \le g \le h$ for dRi function f, h. Hence (dRi-5) holds true.

Finally assuming (dRi-8), note first that g^-, g^+ both satisfy (dRi-1) because this is true for g. Moreover,

$$0 \;\le\; g^{\pm} \;\le\; (g^{\delta})^{+}+(g_{\delta})^{-} \;\le\; \sum_{n\in\mathbb{Z}} \big(|M_{n,\delta}|+|m_{n,\delta}|\big)\, \mathbf{1}_{I_{n,\delta}} \quad \text{for all } \delta>0$$

whence g^{-},g^{+} both satisfy (dRi-5). □

For later purposes, we give one further criterion for direct Riemann integrability, but leave the simple proof to the reader [☞ Problem 2.35 and also Problem 4.8 for an extension].

Lemma 2.28. *Let g be a function on $\mathbb{R}$ that vanishes on $\mathbb{R}_<$ and is nondecreasing on $\mathbb{R}_\ge$. Then $g_\theta(x) := e^{\theta x}g(x)$ is dRi for any $\theta\in\mathbb{R}$ such that g_θ is $\lambda\!\!\lambda_0$-integrable.*

Given a measurable function $f:\mathbb{R}\to\mathbb{R}$ and a standard exponential random variable X, define the *exponential smoothing of f* by

$$\overline{f}(t) \;:=\; \int_{(-\infty,t]} e^{-(t-x)} f(x)\; \lambda\!\!\lambda_0(dx) \;=\; \mathbb{E}f(t-X), \quad t\in\mathbb{R}, \tag{2.22}$$

whenever this function is well-defined, which is obviously the case if $f\in L^1$. It then also has the same integral because

$$\int \overline{f}\; d\lambda\!\!\lambda_0 \;=\; \mathbb{E}\left(\int f(t-X)\; d\lambda\!\!\lambda_0\right) \;=\; \int f\; d\lambda\!\!\lambda_0. \tag{2.23}$$

In Chapter 4, we will use the fact that *exponential smoothing* of a L^1-function always provides us with a dRi function. This is stated as Lemma 2.30 after the following auxiliary result.

Lemma 2.29. *Suppose that $f\in L^1$ satisfies $f\ge 0$ and $f(t+\varepsilon)\ge r(\varepsilon)f(t)$ for all $t\in\mathbb{R}$, $\varepsilon>0$ and a function $r:\mathbb{R}_>\to\mathbb{R}_\ge$ satisfying* $\lim_{\varepsilon\downarrow 0} r(\varepsilon)=1$. *Then f is dRi.*

Proof. W.l.o.g. let r be nonincreasing. Then

$$r(\delta)f(n\delta)\le r(x-n\delta)f(n\delta)\le f(x)\le \frac{f((n+1)\delta)}{r((n+1)\delta-x)}\le \frac{f((n+1)\delta)}{r(\delta)}$$

for all $n\in\mathbb{Z}$, $\delta>0$ and $x\in I_{n,\delta}$ implies

$$r(\delta)f(n\delta)\le m_{n,\delta}\le \frac{1}{\delta}\int_{I_{n,\delta}} f\, d\lambda\!\!\lambda_0 \le M_{n,\delta}\le r(\delta)^{-1}f((n+1)\delta)$$

and therefore

$$\delta r(\delta)\sum_{n\in\mathbb{Z}} f(n\delta) \;\le\; \underline{\sigma}(\delta) \;\le\; \int f\,d\lambda\!\!\lambda_0 \;\le\; \overline{\sigma}(\delta) \;\le\; \frac{\delta}{r(\delta)}\sum_{n\in\mathbb{Z}} f(n\delta)$$

for any $\delta > 0$. Hence, $\delta\sum_{n\in\mathbb{Z}} f(n\delta)$ stays bounded as $\delta\downarrow 0$ and so

$$\overline{\sigma}(\delta)-\underline{\sigma}(\delta) \;\le\; \left(\frac{1}{r(\delta)} - r(\delta)\right)\delta\sum_{n\in\mathbb{Z}} f(n\delta) \;\stackrel{\delta\downarrow 0}{\longrightarrow}\; 0$$

as required. □

Lemma 2.30. *For each $f \in L^1$, its exponential smoothing $\overline{f}$ is dRi.*

Proof. By considering f^+ and f^-, we may assume w.l.o.g. that $f \ge 0$. Then

$$\begin{aligned}\overline{f}(t+\varepsilon) \;&=\; e^{-\varepsilon}\int_{(-\infty,t+\varepsilon]} e^{-(t-x)} f(x)\;\lambda\!\!\lambda_0(dx)\\ &\ge\; e^{-\varepsilon}\int_{(-\infty,t]} e^{-(t-x)} f(x)\;\lambda\!\!\lambda_0(dx) \;=\; e^{-\varepsilon}\overline{f}(t)\end{aligned}$$

for all $t \in \mathbb{R}$ and $\varepsilon > 0$, whence we may invoke the previous lemma to infer that f is dRi. □

2.6.2 The key renewal theorem: statement and proof

We are now ready to formulate and prove the announced extension of Blackwell's renewal theorem. In allusion to its eminent importance in applications Smith [101] called it *key renewal theorem*. The proof presented here is essentially due to FELLER [51].

Theorem 2.31. [Key renewal theorem] *Let $(S_n)_{n\ge 0}$ be a RP with drift μ, lattice-span $d \in \{0,1\}$ and renewal measure $\mathbb{U}$. Then*

$$d\text{-}\lim_{t\to\infty} g*\mathbb{U}(t) \;=\; \frac{1}{\mu}\int_{\mathbb{R}_\ge} g\;d\lambda\!\!\lambda_d \qquad (2.24)$$

for every dRi function $g : \mathbb{R}\to\mathbb{R}$ vanishing on the negative halfline.

Listing non- and d-arithmetic case separately, (2.24) takes the form

$$d\text{-}\lim_{t\to\infty} g*\mathbb{U}(t) \;=\; \frac{1}{\mu}\int_0^\infty g(x)\;dx \qquad (2.25)$$

if $d=0$ where the right-hand integral is meant as an improper Riemann integral. In the case $d=1$, we have accordingly

$$\lim_{n\to\infty} g*\mathbb{U}(n) \;=\; \frac{d}{\mu}\sum_{n\ge 0} g(n) \tag{2.26}$$

and, furthermore, for any $a\in\mathbb{R}$,

$$d\text{-}\lim_{t\to\infty} g*\mathbb{U}(nd+a) \;=\; \frac{1}{\mu}\sum_{n\ge 0} g(n+a), \tag{2.27}$$

because $g(\cdot+a)$ is clearly dRi as well.

Proof. We restrict ourselves to the more difficult nonarithmetic case. Given a dRi function g vanishing on $\mathbb{R}_<$, let g_δ, g^δ be as in (2.21) for $\delta>0$. Plainly, these functions vanish on $\mathbb{R}_<$ as well, so that we have

$$g_\delta \le g \le g^\delta, \quad \underline{\sigma}(\delta) = \int_0^\infty g_\delta(x)\,dx \quad \text{and} \quad \overline{\sigma}(\delta) = \int_0^\infty g^\delta(x)\,dx.$$

Fix any $\delta\in(0,1)$ and $m\in\mathbb{N}$ large enough such that $\sum_{n>m}|M_{n,\delta}|<\delta$. Then, using inequality (2.7), we infer

$$g^\delta*\mathbb{U}(t) \;=\; \sum_{n\ge 0} M_{n,\delta}\mathbb{U}(t-I_{n,\delta}) \;\le\; \sum_{n=0}^{m} M_{n,\delta}\,\mathbb{U}(t-n\delta-I_{0,\delta}) \;+\; \delta\,\mathbb{U}(1)$$

and therefore with Blackwell's theorem

$$\begin{aligned}
\limsup_{t\to\infty} g^\delta*\mathbb{U}(t) \;&\le\; \sum_{n=0}^{m} M_{n,\delta}\lim_{t\to\infty}\mathbb{U}(t-n\delta-I_{0,\delta}) \;+\; \delta\,\mathbb{U}(1)\\
&=\; \frac{\delta}{\mu}\sum_{n=0}^{m} M_{n,\delta} \;+\; \delta\,\mathbb{U}(1)\\
&\le\; \frac{1}{\mu}\int_0^\infty g^\delta(x)\,dx \;+\; \frac{\delta^2}{\mu} \;+\; \delta\,\mathbb{U}(1)\\
&=\; \frac{1}{\mu}\overline{\sigma}(\delta) \;+\; \frac{\delta^2}{\mu} \;+\; \delta\,\mathbb{U}(1).
\end{aligned} \tag{2.28}$$

Consequently, as $g*\mathbb{U}\le g^\delta*\mathbb{U}$ for all $\delta>0$,

$$\limsup_{t\to\infty} g*\mathbb{U}(t) \;\le\; \lim_{\delta\downarrow 0}\limsup_{t\to\infty} g^\delta*\mathbb{U}(t) \;\le\; \frac{1}{\mu}\int_0^\infty g(x)\,dx.$$

Replace g with $-g$ in the above estimation to obtain

$$\liminf_{t\to\infty} g*\mathbb{U}(t) \;\ge\; \frac{1}{\mu}\int_0^\infty g(x)\,dx.$$

This completes the proof of (2.24). □

Remark 2.32. In the 1-arithmetic and thus discrete case, the convolution $g*\mathbb{U}$ may actually be considered as a function on the discrete group $\mathbb{Z}$ and thus requires g to be considered on this set only which reduces it to a sequence $(g_n)_{n\in\mathbb{Z}}$. Doing so merely absolute summability, i.e. $\sum_{n\in\mathbb{Z}}|g_n|<\infty$, is needed instead of direct Riemann integrability. With this observation the result reduces to a straightforward consequence of Blackwell's renewal theorem and explains that much less attention has been paid to it in the literature.

Remark 2.33. The following counterexample shows that in the nonarithmetic case $\lambda\!\!\!\lambda_0$-integrability of g does not suffice to ensure (2.24). Consider a distribution F on $\mathbb{R}_\geq$ with positive mean $\mu=\int x\,F(dx)$ and renewal measure $\mathbb{U}=\sum_{n\geq 0}F^{*n}$. The function $g:=\sum_{n\geq 1}n^{1/2}\mathbf{1}_{[n.n+n^{-2})}$ is obviously $\lambda\!\!\!\lambda_0$-integrable, but

$$g*\mathbb{U}(n) \;=\; \sum_{k\geq 0}g*F^{*k}(n) \;\geq\; g*F^{*0}(n) \;=\; g(n) \;=\; n^{1/2}$$

diverges to ∞ as $n\to\infty$. Here the atom at 0, which any renewal measure of a SRP possesses, already suffices to demonstrate that $g(x)$ must not have unbounded oscillations as $x\to\infty$. But there are also examples of renewal measures with no atom at 0 (thus pertaining to a delayed RP) such that the key renewal theorem fails to hold for $\lambda\!\!\!\lambda_0$-integrable g. FELLER [51, p. 368] provides an example of a $\lambda\!\!\!\lambda_0$-continuous distribution F with finite positive mean such that $\mathbb{U}=\sum_{n\geq 1}F^{*n}$ satisfies $\limsup_{t\to\infty}g*\mathbb{U}(t)=\infty$ for some $\lambda\!\!\!\lambda_0$-integrable g.

Example 2.34. **[Forward and backward recurrence times]** Let $(S_n)_{n\geq 0}$ be a SRP with increment distribution F, lattice-span $d\in\{0,1\}$ and finite drift μ. For $t\geq 0$, let $\tau(t):=\inf\{n\geq 0:S_n>t\}$ denote the *first passage time* beyond level t and consider the first renewal epoch after t and the last renewal epoch before t, more precisely

$$R(t):=S_{\tau(t)}-t \quad\text{and}\quad \widehat{R}(t):=t-S_{\tau(t)-1}$$

called *forward* and *backward recurrence time*, respectively. Other names for $R(t)$, depending on the context in which it is discussed, are *overshoot, excess (over the boundary)* or *residual waiting time*. Other names for $\widehat{R}(t)$ are *age* and *spent waiting time*. We are interested in the asymptotic behavior of $R(t)$ and $\widehat{R}(t)$. It follows by a standard renewal argument that

$$\begin{aligned} \mathbb{P}(R(t)>r) \;&=\; \int_{[0,t]}\mathbb{P}(X>t+r-x)\;\mathbb{U}(dx)\\ \text{and}\quad \mathbb{P}(\widehat{R}(t)>r) \;&=\; \int_{[0,t]}\mathbb{P}(X>t-x)\,\mathbf{1}_{(r,\infty)}(t-x)\;\mathbb{U}(dx)\end{aligned}$$

for all $r,t\geq 0$. To both right-hand expressions the key renewal theorem applies and yields that

$$R(t)\stackrel{d}{\to}R(\infty)\quad\text{and}\quad \widehat{R}(t)+d\stackrel{d}{\to}R(\infty) \tag{2.29}$$

as $t \to \infty$ (through $\mathbb{Z}$ if $d = 1$), where $\mathscr{L}(R(\infty)) = F^s$. Details are left as an exercise to the reader [☞ Problem 2.36].

Problems

Problem 2.35. Prove Lemma 2.28.

Problem 2.36. Under the assumptions of Example 2.34, prove (2.29) by filling in the details of the argument outlined there. Then proceed in a similar manner to find the asymptotic joint distribution of $(R(t), \widehat{R}(t))$ and of $X_{\tau(t)} = R(t) + \widehat{R}(t)$ as $t \to \infty$ (through $\mathbb{Z}$ if $d = 1$). Do the results persist if the distribution of S_0 is arbitrarily chosen from $\mathscr{P}(\mathbb{G}_{d,\geq})$?

Problem 2.37. Still in the situation of Example 2.34, suppose that $F = Exp(\theta)$ for some $\theta > 0$. Compute the asymptotic joint distribution of $(R(t), \widehat{R}(t))$ and of $X_{\tau(t)} = R(t) + \widehat{R}(t)$ in this case.

2.7 The renewal equation

Almost every renewal quantity may be described as the solution to a convolution equation of the general form

$$\Psi \ = \ \psi \ + \ \Psi * Q, \tag{2.30}$$

where Q is a given locally finite measure and ψ a given locally bounded function on $\mathbb{R}_\geq$ (standard case) or $\mathbb{R}$ (general case). For reasons that will become apparent soon it is called *renewal equation*. If $\psi = 0$, then (2.30) is also a well-known object in harmonic analysis where its solutions are called *Q-harmonic functions*. It has been studied in the more general framework of Radon measures on separable locally compact Abelian groups by CHOQUET & DENY [27] and is therefore also known as the *Choquet-Deny equation*. Here we will focus on the standard case where functions and measures vanish on the negative halfline. Eq. (2.30) then takes the form

$$\Psi(x) \ = \ \psi(x) \ + \ \int_{[0,x]} \Psi(x-y)\, Q(dy), \quad x \in \mathbb{R}_\geq, \tag{2.31}$$

and is called *standard renewal equation* because it is the one encountered in most applications. Regarding the total mass of Q, a renewal equation is called *defective* if $\|Q\| < 1$, *proper* if $\|Q\| = 1$, and *excessive* if $\|Q\| > 1$.

2.7.1 Getting started

Some further notation is needed hereafter and therefore introduced first. Recall that Q is assumed to be locally finite, thus $Q(t) = Q([0,t]) < \infty$ for all $t \in \mathbb{R}_\geq$. We denote its mean value by $\mu(Q)$ and its mgf by ϕ_Q, that is

$$\mu(Q) \ := \ \int_{\mathbb{R}_\geq} x\, Q(dx)$$

and

$$\phi_Q(\theta) \ := \ \int_{\mathbb{R}_\geq} e^{\theta x}\, Q(dx).$$

The latter function is nondecreasing and convex on its natural domain

$$\mathbb{D}_Q \ := \ \{\theta \in \mathbb{R} : \phi_Q(\theta) < \infty\}$$

for which one of the four alternatives

$$\mathbb{D}_Q \ = \ \emptyset,\ (-\infty,\theta^*),\ (-\infty,\theta^*],\ \text{or}\ \mathbb{R}$$

with $\theta^* \in \mathbb{R}$ must hold. If $\mathbb{D}_Q$ has interior points, then ϕ_Q is infinitely often differentiable on $\operatorname{int}(\mathbb{D}_Q)$ with n^{th} derivative given by

$$\phi_Q^{(n)}(\theta) \ = \ \int_{\mathbb{R}_\geq} x^n e^{\theta x}\, Q(dx) \quad \text{for all } n \in \mathbb{N}.$$

In the following we will focus on measures Q on $\mathbb{R}_\geq$, called *admissible*, for which $\mu(Q) > 0$, $Q(0) < 1$ and $\mathbb{D}_Q \neq \emptyset$ holds true. Note that the last condition is particularly satisfied if $\|Q\| < \infty$ or, more generally, Q is uniformly locally bounded, i.e.

$$\sup_{t\geq 0} Q([t,t+1]) \ < \ \infty.$$

Moreover, ϕ_Q is increasing and strictly convex for such Q. Hence, there exists at most one value $\vartheta \in \mathbb{D}_Q$ such that $\phi_Q(\vartheta) = 1$. It is called the *characteristic exponent of Q* hereafter.

Let $\mathbb{U} := \sum_{n\geq 0} Q^{*n}$ with $Q^{*0} := \delta_0$ be the renewal measure of Q. Put further

$$Q_\theta(dx) \ := \ e^{\theta x}\, Q(dx)$$

again a locally finite measure for any $\theta \in \mathbb{R}$, and let $\mathbb{U}_\theta$ be its renewal measure.[1] Then

$$\mathbb{U}_\theta(dx) \ = \ \sum_{n\geq 0} Q_\theta^{*n}(dx) \ = \ \sum_{n\geq 0} e^{\theta x}\, Q^{*n}(dx) \ = \ e^{\theta x}\, \mathbb{U}(dx). \qquad (2.32)$$

[1] The reader will notice here a notational conflict because in all previous and almost all subsequent sections $\mathbb{U}_\theta = \delta_\theta * \mathbb{U}$. On the other hand, whenever the current definition is meant, this will be clearly pointed out and should therefore not lead to any confusion.

Moreover, $\phi_{Q_\theta} = \phi_Q(\cdot + \theta)$ and $\phi_{\mathbb{U}_\theta} = \phi_{\mathbb{U}}(\cdot + \theta)$.

Lemma 2.38. *Given an admissible measure Q on $\mathbb{R}_\geq$, the following assertions hold true for any $\theta \in \mathbb{R}$:*

(a) Q_θ^{*n} *is admissible for all $n \in \mathbb{N}$.*
(b) $\mathbb{U}_\theta$ *is locally finite, that is $\mathbb{U}_\theta(t) < \infty$ for all $t \in \mathbb{R}_\geq$.*
(c) $\lim_{n\to\infty} Q^{*n}(t) = 0$ *for all $t \in \mathbb{R}_\geq$.*

Proof. Assertion (a) is trivial when noting that $Q_\theta^{*n}(0) = Q^{*n}(0) = Q(0)^n$ for all $\theta \in \mathbb{R}$ and $n \in \mathbb{N}$. As for (b), it clearly suffices to show that $\mathbb{U}_\theta$ is locally finite for *some* $\theta \in \mathbb{R}$. To this end note that $\mathbb{D}_Q \neq \emptyset$ implies $\phi_Q(t) \to 0$ as $t \to -\infty$ and thus the existence of $\theta \in \mathbb{R}$ such that $\|Q_\theta\| = \phi_Q(\theta) < 1$. Hence $\mathbb{U}_\theta$ is the renewal measure of the defective probability measure Q_θ and thus finite, for

$$\|\mathbb{U}_\theta\| \;=\; \sum_{n\geq 0} \|Q_\theta^{*n}\| \;=\; \sum_{n\geq 0} \|Q_\theta\|^n \;=\; \frac{1}{1-\phi_Q(\theta)} \;<\; \infty.$$

Finally, the local finiteness of $\mathbb{U} = \mathbb{U}_0$ gives $\mathbb{U}(t) = \sum_{n\geq 0} Q^{*n}(t) < \infty$ for all $t \in \mathbb{R}_\geq$ from which (c) directly follows. ☐

2.7.2 Existence and uniqueness of a locally bounded solution

We are now ready to prove the fundamental theorem about existence and uniqueness of solutions in the standard case (2.31) under the assumption that the measure Q is regular and the function ψ is *locally bounded* on $\mathbb{R}_\geq$, i.e.

$$\sup_{x\in[0,t]} |\psi(x)| \;<\; \infty \quad \text{for all } t \in \mathbb{R}_\geq.$$

Before stating the result let us note that n-fold iteration of equation (2.31) leads to

$$\Psi(x) \;=\; \sum_{k=0}^{n} \psi * Q^{*k}(x) \;+\; \Psi * Q^{*(n+1)}(x)$$

which in view of part (c) of the previous lemma suggests that $\Psi = \psi * \mathbb{U}$ forms the unique solution of (2.31).

Theorem 2.39. *Let Q be an admissible measure on $\mathbb{R}_\geq$ and $\psi : \mathbb{R}_\geq \to \mathbb{R}$ a locally bounded function. Then there exists a unique locally bounded solution Ψ of the renewal equation* (2.31)*, viz.*

$$\Psi(x) \;=\; \psi * \mathbb{U}(x) \;=\; \int_{[0,x]} \psi(x-y)\,\mathbb{U}(dy), \quad x \in \mathbb{R}_{\geq}$$

where $\mathbb{U}$ denotes the renewal measure of Q. Moreover, Ψ is nondecreasing if the same holds true for ψ.

Proof. Since $\mathbb{U}$ is locally finite, the local boundedness of ψ entails the same for the function $\psi * \mathbb{U}$, and the latter function satisfies (2.31) as

$$\psi * \mathbb{U} \;=\; \psi * \delta_0 \;+\; \left(\sum_{n\geq 1} \psi * Q^{*(n-1)}\right) * Q \;=\; \psi \;+\; (\psi * \mathbb{U}) * Q.$$

Moreover, $\psi * \mathbb{U}$ is nondecreasing if ψ has this property.

Turning to uniqueness, suppose we have two locally bounded solutions Ψ_1, Ψ_2 of (2.31). Then its difference Δ, say, satisfies the very same equation with $\psi \equiv 0$, that is $\Delta = \Delta * Q$. By iteration,

$$\Delta \;=\; \Delta * Q^{*n} \quad \text{for all } n \in \mathbb{N}.$$

Since Δ is locally bounded, it follows upon setting $\|\Delta\|_{x,\infty} := \sup_{y\in[0,x]} |\Delta(x)|$ and an appeal to Lemma 2.38(c) that

$$|\Delta(x)| \;=\; \lim_{n\to\infty} |\Delta * Q^{*n}(x)| \;\leq\; \|\Delta\|_{x,\infty} \lim_{n\to\infty} Q^{*n}(x) \;=\; 0 \quad \text{for all } x \in \mathbb{R}_{\geq}$$

which proves $\Psi_1 = \Psi_2$. □

The following version of the Choquet-Deny lemma is a direct consequence of the previous result.

Corollary 2.40. *If Q is an admissible measure on $\mathbb{R}_{\geq}$, then $\Psi \equiv 0$ is the only locally bounded solution to the Choquet-Deny equation $\Psi = \Psi * Q$.*

2.7.3 Asymptotics

Continuing with a study of the asymptotic behavior of solutions $\psi * \mathbb{U}$ a distinction of the cases $\|Q\| < 1$, $\mathbb{Q}\| = 1$, and $\|Q\| > 1$ is required. Put $I_d := \{0\}$ if $d = 0$, and $I_d := [0,d)$ if $d > 0$.

We begin with the defective case when $\phi_Q(0) = \|Q\| < 1$ and thus $\mathbb{U}$ is finite with total mass $\|U\| = (1-\phi_Q(0))^{-1}$.

Theorem 2.41. *Given a defective renewal equation of the form* (2.31) *with locally bounded ψ such that $\psi(\infty) := \lim_{x\to\infty}\psi(x) \in [-\infty,\infty]$ exists, the same holds true for $\Psi = \psi * \mathbb{U}$, namely*

$$\Psi(\infty) \;=\; \frac{\psi(\infty)}{1-\phi_Q(0)}.$$

Proof. If $\psi(\infty) = \infty$, then the local boundedness of ψ implies $\inf_{x\ge 0}\psi(x) > -\infty$. Consequently, by an appeal to Fatou's lemma,

$$\begin{aligned}\liminf_{x\to\infty}\Psi(x) \;&=\; \liminf_{x\to\infty}\int_{[0,x]}\psi(x-y)\;\mathbb{U}(dy)\\ &\ge\; \int_{\mathbb{R}_\ge}\liminf_{x\to\infty}\mathbf{1}_{[0,x]}(y)\,\psi(x-y)\;\mathbb{U}(dy)\\ &=\; \psi(\infty)\,\|U\| \;=\; \infty.\end{aligned}$$

A similar argument shows $\limsup_{x\to\infty}\Psi(x) = -\infty$ if $\psi(\infty) = -\infty$. But if $\psi(\infty)$ is finite then ψ is necessarily bounded and we obtain by the dominated convergence theorem that

$$\lim_{x\to\infty}\Psi(x) \;=\; \int_{\mathbb{R}_\ge}\lim_{x\to\infty}\mathbf{1}_{[0,x]}(y)\,\psi(x-y)\;\mathbb{U}(dy) \;=\; \psi(\infty)\,\|\mathbb{U}\| \;=\; \frac{\psi(\infty)}{1-\phi_Q(0)}$$

as claimed. □

Turning to the case where $Q \ne \delta_0$ is a probability distribution on $\mathbb{R}_\ge$ (proper case) a statement about the asymptotic behavior of solutions $\psi * \mathbb{U}$ can be directly deduced with the help of the key renewal theorem 2.31.

Theorem 2.42. *Given a proper renewal equation of the form* (2.31) *with dRi function ψ, it follows for all $a \in I_d$ that*

$$d\text{-}\lim_{x\to\infty}\Psi(x+a) \;=\; \frac{1}{\mu(Q)}\int_{\mathbb{R}_\ge}\psi(x+a)\;\lambda\!\!\lambda_d(dx), \tag{2.33}$$

where d denotes the lattice-span of Q.

Our further investigations will rely on the subsequent lemma which shows that a renewal equation preserves its structure under the exponential transform $Q(dx) \mapsto Q_\theta(dx) = e^{\theta x}Q(dx)$ for any $\theta \in \mathbb{R}$. Plainly, Q_θ is a probability measure iff θ equals the characteristic exponent of Q. Given a function ψ on $\mathbb{R}_\ge$, put

$$\psi_\theta(x) \;:=\; e^{\theta x}\psi(x), \quad x \in \mathbb{R}_\ge$$

for any $\theta \in \mathbb{R}$.

Lemma 2.43. *Let Q be an admissible measure on $\mathbb{R}_{\geq}$, $\psi : \mathbb{R}_{\geq} \to \mathbb{R}$ a locally bounded function and Ψ any solution to the pertinent renewal equation* (2.31). *Then, for any $\theta \in \mathbb{R}$, Ψ_θ forms a solution to* (2.31) *for the pair (ψ_θ, Q_θ), i.e.*

$$\Psi_\theta \;=\; \psi_\theta + \Psi_\theta * Q_\theta. \tag{2.34}$$

Moreover, if $\Psi = \psi * \mathbb{U}$, then $\Psi_\theta = \psi_\theta * \mathbb{U}_{Q_\theta}$ is the unique locally bounded solution to (2.34).

Proof. For the first assertion, it suffices to note that $\Psi = \psi + \Psi * Q$ obviously implies (2.34), for

$$e^{\theta x}\Psi(x) \;=\; e^{\theta x}\psi(x) \;+\; \int_{[0,x]} e^{\theta(x-y)}\,\Psi(x-y)\,e^{\theta y}\,Q(dy)$$

for all $x \in \mathbb{R}_{\geq}$. Since Q_θ is admissible for any $\theta \in \mathbb{R}$, the second assertion follows by Thm. 2.39. □

With the help of this lemma we are now able to derive the following general result on the asymptotic behavior of $\psi * \mathbb{U}$ for a standard renewal equation of the form (2.31). It covers the excessive as well as the defective case.

Theorem 2.44. *Given a renewal equation of the form* (2.31) *with admissible Q with lattice-span d and locally bounded function ψ, the following assertions hold true for its unique locally bounded solution $\Psi = \psi * \mathbb{U}$:*

(a) If $\theta \in \mathbb{R}$ is such that $\|Q_\theta\| < 1$ and $\psi_\theta(\infty)$ exists, then

$$\lim_{x\to\infty} e^{\theta x}\,\Psi(x) \;=\; \frac{\psi_\theta(\infty)}{1-\phi_Q(\theta)} \tag{2.35}$$

(b) If Q possesses a characteristic exponent ϑ, then

$$d\text{-}\lim_{x\to\infty} e^{\vartheta x}\,\Psi(x+a) \;=\; \frac{1}{\mu(Q_\vartheta)}\int_{\mathbb{R}_{\geq}} e^{\vartheta x}\,\psi(x+a)\,\lambda\!\!\lambda_d(dx) \tag{2.36}$$

for all $a \in I_d$ if ψ_ϑ is dRi.

Proof. All assertions are direct consequences of the previous results. □

Remark 2.45. If $1 < \|Q\| < \infty$ in the previous theorem, then $\mathbb{D}_Q \supset (-\infty, 0]$ and the continuity ϕ_Q together with $\lim_{\theta\to-\infty}\phi_Q(\theta) = 0$ *always* ensures the existence of

$\vartheta < 0$ with $\phi_Q(\vartheta) = \|Q_\vartheta\| = 1$ by the intermediate value theorem. On the other hand, if Q is an infinite admissible measure, then it is possible that $\phi_Q(\theta) < 1$ *for all* $\theta \in \mathbb{D}_Q$.

There is yet another situation uncovered so far where further information on the asymptotic behavior of $\psi * \mathbb{U}$ may be obtained. Suppose that, for some $\theta \in \mathbb{R}$, $\psi_\theta(\infty)$ exists but is nonzero and that Q_θ is defective. Then Thm. 2.41 provides us with

$$\Psi_\theta(\infty) \;=\; \lim_{x\to\infty} e^{\theta x}\Psi(x) \;=\; \frac{\psi_\theta(\infty)}{1-\phi_Q(\theta)} \;\neq\; 0$$

which in turn raises the question whether the rate of convergence of $\Psi_\theta(x)$ to $\Psi_\theta(\infty)$ may be studied by finding a renewal equation satisfied by the difference $\Psi_\theta^0 := \Psi_\theta(\infty) - \Psi_\theta$. An answer is given by the next theorem for which $\theta = 0$ is assumed without loss of generality. For $d \in \mathbb{R}_\geq$ and $\theta \in \mathbb{R}$, let us define

$$e(d,\theta) \;:=\; \begin{cases} \theta, & \text{if } d = 0,\\ (e^{\theta d}-1)/d, & \text{if } d > 0. \end{cases} \tag{2.37}$$

which is a continuous function on $\mathbb{R}_\geq \times \mathbb{R}$.

Theorem 2.46. *Given a defective renewal equation of the form* (2.31) *with locally bounded* ψ *such that* $\psi(\infty) \neq 0$*, it follows that* $\Psi^0 := \Psi(\infty) - \Psi$ *forms the unique locally bounded solution to the renewal equation* $\Psi^0 = \widehat{\psi} + \Psi^0 * Q$ *with*

$$\widehat{\psi}(x) \;:=\; \psi^0(x) + \psi(\infty)\,\frac{Q((x,\infty))}{1-\phi_Q(0)}, \quad x \in \mathbb{R}.$$

Furthermore, if Q *has characteristic exponent* ϑ *(necessarily positive) and lattice-span* d*, then*

$$d\text{-}\lim_{x\to\infty} \Psi_\vartheta^0(x+a) \;=\; \frac{e^{\vartheta a}}{\mu(Q_\vartheta)}\left(\frac{\psi(\infty)}{e(d,\vartheta)} + \int_{\mathbb{R}_\geq} e^{\vartheta y}\,\psi^0(y+a)\;\lambda\!\!\lambda_d(dy)\right) \tag{2.38}$$

for any $a \in I_d$ *provided that* $\widehat{\psi}_\vartheta$ *is dRi.*

Proof. A combination of

$$\Psi(\infty) - \int_{[0,x]} \Psi(\infty)\,Q(dx) \;=\; \Psi(\infty)\,Q((x,\infty)) \;=\; \frac{\psi(\infty)\,Q((x,\infty))}{1-\phi_Q(0)} \;=\; \widehat{\psi}(x) - \psi^0(x)$$

and $\Psi = \psi + \Psi * Q$ shows the asserted renewal equation for Ψ^0. By the previous results, we then infer under the stated conditions on $\widehat{\psi}$ and Q that

$$d\text{-}\lim_{x\to\infty} \Psi_\vartheta^0(x+a) \;=\; \frac{1}{\mu(Q_\vartheta)} \int_{\mathbb{R}_\geq} \widehat{\psi}_\vartheta(y+a)\;\lambda\!\!\lambda_d(dy) \quad \text{for any } a \in [0,d).$$

Hence it remains to verify that the right-hand side equals the right-hand side of (2.38).

Let us first consider the case $d=0$: Using $\phi_Q(\vartheta)=1$, we find that

$$\begin{aligned}\int_{\mathbb{R}_\geq} e^{\vartheta y}(\widehat{\psi}(y)-\psi^0(y))\,\lambda\!\!\lambda_0(dy) \ &= \ \frac{\psi(\infty)}{1-\phi_Q(0)}\int_{\mathbb{R}_\geq} e^{\vartheta y}Q((y,\infty))\,\lambda\!\!\lambda_0(dy)\\ &= \ \frac{\psi(\infty)}{\vartheta(1-\phi_Q(0))}\int_{\mathbb{R}_\geq}(e^{\vartheta y}-1)\,Q(dy) \ = \ \frac{\psi(\infty)}{\vartheta}\end{aligned}$$

which is the desired result.

If $d>0$ and $a\in[0,d)$, use $Q((y+a,\infty))=Q((y,\infty))$ for any $y\in d\mathbb{Z}$ to see that

$$\begin{aligned}\int_{d\mathbb{N}_0} e^{\vartheta y}(\widehat{\psi}(y+a)-\psi^0(y+a))\,\lambda\!\!\lambda_d(dy) \ &= \ \frac{\psi(\infty)}{1-\phi_Q(0)}\int_{d\mathbb{N}_0} e^{\vartheta y}Q((y,\infty))\,\lambda\!\!\lambda_d(dy)\\ &= \ \frac{d\psi(\infty)}{1-\phi_Q(0)}\sum_{n\geq 0}\sum_{k>n} e^{\vartheta nd}Q(\{kd\})\\ &= \ \frac{d\psi(\infty)}{1-\phi_Q(0)}\sum_{k\geq 1} Q(\{kd\})\sum_{n=0}^{k-1} e^{\vartheta nd}\\ &= \ \frac{d\psi(\infty)}{1-\phi_Q(0)}\sum_{k\geq 1}\frac{e^{\vartheta kd}-1}{e^{\vartheta d}-1}\,Q(\{kd\})\\ &= \ \frac{\psi(\infty)}{(1-\phi_Q(0))e(d,\vartheta)}\sum_{k\geq 0}(e^{\vartheta kd}-1)\,Q(\{kd\}) \ = \ \frac{\psi(\infty)}{e(d,\vartheta)}.\end{aligned}$$

The proof is herewith complete. □

It is worthwile to give the following corollary that provides information on the behavior of the renewal function $\mathbb{U}(t)$ pertaining to an admissible measure Q that possesses a characteristic exponent $\vartheta\neq 0$. The proper renewal case $\vartheta=0$ will be considered more carefully later in the upcoming section.

Corollary 2.47. *Let Q be an admissible measure on $\mathbb{R}_\geq$ with lattice-span d and characteristic exponent ϑ. Then its renewal function $\mathbb{U}(x)$ satisfies*

(a) *in the defective case ($\vartheta>0$):*

$$d\text{-}\lim_{x\to\infty} e^{\vartheta x}\left(\frac{1}{1-\phi_Q(0)}-\mathbb{U}(x)\right) \ = \ d\text{-}\lim_{x\to\infty}\mathbb{U}((x,\infty)) \ = \ \frac{1}{\mu(Q_\vartheta)e(d,\vartheta)}.$$

(b) *in the excessive case ($\vartheta<0$):*

$$d\text{-}\lim_{x\to\infty} e^{\vartheta x}\,\mathbb{U}(x) \ = \ \frac{1}{\mu(Q_\vartheta)|e(d,\vartheta)|}$$

Proof. Since $\mathbb{U}(x) = I(x) + \mathbb{U} * Q(x)$ for $x \in \mathbb{R}_\geq$ with $I := \mathbf{1}_{[0,\infty)}$, we infer from Thm. 2.46 that in the defective case $\mathbb{U}^0(x) = \|U\| - \mathbb{U}((x,\infty))$ satisfies the renewal equation

$$\mathbb{U}^0(x) \;=\; \widehat{I}(x) + \mathbb{U}^0 * Q(x) \quad \text{with} \quad \widehat{I}(x) \;:=\; \|\mathbb{U}\|\, Q((x,\infty)).$$

The function $\widehat{I}_\vartheta$ is dRi by Lemma 2.28 because $\widehat{I}$ is nondecreasing on $\mathbb{R}_\geq$ and $\int_0^\infty \vartheta e^{\vartheta y} Q((y,\infty))\, dy = \phi_Q(\vartheta) - \phi_Q(0) < \infty$. Hence we obtain the asserted result by an appeal to (2.38) of Thm. 2.46.

In the excessive case, $\vartheta < 0$ implies that $I_\vartheta(x) = e^{\vartheta x}\mathbf{1}_{[0,\infty)}(x)$ is dRi so that, by (2.36) of Thm. 2.44(b),

$$d\text{-}\lim_{x\to\infty} e^{\vartheta x}\,\mathbb{U}(x) \;=\; \frac{1}{\mu(Q_\vartheta)} \int_{\mathbb{R}_\geq} e^{\vartheta x}\, \lambda\!\!\lambda_d(dx) \;=\; \frac{1}{\mu(Q_\vartheta)|e(d,\vartheta)|}$$

as claimed. □

2.8 Renewal function and first passage times

Let $(S_n)_{n\geq 0}$ be a RP in a standard model with increment distribution F, lattice-span $d \in \{0,1\}$, *finite* drift μ and renewal measure $\mathbb{U}_\lambda$ under $\mathbb{P}_\lambda$. Recall that $\tau(t) = \inf\{n \geq 0 : S_n > t\}$ denotes the associated first passage time beyond level t for $t \geq 0$. The rather crude asymptotic $t^{-1}\mathbb{U}_0(t) \to \mu^{-1}$, known as the elementary renewal theorem [☞ Lemma 2.1(f)], in combination with $d\text{-}\lim_{t\to\infty}(\mathbb{U}_0(t+h) - \mathbb{U}_0(t)) = \mu^{-1}\lambda\!\!\lambda_d((0,h])$ for any $h > 0$ from Blackwell's theorem 2.20 provides some evidence for the assertion that

$$\mathbb{U}_0(t) \;=\; \frac{t}{\mu} + \Delta + o(1) \quad \text{as } t \to \infty \text{ through } \mathbb{G}_d,$$

where Δ denotes a suitable constant depending on F. This will now in fact be derived via another standard renewal equation and requires the assumption that F has finite variance $\sigma^2 = \mathbb{E}(X - \mu)^2$. The result will also lead to an asymptotic expansion of $\mathbb{E}_0\tau(t)$ up to vanishing terms as $t \to \infty$ because of the simple but important relationship between renewal function and mean first passage times.

Lemma 2.48. *Given a RP $(S_n)_{n\geq 0}$ in a standard model with associated first passage times $\tau(t)$, the identity*

$$\mathbb{U}_\lambda(t) \;=\; \mathbb{E}_\lambda \tau(t) \tag{2.39}$$

holds true for any $t \geq 0$ and $\lambda \in \mathscr{P}(\mathbb{R}_\geq)$.

Proof. It suffices to note that

$$\mathbb{E}_\lambda \tau(t) \;=\; \sum_{n\ge 0} \mathbb{P}_\lambda(\tau(t)>n) \;=\; \sum_{n\ge 0} \mathbb{P}_\lambda(S_n \le t) \;=\; \mathbb{U}_\lambda(t)$$

for any $t \ge 0$ and $\lambda \in \mathscr{P}(\mathbb{R}_\ge)$. □

Observe that, by an appeal to Wald's identity [☞ Prop. 2.53],

$$\mathbb{U}_0(t) \;=\; \frac{1}{\mu}\,\mathbb{E}_0 S_{\tau(t)} \;=\; \frac{t}{\mu} + \frac{1}{\mu}\,\mathbb{E}(S_{\tau(t)} - t) \;\ge\; \frac{t}{\mu}$$

for all $t \ge 0$. Therefore, the function $\Psi(t) := \mathbb{U}_0(t) - \mathbb{U}_{F^s}(t)$ is nonnegative, locally bounded, vanishes on $\mathbb{R}_<$ and equals $U_0(t) - \mu^{-1}t$ for $t \in \mathbb{G}_{d,\ge}$. By (2.8), it satisfies the renewal equation

$$\Psi(t) \;=\; \psi(t) \;+\; \int_{[0,t]} \Psi(t-x)\,F(dx), \quad t \ge 0,$$

where

$$\psi(t) \;:=\; \delta_0(t) - F^s(t) \;=\; \overline{F}^s(t) \quad t \ge 0,$$

is clearly locally bounded. Consequently, Ψ forms the unique locally bounded solution to the renewal equation by Thm. 2.39 and must hence equal $\psi * \mathbb{U}_0$. The function ψ is dRi by (dRi-6) of Prop. 2.27, for it is nonincreasing on $\mathbb{R}_\ge$ and satisfies

$$\int_0^\infty \overline{F}^s(t)\,dt \;=\; \mu^s \;=\; \frac{\sigma^2 + \mu^2}{2\mu} + \frac{d}{2} \;<\; \infty \tag{2.40}$$

by (2.14). The following result, giving the announced second order approximation of the renewal function, is now easily inferred with the help of Thm. 2.42.

Theorem 2.49. *Let $(S_n)_{n\ge 0}$ be a RP in a standard model with lattice-span $d \in \{0,1\}$, drift μ and finite increment variance σ^2. Then*

$$d\text{-}\lim_{t\to\infty} \left(\mathbb{U}_\lambda(t) - \frac{t}{\mu}\right) \;=\; \frac{d}{2\mu} + \frac{\mu^2+\sigma^2}{2\mu^2} - \frac{\mu_0}{\mu} \tag{2.41}$$

for any $\lambda \in \mathscr{P}(\mathbb{R}_\ge)$ having finite mean μ_0.

Proof. The result follows from the previous considerations if $\lambda = \delta_0$ and for general λ with finite mean μ_0 upon using $\mathbb{U}_\lambda = \lambda * \mathbb{U}_0$. Details are left to the reader [☞ Problem 2.51]. □

Regarding the forward recurrence time $R(t) = S_{\tau(t)} - t$, let us finally point out that a combination of the previous result with (2.39), (2.40) and Wald's identity [☞ Prop. 2.53] implies

$$d\text{-}\lim_{t\to\infty} \mathbb{E}_0 R(t) \;=\; d\text{-}\lim_{t\to\infty} \Big(\mu\,\mathbb{E}_0\tau(t) - t\Big) \;=\; d\text{-}\lim_{t\to\infty} \Big(\mu\,\mathbb{U}_0(t) - t\Big) \;=\; \mu^s.$$

But we further know from (2.29) that $R(t)$ converges in distribution to F^s, whence the following result is immediate.

Corollary 2.50. *Let $(S_n)_{n\ge0}$ be a RP in a standard model with lattice-span $d\in\{0,1\}$, drift μ and finite increment variance σ^2. Then the family of forward recurrence times $\{R(t): t\ge 0\}$ is ui.*

Problems

Problem 2.51. Prove Thm. 2.49.

2.9 An intermezzo: random walks, stopping times and ladder variables

Before giving a brief account of the most important extensions of previous results to random walks on the line with positive drift, we collect some important facts about random walks and stopping times including the crucial concept of *ladder variables*. We skip some of the proofs and refer instead to [2].

In the following, let $(S_n)_{n\ge0}$ be a RW in a standard model with increments $X_1,X_2,...$ and increment distribution F. For convenience, it may take values in any $\mathbb{R}^d$, $d\ge1$. We will use $\mathbb{P}$ for probabilities that do not depend on the distribution of S_0. Let further $(\mathscr{F}_n)_{n\ge0}$ be a filtration such that

(F1) $(S_n)_{n\ge0}$ is adapted to $(\mathscr{F}_n)_{n\ge0}$, i.e., $\sigma(S_0,...,S_n)\subset\mathscr{F}_n$ for all $n\in\mathbb{N}_0$.
(F2) $\mathscr{F}_n$ is independent of $(X_{n+k})_{k\ge1}$ for each $n\in\mathbb{N}_0$.

Let also $\mathscr{F}_\infty$ be the smallest σ-field containing all $\mathscr{F}_n$. Condition (F2) ensures that $(S_n)_{n\ge0}$ is a temporally homogeneous Markov chain with respect to $(\mathscr{F}_n)_{n\ge0}$, viz.

$$\mathbb{P}(S_{n+1}\in B|\mathscr{F}_n) \;=\; \mathbb{P}(S_{n+1}\in B|S_n) \;=\; F(B-S_n) \quad \mathbb{P}_\lambda\text{-a.s.}$$

for all $n\in\mathbb{N}_0$, $\lambda\in\mathscr{P}(\mathbb{R}^d)$ and $B\in\mathscr{B}(\mathbb{R}^d)$. A more general, but in fact equivalent statement is that

$$\mathbb{P}((S_{n+k})_{k\ge0}\in C|\mathscr{F}_n) \;=\; \mathbb{P}((S_{n+k})_{k\ge0}\in C|S_n) \;=\; \mathbf{P}(S_n,C) \quad \mathbb{P}_\lambda\text{-a.s.}$$

for all $n\in\mathbb{N}_0$, $\lambda\in\mathscr{P}(\mathbb{R})$ and $C\in\mathscr{B}(\mathbb{R}^d)^{\mathbb{N}_0}$, where

$$\mathbf{P}(x,C) \;:=\; \mathbb{P}_x((S_k)_{k\ge0}\in C) \;=\; \mathbb{P}_0((S_k)_{k\ge0}\in C-x) \quad \text{for } x\in\mathbb{R}^d.$$

Let us recall that, if τ is any stopping time with respect to $(\mathscr{F}_n)_{n\ge 0}$, also called $(\mathscr{F}_n)$-*time* hereafter, then

$$\mathscr{F}_\tau \;=\; \{A\in\mathscr{F}_\infty : A\cap\{\tau\le n\}\in\mathscr{F}_n \text{ for all } n\in\mathbb{N}_0\},$$

and the random vector $(\tau,S_0,...,S_\tau)\mathbf{1}_{\{\tau<\infty\}}$ is $\mathscr{F}_\tau$-measurable. The following basic result combines the strong Markov property and temporal homogeneity of $(S_n)_{n\ge 0}$ as a Markov chain with its additional *spatial homogeneity* owing to its iid increments.

Proposition 2.52. *Under the stated assumptions, let τ be a $(\mathscr{F}_n)$-time. Then, for all $\lambda\in\mathscr{P}(\mathbb{R}^d)$, the following equalities hold $\mathbb{P}_\lambda$-a.s. on $\{\tau<\infty\}$:*

$$\mathbb{P}((S_{\tau+n}-S_\tau)_{n\ge 0}\in\cdot|\mathscr{F}_\tau) = \mathbb{P}((S_n-S_0)_{n\ge 0}\in\cdot) = \mathbb{P}_0((S_n)_{n\ge 0}\in\cdot). \quad (2.42)$$

$$\mathbb{P}((X_{\tau+n})_{n\ge 1}\in\cdot|\mathscr{F}_\tau) = \mathbb{P}((X_n)_{n\ge 1}\in\cdot). \quad (2.43)$$

If $\mathbb{P}_\lambda(\tau<\infty)=1$, then furthermore (under $\mathbb{P}_\lambda$)

(a) $(S_{\tau+n}-S_\tau)_{n\ge 0}$ *and $\mathscr{F}_\tau$ are independent.*
(b) $(S_{\tau+n}-S_\tau)_{n\ge 0}\stackrel{d}{=}(S_n-S_0)_{n\ge 0}$.
(c) $X_{\tau+1},X_{\tau+2},...$ *are iid with the same distribution as X_1.*

Proof. It suffices to prove (2.43) for which we pick any $k\in\mathbb{N}_0$, $n\in\mathbb{N}$, $B_1,...,B_n\in\mathscr{B}(\mathbb{R}^d)$ and $A\in\mathscr{F}_\tau$. Using $A\cap\{\tau=k\}\in\mathscr{F}_k$ and (F2), it follows for each $\lambda\in\mathscr{P}(\mathbb{R}^d)$ that

$$\begin{aligned}&\mathbb{P}_\lambda(A\cap\{\tau=k,X_{k+1}\in B_1,...,X_{k+n}\in B_n\})\\ &\quad=\;\mathbb{P}_\lambda(A\cap\{\tau=k\})\,\mathbb{P}(X_{k+1}\in B_1,...,X_{k+n}\in B_n)\\ &\quad=\;\mathbb{P}_\lambda(A\cap\{\tau=k\})\,\mathbb{P}(X_1\in B_1,...,X_n\in B_n),\end{aligned}$$

and this yields the desired conclusion. □

We continue with a statement of two very useful identities originally due to A. WALD [112] for the first and second moment of stopped sums S_τ for finite mean stopping times τ, known as *Wald's equations* or *Wald's identities*. The first of these has already been used before.

Proposition 2.53. [Wald's equations] *Let $(S_n)_{n\ge 0}$ be a SRW adapted to a filtration $(\mathscr{F}_n)_{n\ge 0}$ satisfying (F1) and (F2). Let further τ be an a.s. finite $(\mathscr{F}_n)$-time and suppose that $\mu:=\mathbb{E}X_1$ exists. Then*

$$\mathbb{E}S_\tau \;=\; \mu\,\mathbb{E}\tau \qquad \textbf{(Wald's equation)}$$

provided that either the X_n are a.s. nonnegative, or τ has finite mean. If the X_n have also finite variance σ^2, then furthermore

$$\mathbb{E}(S_\tau - \mu\tau)^2 \;=\; \sigma^2\,\mathbb{E}\tau \qquad\qquad \textbf{(Wald's 2nd equation)}$$

for any $(\mathscr{F}_n)$-time τ with finite mean.

Proof. ☞ [2, Prop. 2.11 and 2.12]. □

Assuming $S_0 = 0$ hereafter, let us now turn to the concept of formally copying a stopping time τ for $(S_n)_{n\ge1}$. The latter means that there exist $B_n \in \mathscr{B}(\mathbb{R}^{nd})$ for $n \ge 1$ such that

$$\tau \;=\; \inf\{n \ge 1 : (S_1,...,S_n) \in B_n\}, \tag{2.44}$$

where as usual $\inf\emptyset := \infty$. With the help of the B_n we can copy this stopping rule to the *post-τ process* $(S_{\tau+n} - S_\tau)_{n\ge1}$ if $\tau < \infty$. For this purpose put $S_{n,k} := S_{n+k} - S_n$,

$$\begin{aligned}\mathbf{S}_{n,k} \;&:=\; (S_{n+1} - S_n,...,S_{n+k} - S_n) \;=\; (S_{n,1},...,S_{n,k}) \quad \text{and}\\ \mathbf{X}_{n,k} \;&:=\; (X_{n+1},...,X_{n+k})\end{aligned}$$

for $k \in \mathbb{N}$ and $n \in \mathbb{N}_0$.

Definition 2.54. Let τ be a stopping time for $(S_n)_{n\ge1}$ as in (2.44). Then the sequences $(\tau_n)_{n\ge1}$ and $(\sigma_n)_{n\ge0}$, defined by $\sigma_0 := 0$ and

$$\tau_n \;:=\; \begin{cases}\inf\{k \ge 1 : \mathbf{S}_{\sigma_{n-1},k} \in B_k\}, & \text{if } \sigma_{n-1} < \infty\\ \infty, & \text{if } \sigma_{n-1} = \infty\end{cases} \quad \text{and} \quad \sigma_n \;:=\; \sum_{k=1}^{n} \tau_k$$

for $n \ge 1$ (thus $\tau_1 = \tau$) are called the *sequence of formal copies of* τ and its associated *sequence of copy sums*, respectively.

The following proposition summarizes the most important properties of the τ_n, σ_n and $S_{\sigma_n}\mathbf{1}_{\{\sigma_n<\infty\}}$.

Proposition 2.55. *Given the previous notation, put further $\beta := \mathbb{P}(\tau < \infty)$ and $\mathbf{Z}_n := (\tau_n, \mathbf{X}_{\sigma_{n-1},\tau_n})$ for $n \in \mathbb{N}$. Then the following assertions hold true:*

(a) $\sigma_0, \sigma_1, \ldots$ *are stopping times for* $(S_n)_{n\ge0}$.
(b) τ_n *is a stopping time with respect to* $(\mathscr{F}_{\sigma_{n-1}+k})_{k\ge0}$ *and* $\mathscr{F}_{\sigma_{n-1}}$*-measurable for each* $n \in \mathbb{N}$.
(c) $\mathbb{P}(\tau_n \in \cdot|\mathscr{F}_{\sigma_{n-1}}) = \mathbb{P}(\tau < \infty)$ *a.s. on* $\{\sigma_{n-1} < \infty\}$ *for each* $n \in \mathbb{N}$.
(d) $\mathbb{P}(\tau_n < \infty) = \mathbb{P}(\sigma_n < \infty) = \beta^n$ *for all* $n \in \mathbb{N}$.

(e) $\mathbb{P}(\mathbf{Z}_n \in \cdot, \tau_n < \infty | \mathscr{F}_{\sigma_{n-1}}) = \mathbb{P}(\mathbf{Z}_1 \in \cdot, \tau_1 < \infty)$ *a.s. on* $\{\sigma_{n-1} < \infty\}$ *for all* $n \in \mathbb{N}$.
(f) *Given* $\sigma_n < \infty$*, the random vectors* $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ *are conditionally iid with the same distribution as* $\mathbf{Z}_1$ *conditioned upon* $\tau_1 < \infty$.
(g) *If* $G := \mathbb{P}((\tau, S_\tau) \in \cdot | \tau < \infty)$*, then* $\mathbb{P}((\sigma_n, S_{\sigma_n}) \in \cdot | \sigma_n < \infty) = G^{*n}$ *a.s. for all* $n \in \mathbb{N}$.

In the case where τ *is a.s. finite* ($\beta = 1$)*, this implies further:*

(h) $\mathbf{Z}_n$ *and* $\mathscr{F}_{\sigma_{n-1}}$ *are independent for each* $n \in \mathbb{N}$.
(i) $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ *are iid.*
(j) $(\sigma_n, S_{\sigma_n})_{n \geq 0}$ *forms a SRW taking values in* $\mathbb{N}_0 \times \mathbb{R}^d$.

Proof. The simple proof of (a) and (b) is left to the reader. Assertion (c) and (e) follow from (2.42) when observing that, on $\{\sigma_{n-1} < \infty\}$,

$$\tau_n \;=\; \sum_{k \geq 0} \mathbf{1}_{\{\tau_n > k\}} \;=\; \sum_{k \geq 0} \prod_{j=1}^{k} \mathbf{1}_{B_j^c}(\mathbf{S}_{\sigma_{n-1},j}) \quad \text{and} \quad \mathbf{Z}_n \, \mathbf{1}_{\{\tau_n < \infty\}}$$

are measurable functions of $(S_{\sigma_{n-1},k})_{k \geq 0}$. Since $\mathbb{P}(\tau_n < \infty) = \mathbb{P}(\tau_1 < \infty, \ldots, \tau_n < \infty)$, we infer (d) by an induction over n and use of (c). Another induction in combination with (d) gives assertion (f) once we have proved that

$$\begin{aligned}
&\mathbb{P}((\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in A_n, \mathbf{Z}_{n+1} \in B, \sigma_{n+1} < \infty) \\
&\qquad = \mathbb{P}((\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in A_n, \sigma_n < \infty) \, \mathbb{P}(\mathbf{Z}_1 \in B, \tau < \infty)
\end{aligned}$$

for all $n \in \mathbb{N}$ and A_n, B from the σ-fields obviously to be chosen here. But with the help of (e), this is inferred as follows:

$$\begin{aligned}
&\mathbb{P}((\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in A_n, \mathbf{Z}_{n+1} \in B, \sigma_{n+1} < \infty) \\
&\quad = \mathbb{P}((\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in A_n, \mathbf{Z}_{n+1} \in B, \sigma_n < \infty, \tau_{n+1} < \infty) \\
&\quad = \int_{\{(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in A_n, \sigma_n < \infty\}} \mathbb{P}(\mathbf{Z}_{n+1} \in B, \tau_{n+1} < \infty | \mathscr{F}_{\sigma_n}) \, d\mathbb{P} \\
&\quad = \mathbb{P}((\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \in A_n, \sigma_n < \infty) \, \mathbb{P}(\mathbf{Z}_1 \in B, \tau < \infty).
\end{aligned}$$

Assertion (g) is a direct consequence of (f), and the remaining assertion (h),(i) and (j) in the case $\beta = 1$ are just the specializations of (e),(f) and (g) to this case. □

The most prominent sequences of copy sums in the theory of RW's are obtained by looking at the record epochs and record values of a RW $(S_n)_{n \geq 0}$, or its reflection $(-S_n)_{n \geq 0}$. They also provide a key tool for the extension of renewal theory to random walks with positive drift

Definition 2.56. Given a SRW $(S_n)_{n\ge 0}$, the stopping times

$$\sigma^> := \inf\{n \ge 1 : S_n > 0\}, \quad \sigma^\ge := \inf\{n \ge 1 : S_n \ge 0\},$$
$$\sigma^< := \inf\{n \ge 1 : S_n < 0\}, \quad \sigma^\le := \inf\{n \ge 1 : S_n \le 0\},$$

are called *first strictly ascending, weakly ascending, strictly descending* and *weakly descending ladder epoch*, respectively, and

$$S_1^> := S_{\sigma^>} \mathbf{1}_{\{\sigma^> < \infty\}}, \quad S_1^\ge := S_{\sigma^\ge} \mathbf{1}_{\{\sigma^\ge < \infty\}},$$
$$S_1^< := S_{\sigma^<} \mathbf{1}_{\{\sigma^< < \infty\}}, \quad S_1^\le := S_{\sigma^\le} \mathbf{1}_{\{\sigma^\le < \infty\}}$$

their respective *ladder heights*. The associated sequences of copy sums $(\sigma_n^>)_{n\ge 0}$, $(\sigma_n^\ge)_{n\ge 0}$, $(\sigma_n^<)_{n\ge 0}$ and $(\sigma_n^\le)_{n\ge 0}$ are called *sequences of strictly ascending, weakly ascending, strictly descending* and *weakly descending ladder epochs*, respectively, and

$$S_n^> := S_{\sigma_n^>} \mathbf{1}_{\{\sigma_n^> < \infty\}},\ n \ge 0, \quad S_n^\ge := S_{\sigma_n^\ge} \mathbf{1}_{\{\sigma_n^\ge < \infty\}},\ n \ge 0,$$
$$S_n^< := S_{\sigma_n^<} \mathbf{1}_{\{\sigma_n^< < \infty\}},\ n \ge 0, \quad S_n^\le := S_{\sigma_n^\le} \mathbf{1}_{\{\sigma_n^\le < \infty\}},\ n \ge 0$$

the respective *sequences of ladder heights*.

Plainly, if $(S_n)_{n\ge 0}$ has nonnegative (positive) increments, then $\sigma_n^\ge = n$ ($\sigma_n^> = n$) for all $n \in \mathbb{N}$. Moreover, $\sigma_n^\ge = \sigma_n^>$ and $\sigma_n^\le = \sigma_n^<$ a.s. for all $n \in \mathbb{N}$ in the case where the increment distribution is continuous, for then $\mathbb{P}(S_m = S_n) = 0$ for all $m, n \in \mathbb{N}$.

The following proposition provides some basic information on the ladder variables and is a consequence of the SLLN and Prop. 2.55.

Proposition 2.57. *Let $(S_n)_{n\ge 0}$ be a nontrivial SRW. Then the following assertions are equivalent:*

(a) $\left(\sigma_n^\alpha, S_{\sigma_n^\alpha}\right)_{n\ge 0}$ *is a SRW taking values in $\mathbb{N}_0 \times \mathbb{R}$ for any $\alpha \in \{>, \ge\}$ (resp. $\{<, \le\}$).*
(b) $\sigma^\alpha < \infty$ *a.s. for $\alpha \in \{>, \ge\}$ (resp. $\{<, \le\}$).*
(c) $\limsup_{n\to\infty} S_n = \infty$ *a.s. (resp. $\liminf_{n\to\infty} S_n = -\infty$ a.s.)*

Proof. It clearly suffices to prove equivalence of the assertions outside parentheses. The implications "(a)⇒(b)" and "(c)⇒(b)" are trivial, while "(b)⇒(a)" follows from Prop. 2.55(j). This leaves us with a proof of "(a),(b)⇒(c)". But $\mathbb{E}S^> > 0$ in combination with the SLLN applied to $(S_n^>)_{n\ge 0}$ implies

$$\limsup_{n\to\infty} S_n \ \ge\ \lim_{n\to\infty} S_n^> = \infty \quad \text{a.s.}$$

and thus the assertion. □

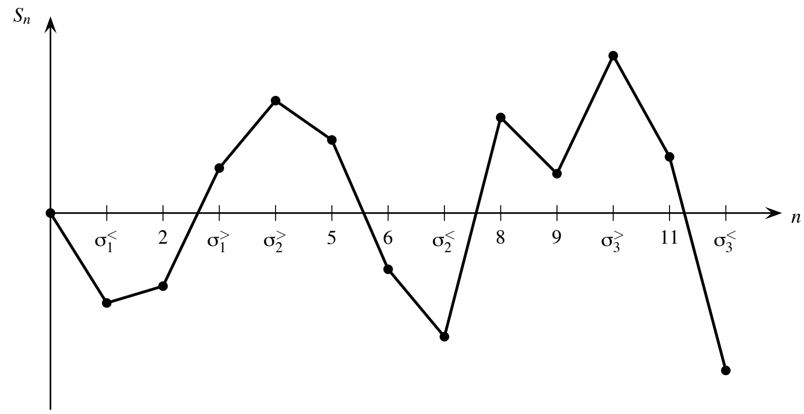


Fig. 2.2 Path of a RW with strictly ascending ladder epochs $\sigma_1^> = 3$, $\sigma_2^> = 4$ and $\sigma_3^> = 10$, and strictly descending ladder epochs $\sigma_1^< = 1$, $\sigma_2^< = 5$ and $\sigma_3^< = 12$.

If $\mathbb{E}X_1 > 0$ (resp. < 0) we thus have that $\sigma^>, \sigma^\geq$ (resp. $\sigma^<, \sigma^\leq$) are a.s. finite whence the associated sequences of ladder epochs and ladder heights each constitute nondecreasing (resp. nonincreasing) zero-delayed RW's. Much deeper information, however, is provided by Prop. 2.59 below which discloses a quite unexpected duality between ascending and descending ladder epochs that will enable us to derive a further classification of RW's as to their asymptotic behavior including the *Chung-Fuchs theorem* on the asymptotic behavior of a RW with drift zero. We pause for the following lemma about the lattice-type of a ladder height.

Lemma 2.58. *Let $(S_n)_{n\geq 0}$ be a nontrivial SRW and σ an a.s. finite first ladder epoch. Then $d(X_1) = d(S_\sigma)$.*

Proof. ☞ [2, Lemma 2.33] or Problem 2.62. □

Proposition 2.59. *Given a SRW $(S_n)_{n\geq 0}$ with first ladder epochs $\sigma^\geq$, $\sigma^>$, $\sigma^\leq$, $\sigma^<$, the following assertions hold true:*

$$\mathbb{E}\sigma^\geq = \frac{1}{\mathbb{P}(\sigma^< = \infty)} \quad \textit{and} \quad \mathbb{E}\sigma^> = \frac{1}{\mathbb{P}(\sigma^\leq = \infty)}, \tag{2.45}$$

$$\mathbb{P}(\sigma^\leq = \infty) = (1-\kappa)\,\mathbb{P}(\sigma^< = \infty), \tag{2.46}$$

where

$$\kappa := \sum_{n\ge 1}\mathbb{P}(S_1>0,\dots,S_{n-1}>0,S_n=0) = \sum_{n\ge 1}\mathbb{P}(\sigma^{\le}=n,S_1^{\le}=0).$$

Proof. ☞ [2, Prop. 2.15]. □

We close this section with the announced classification of RW's that provides us with a good understanding of their long-run behavior.

Theorem 2.60. *Let $(S_n)_{n\ge 0}$ be a nontrivial SRW. Then exactly one of the following three cases holds true:*

(i) $\sigma^{\le},\sigma^{<}$ *are both defective and* $\mathbb{E}\sigma^{\ge},\mathbb{E}\sigma^{>}$ *are both finite.*
(ii) $\sigma^{\ge},\sigma^{>}$ *are both defective and* $E\sigma^{\le},E\sigma^{<}$ *are both finite.*
(iii) $\sigma^{\ge},\sigma^{>},\sigma^{\le},\sigma^{<}$ *are all a.s. finite with infinite expectation.*

In terms of the asymptotic behavior of S_n as $n\to\infty$, these three alternatives are characterized as follows:

(i) $\lim_{n\to\infty}S_n=\infty$ *a.s.*
(ii) $\lim_{n\to\infty}S_n=-\infty$ *a.s.*
(iii) $\liminf_{n\to\infty}S_n=-\infty$ *and* $\limsup_{n\to\infty}S_n=\infty$ *a.s.*

Finally, if $\mu := EX_1$ exists, thus $EX^+<\infty$ or $EX^-<\infty$, then (i), (ii), and (iii) are equivalent to $\mu>0$, $\mu<0$, and $\mu=0$, respectively.

The last stated fact that alternative (iii) occurs for any SRW with drift $\mu=0$ is usually referred to as the ***Chung-Fuchs theorem***.

Proof. Notice first that $\mathbb{P}(X_1=0)<1$ is equivalent to $\kappa<1$, whence (2.46) ensures that $\sigma^{>},\sigma^{\ge}$ as well as $\sigma^{<},\sigma^{\le}$ are always defective simultaneously in which case the respective dual ladder epochs have finite expectation by (2.45). Hence, if neither (a) nor (b) holds true, the only remaining alternative is that all four ladder epochs are a.s. finite with infinite expectation. By combining the three alternatives for the ladder epochs just proved with Prop. 2.57, the respective characterizations of the behavior of S_n for $n\to\infty$ are immediate.

Suppose now that $\mu=\mathbb{E}X_1$ exists. In view of Prop. 2.57 it then only remains to verify that (iii) holds true in the case $\mu=0$. But any of the alternatives (i) or (ii) would lead to the existence of a ladder epoch σ such that $\mathbb{E}\sigma<\infty$ and S_σ is a.s. positive or negative. On the other hand, $\mathbb{E}S_\sigma=\mu\,\mathbb{E}\sigma=0$ would follow by an appeal to Wald's identity which is impossible. Hence $\mu=0$ entails (iii). □

The following definition gives names to the three above alternatives (i), (ii) and (iii) that may occur for a RW $(S_n)_{n\ge 0}$.

Definition 2.61. A RW $(S_n)_{n\ge 0}$ is called

– *positive divergent* if $\lim_{n\to\infty} S_n = \infty$ a.s.
– *negative divergent* if $\lim_{n\to\infty} S_n = -\infty$ a.s.
– *oscillating* if $\liminf_{n\to\infty} S_n = -\infty$ and $\limsup_{n\to\infty} S_n = \infty$ a.s.

Problems

Problem 2.62. Prove Lemma 2.58.

2.10 Two-sided renewal theory: a short path to extensions

Now we are ready to extend some of the previously stated renewal theorems to RW's with positive drift. The basic idea is as simple as effective and based upon the use of the embedded RP of strictly ascending ladder heights.

For most of the following derivations it suffices to consider the zero-delayed case when $S_0 = 0$, for the result in the general case then usually follows by a straightforward argument. So let $(S_n)_{n\ge 0}$ be a SRW with increment distribution F, positive drift μ and embedded SRP $(S_n^>)_{n\ge 0}$ of strictly ascending ladder heights the drift of which we denote by $\mu^> = \mathbb{E}S_1^>$. Since $\mathbb{E}\sigma^>$ is finite, Wald's equation implies

$$\mu^> \;=\; \mathbb{E}S_1^> \;=\; \mathbb{E}S_{\sigma^>} \;=\; \mu\,\mathbb{E}\sigma^>$$

even if $\mu = \infty$. As before, let $\mathbb{U} = \sum_{n\ge 0} F^{*n} = \sum_{n\ge 0} \mathbb{P}(S_n \in \cdot)$ be the renewal measure of $(S_n)_{n\ge 0}$ so that $\mathbb{U}(A)$ gives the expected number of visits of the RW to $A \in \mathscr{B}(\mathbb{R})$. We remark that it is not clear at this point whether $\mathbb{U}(A)$ is always finite for any bounded B as in the renewal case. The renewal measure of $(S_n^>)_{n\ge 0}$ is denoted $\mathbb{U}^>$.

2.10.1 The key tool: cyclic decomposition

Let σ be an a.s. finite stopping time for $(S_n)_{n\ge 1}$ with associated sequence $(\sigma_n)_{n\ge 0}$ of copy sums. Denote by $\mathbb{U}^{(\sigma)}$ the renewal measure of the RW $(S_{\sigma_n})_{n\ge 0}$ and define the *pre-σ occupation measure of* $(S_n)_{n\ge 0}$ by

$$\mathbb{V}^{(\sigma)}(A) \;:=\; \mathbb{E}\left(\sum_{n=0}^{\sigma-1} \mathbf{1}_A(S_n)\right) \quad \text{for } A \in \mathscr{B}(\mathbb{R}), \tag{2.47}$$

which has total mass $\|V^{(\sigma)}\| = \mathbb{E}\sigma$ and is hence finite if σ has finite mean. The next lemma shows that $\mathbb{U}$ and $\mathbb{U}^{(\sigma)}, \mathbb{V}^{(\sigma)}$ are related in a nice way and holds true even without any assumption on the drift of $(S_n)_{n\ge 0}$.

Lemma 2.63. [Cyclic decomposition formula] *Under the stated assumptions,*

$$\mathbb{U} \;=\; \mathbb{V}^{(\sigma)} * \mathbb{U}^{(\sigma)}$$

for any a.s. finite stopping time σ for $(S_n)_{n\ge 1}$.

Proof. Using cyclic decomposition with the help of the σ_n, we obtain

$$\begin{aligned}
\mathbb{U}(A) \;&=\; \mathbb{E}\left(\sum_{k\ge 0} \mathbf{1}_A(S_k)\right) \;=\; \sum_{n\ge 0} \mathbb{E}\left(\sum_{k=\sigma_n}^{\sigma_{n+1}-1} \mathbf{1}_A(S_k)\right)\\
&=\; \sum_{n\ge 0} \int_{\mathbb{R}} E\left(\sum_{k=\sigma_n}^{\sigma_{n+1}-1} \mathbf{1}_{A-x}(S_k - S_{\sigma_n})\middle| S_{\sigma_n} = x\right) \mathbb{P}(S_{\sigma_n} \in dx)\\
&=\; \sum_{n\ge 0} \int_{\mathbb{R}} \mathbb{V}^{(\sigma)}(A-x)\, \mathbb{P}(S_{\sigma_n} \in dx)\\
&=\; \mathbb{V}^{(\sigma)} * \mathbb{U}^{(\sigma)}(A) \quad \text{for all } A \in \mathscr{B}(\mathbb{R}),
\end{aligned}$$

where (2.42) of Prop. 2.52 has been utilized in the penultimate line. □

Specializing to $\sigma = \sigma^>$ and writing $\mathbb{V}^>$ for $\mathbb{V}^{(\sigma^>)}$, the cyclic decomposition formula takes the form

$$\mathbb{U} \;=\; \mathbb{V}^> * \mathbb{U}^>. \tag{2.48}$$

We thus have a convolution formula for the renewal measure $\mathbb{U}$ that involves a finite measure concentrated on $\mathbb{R}_\le$, viz. $\mathbb{V}^>$, and the renewal measure of a SRP, namely $\mathbb{U}^>$, for which the asymptotic behavior has been found in the previous sections. Various results for RP's including Blackwell's theorem and the key renewal theorem will now easily be extended to RW's with positive drift with help of this formula.

If $(S_n)_{n\ge 0}$, given in a standard model, has arbitrary initial distribution λ, then Lemma 2.63 in combination with $\mathbb{U}_\lambda = \lambda * \mathbb{U}_0$ immediately implies

$$\mathbb{U}_\lambda \;=\; \lambda * \mathbb{V}^{(\sigma)} * \mathbb{U}^{(\sigma)} \;=\; \mathbb{V}^{(\sigma)} * \mathbb{U}_\lambda^{(\sigma)} \tag{2.49}$$

where $\mathbb{V}^{(\sigma)}, \mathbb{U}^{(\sigma)}$ are defined as before under $\mathbb{P}_0$.

Returning to the zero-delayed situation, let us further note that a simple computation shows that

$$\mathbb{V}^{(\sigma)} \;=\; \sum_{n\ge 0} \mathbb{P}(\sigma > n, S_n \in \cdot) \tag{2.50}$$

[☞ Problem 2.69] and that, for any real- or complex-valued function f

$$\int f \, d\mathbb{V}^{(\sigma)} \;=\; \sum_{n\ge 0} \int_{\{\sigma>n\}} f(S_n)\, d\mathbb{P} \;=\; \mathbb{E}\left(\sum_{n=0}^{\sigma-1} f(S_n)\right) \tag{2.51}$$

whenever one of the three expressions exist.

2.10.2 Uniform local boundedness and stationary delay distribution

The following lemma showing uniform local boundedness of the renewal measure for any random walk with positive drift is the partial extension of Lemma 2.14. A full extension by extending the argument given there is stated as Problem 2.71.

Lemma 2.64. *Let $(S_n)_{n\ge 0}$ be a RW with positive drift in a standard model. Then $\mathbb{U}_\lambda$ is uniformly locally bounded for each $\lambda \in \mathscr{P}(\mathbb{R})$, in fact*

$$\sup_{t\in\mathbb{R}} \mathbb{U}_\lambda([t,t+h]) \;\le\; \mathbb{E}\sigma^{>}\, \mathbb{U}_0^{>}(h) \tag{2.52}$$

for all $h > 0$.

Proof. For any $\lambda \in \mathscr{P}(\mathbb{R})$ and $h > 0$, the cyclic decomposition formula (2.49) with $\sigma = \sigma^{>}$ in combination with $\|\mathbb{V}^{>}\| = \mathbb{E}\sigma^{>}$ and

$$\sup_{t\in\mathbb{R}} \mathbb{U}_\lambda^{>}([t,t+h]) \;\le\; \mathbb{U}_0^{>}(h)$$

by Lemma 2.14 yields

$$\begin{aligned} \mathbb{U}_\lambda([t,t+h]) \;&=\; \mathbb{V}^{>} * \mathbb{U}_\lambda^{>}([t,t+h]) \\ &=\; \int \mathbb{U}_\lambda^{>}([t-x,t-x+h])\, \mathbb{V}^{>}(dx) \\ &\le\; \mathbb{E}\sigma^{>}\, \mathbb{U}_0^{>}(h) \end{aligned}$$

as claimed. □

Cyclic decomposition also allows us to generalize the results from Subections 2.4.2 and 2.4.3 about the stationary delay distribution. This is accomplished by considering F^s, F_a^s and ξ as defined there, but for the associated ladder height RP $(S_n^{>})_{n\ge 0}$. Hence we put

$$\xi(t) \;:=\; \begin{cases} \int_0^t \mathbb{P}(S_1^{>} > x)\, dx, & \text{if } d = 0, \\ \sum_{k=1}^{n(t)} \mathbb{P}(S_1^{>} \ge k), & \text{if } d = 1 \end{cases} \tag{2.53}$$

for $t \in \mathbb{R}_\geq$ (with $n(t)$ as in Prop. 2.15) and then again F_a^s by (2.11) for $a \in \mathbb{R}_>$. If $S_1^>$ has finite mean $\mu^>$ and hence ξ is finite, then let F^s be its normalization, i.e. $F^s = (\mu^>)^{-1}\xi$. Recall from Lemma 2.58 that $S_1^>$ and X_1 are of the same lattice-type.

Theorem 2.65. *Let $(S_n)_{n\geq 0}$ be a RW in a standard model with positive drift μ and lattice-span $d \in \{0,1\}$. Then the following assertions hold with ξ, F_a^s and F^s as defined in* (2.53) *and thereafter.*

(a) $\mathbb{U}_\xi^+ = \mathbb{E}\sigma^> \lambda_d^+$.
(b) $\mathbb{U}_{F_a^s}^+ \leq \xi(a)^{-1}\mathbb{E}\sigma^> \lambda_d^+$ *for all* $a \in \mathbb{R}_>$.
(c) *If μ is finite, then* $\mathbb{U}_{F^s}^+ = \mu^{-1}\lambda_d^+$.

Proof. First note that $\mu > 0$ implies $\mathbb{E}\sigma^> < \infty$ [☞ Thm. 2.60] and $\mu^> = \mathbb{E}S_1^> = \mu\,\mathbb{E}\sigma^>$ by Wald's identity. By (2.49), $\mathbb{U}_\lambda = \mathbb{V}^> * \mathbb{U}_\lambda^>$ for any distribution λ and this obviously extends to arbitrary locally finite measures λ. Therefore,

$$\begin{aligned}
\mathbb{U}_\xi(A) \;&=\; \mathbb{V}^> * \mathbb{U}_\xi^>(A) \;=\; \int_{\mathbb{R}_\leq} \mathbb{U}_\xi^>(A-x)\,\mathbb{V}^>(dx) \\
&=\; \int_{\mathbb{R}_\leq} \lambda_d^+(A-x)\,\mathbb{V}^>(dx) \;=\; \mathbb{V}^>(\mathbb{R}_\leq)\,\lambda_d^+(A) \\
&=\; \mathbb{E}\sigma^>\,\lambda_d^+(A) \quad \text{for all } A \in \mathscr{B}(\mathbb{R}_\geq)
\end{aligned}$$

where Cor. 2.16, the translation invariance of λ_d and $A - x \subset \mathbb{R}_\geq$ for all $x \in \mathbb{R}_\leq$ have been utilized. Hence assertion (a) is proved. As (b) and (c) are shown in a similar manner, we omit supplying the details again and only note for (c) that, if $\mu < \infty$, $\mathbb{U}_{F^s}^+ = (\mu^>)^{-1}\,\mathbb{E}\sigma^>\,\lambda_d^+$ really equals $\mu^{-1}\lambda_d^+$ because $\mu^> = \mu\,\mathbb{E}\sigma^>$ as mentioned above. □

2.10.3 Extensions of Blackwell's and the key renewal theorem

Extensions of the two main renewal theorems to RW's with positive drift are now obtained in a straightforward manner by combination of these results for the ladder height RP with cyclic decomposition.

Theorem 2.66. [Blackwell's renewal theorem] *Let $(S_n)_{n\geq 0}$ be a RW in a standard model with lattice-span $d \geq 0$ and positive drift μ. Then*

$$d\text{-}\lim_{t\to\infty} \mathbb{U}_\lambda([t,t+h]) \;=\; \mu^{-1}\lambda_d([0,h]) \quad \textit{and} \tag{2.54}$$

$$\lim_{t\to-\infty} \mathbb{U}_\lambda([t,t+h]) \;=\; 0 \tag{2.55}$$

for all $h > 0$ and $\lambda \in \mathscr{P}(\mathbb{G}_d)$, where $\mu^{-1} := 0$ if $\mu = \infty$.

Proof. By another use of cyclic decomposition in combination with Lemma 2.64, Blackwell's theorem for $\mathbb{U}_\lambda^>$ and of $\mu\,\|\mathbb{V}^>\| = \mu\,\mathbb{E}\sigma^> = \mu^>$ if $\mu < \infty$, it follows with the help of the dominated convergence theorem that

$$\begin{aligned} d\text{-}\lim_{t\to\infty} \mathbb{U}_\lambda([t,t+h]) &= \int d\text{-}\lim_{t\to\infty} \mathbb{U}_\lambda^>([t-x,t-x+h])\,\mathbb{V}^>(dx) \\ &= \frac{\|\mathbb{V}^>\|\,\lambda\!\!\lambda_d([0,h])}{\mu^>} = \frac{\lambda\!\!\lambda_d([0,h])}{\mu} \end{aligned}$$

for any $h > 0$, i.e. (2.54). But (2.55) follows analogously, for $\mathbb{U}^>$ vanishes on the negative halfline giving $\lim_{t\to-\infty} \mathbb{U}_\lambda^>([t,t+h]) = 0$. □

Theorem 2.67. [Key renewal theorem] *Let $(S_n)_{n\ge0}$ be a RW with positive drift μ, lattice-span $d \in \{0,1\}$ and renewal measure $\mathbb{U}$. Then*

$$d\text{-}\lim_{t\to\infty} g*\mathbb{U}(t) = \frac{1}{\mu}\int g\,d\lambda\!\!\lambda_d \quad \textit{and} \tag{2.56}$$

$$\lim_{t\to-\infty} g*\mathbb{U}(t) = 0 \tag{2.57}$$

for every dRi function $g : \mathbb{R} \to \mathbb{R}$.

Proof. Given a dRi function $g : \mathbb{R} \to \mathbb{R}$, we leave it to the reader [☞ Problem 2.72] to verify that Thm. 2.31 still applies and yields

$$d\text{-}\lim_{t\to\infty} g*\mathbb{U}^>(t) = \frac{1}{\mu^>}\int_{-\infty}^{\infty} g(x)\,\lambda\!\!\lambda_d(dx)$$

as well as

$$\lim_{t\to-\infty} g*\mathbb{U}^>(t) = 0.$$

In particular, $g*\mathbb{U}^>$ is a bounded function. Then use cyclic decomposition and the dominated convergence theorem to infer that

$$g*\mathbb{U}(t) = \int g*\mathbb{U}^>(t-x)\,\mathbb{V}^>(dx)$$

has the asserted limits as $t \to \infty$ (through $\mathbb{G}_d$) and $t \to -\infty$. □

We finish with a brief look at the renewal function $\mathbb{U}(t) = \mathbb{U}((-\infty,t])$ for a SRW $(S_n)_{n\ge0}$ with positive drift μ. We know that $(\mu^>)^{-1}t \le \mathbb{U}^>(t) \le (\mu^>)^{-1}t + C$ for all $t \ge 0$ and a suitable constant $C \in \mathbb{R}_>$, whence cyclic decomposition provides us with the estimate

$$\mathbb{U}(t) \;=\; \int_{\mathbb{R}_\leq} \mathbb{U}^>(t-x)\,\mathbb{V}^>(dx) \begin{cases} \geq\; \mu^{-1}t \;+\; \int_{\mathbb{R}_\leq} |x|\,\mathbb{V}^>(dx) \\ \leq\; \mu^{-1}t \;+\; \int_{\mathbb{R}_\leq} |x|\,\mathbb{V}^>(dx) \;+\; C\,\mathbb{E}\sigma^> \end{cases}$$

for all $t \geq 0$. As a consequence, $\mathbb{U}(t) < \infty$ for some/all $t \in \mathbb{R}$ holds iff (use (2.51))

$$\int_{\mathbb{R}_\leq} |x|\,\mathbb{V}^>(dx) \;=\; \mathbb{E}\left(\sum_{n=0}^{\sigma^>-1} |S_n|\right) \;=\; \mathbb{E}\left(\sum_{n=0}^{\sigma^>-1} S_n^-\right) \;<\; \infty.$$

A nontrivial result not derived here states [☞ [2, Cor. 6.25]] that this is equivalent to the moment condition

$$\mu_2^- := \mathbb{E}(X^-)^2 < \infty.$$

2.10.4 An application: Tail behavior of $\sup_{n\geq 0} S_n$ in the negative drift case

In Applied Probability, the supremum of a SRW with negative drift (or, equivalently, the minimum of a SRW with positive drift) pops up in various contexts like the ruin problem in insurance mathematics or the asymptotic analysis of queuing models. An instance already encountered is Lindley's equation

$$W \stackrel{d}{=} (W+X)^+$$

for a random variable X with negative mean and independent of W. As explained in Section 1.2, the law of W equals the equilibrium distribution of a customer's waiting time in a G/G/1-queue before proceeding to the server if $X = B - A$, the difference of a generic service time B and a generic interarrival time A. If $(S_n)_{n\geq 0}$ denotes a SRW with increments $X_1, X_2, \ldots$ which are copies of X, then

$$W \stackrel{d}{=} \sup_{n\geq 0} S_n$$

as stated in Theorem 1.4 [☞ also Problem 1.6(b)]. The following classical result, which may already be found in FELLER's textbook [51, Ch. XII, (5.13)], provides the exact first-order asymptotics for $\mathbb{P}(W > t)$ as $t \to \infty$ under an exponential moment condition.

Theorem 2.68. *Let $(S_n)_{n\geq 0}$ be a SRW with negative drift μ, lattice-span $d \in \{0,1\}$, $\mathbb{E}e^{\vartheta S_1} = 1$ and $\mu_\vartheta := \mathbb{E}e^{\vartheta S_1^>} S_1^> \mathbf{1}_{\{\sigma^> < \infty\}} < \infty$ for some $\vartheta > 0$. Then*

$$d\text{-}\lim_{t\to\infty} e^{\vartheta t}\,\mathbb{P}\left(\sup_{n\geq 0} S_n > t\right) \;=\; \frac{\mathbb{P}(\sigma^> = \infty)}{e(d,\vartheta)\mu_\vartheta} \;\in\; \mathbb{R}_> \qquad (2.58)$$

with $e(d,\theta)$ as defined in (2.37)*. If $\mu_\vartheta = \infty$, the result remains true when interpreting the right-hand side of* (2.58) *as zero.*

For an alternative approach to this result via implicit renewal theory, we refer to Subsection 4.4.2. Let us further point out that the increments of $(S_n)_{n\ge 0}$ may take values in $\mathbb{R}\cup\{-\infty\}$ as one can readily see from the following proof. In this case, $e^{-\infty} := 0$ as usual.

Proof. By Prop. 2.59, $\mathbb{P}(\sigma^> = \infty) = (\mathbb{E}\sigma^{\le})^{-1} > 0$, for $\mu < \infty$ implies $\mathbb{E}\sigma^{\le} < \infty$. Further, $\mathbb{P}(\sigma^> = \infty) < 1$ and $\mu_\vartheta > 0$, for $\mathbb{E}e^{\vartheta S_1} = 1$ ensures $\mathbb{P}(S_1 > 0) > 0$. Consequently, $Q_> := \mathbb{P}(S_1^> \in \cdot, \sigma^> < \infty)$ is nonzero and defective, i.e. $0 < \|Q_>\| < 1$, and the associated renewal measure

$$\mathbb{U}^> \;=\; \sum_{n\ge 0}\mathbb{P}(S_n^> \in \cdot, \sigma_n^> < \infty) \;=\; \sum_{n\ge 0} Q_>^{*n}$$

[use Prop. 2.55(g) for the second equality] a finite measure.

Since $\mathbb{E}e^{\vartheta S_1} = 1$, the sequence $(e^{\vartheta S_n})_{n\ge 0}$ constitutes a nonnegative martingale with mean one. Let $(\mathscr{F}_n)_{n\ge 0}$ denote its natural filtration and $\mathscr{F}_\infty := \sigma(S_n : n \ge 0)$. Define a new probability measure $\widehat{\mathbb{P}}$ on $(\Omega, \mathscr{F}_\infty)$ by

$$\widehat{\mathbb{P}}(A) := \mathbb{E}e^{\vartheta S_n}\mathbf{1}_A \quad \text{for } A \in \mathscr{F}_n \text{ and } n \ge 0.$$

As one easily see, $X_1, X_2, \ldots$ are still iid under $\widehat{\mathbb{P}}$ with the same lattice-span d, common distribution $\widehat{Q}(B) := \mathbb{E}e^{\vartheta S_1}\mathbf{1}_B(S_1)$ for $B \in \mathscr{B}(\mathbb{R})$, and mean $\widehat{\mu} := \mathbb{E}e^{\vartheta S_1}S_1$. Equivalently, $(S_n)_{n\ge 0}$ is still a SRW with drift $\widehat{\mu}$. The latter is positive because $\phi(\theta) := \mathbb{E}e^{\theta S_1}$ is a convex function on $[0,\vartheta]$ with $\phi(0) = \phi(\vartheta) = 1$ and negative (right) derivative μ at 0. We further infer that

$$1 \;=\; \widehat{\mathbb{P}}(\sigma^> < \infty) \;=\; \sum_{n\ge 1}\widehat{\mathbb{P}}(\sigma^> = n) \;=\; \sum_{n\ge 1}\mathbb{E}e^{\vartheta S_n}\mathbf{1}_{\{\sigma^> = n\}} \;=\; \mathbb{E}e^{\vartheta S_1^>}.$$

Now observe that, with $g(x) := \mathbf{1}_{(-\infty,0)}(x)$,

$$\begin{aligned}\mathbb{P}\left(\sup_{n\ge 0} S_n > t\right) \;&=\; \sum_{n\ge 0}\mathbb{P}(S_n^> > t,\, \sigma_n^> < \infty,\, \sigma_{n+1}^> - \sigma_n^> = \infty)\\ &=\; \mathbb{P}(\sigma^> = \infty)\,\mathbb{U}^>((t,\infty))\\ &=\; \mathbb{P}(\sigma^> = \infty)\, g * \mathbb{U}^>(t)\end{aligned}$$

implying

$$e^{\vartheta t}\,\mathbb{P}\left(\sup_{n\ge 0} S_n > t\right) \;=\; \mathbb{P}(\sigma^> = \infty)\, g_\vartheta * \mathbb{U}_\vartheta^>(t),$$

where, as in Section 2.7, $g_\vartheta(x) = e^{\vartheta x} g(x)$ and $\mathbb{U}_\vartheta^>(dx) = e^{\vartheta x}\,\mathbb{U}^>(dx)$, the latter being the renewal measure of $(S_n^>)_{n\ge 0}$ under $\widehat{\mathbb{P}}$. Since g_ϑ is easily seen to be dRi, the assertion finally follows by an appeal to the key renewal theorem 2.67. □

Problems

Problem 2.69. Prove (2.50).

Problem 2.70. Prove that the cyclic decomposition formula remains true for any a.s. finite σ that is independent of $(S_n)_{n\ge 0}$ and called *randomized stopping time* for this RW. [Hint: Consider a SRP $(\sigma_n)_{n\ge 0}$ independent of $(S_n)_{n\ge 0}$ with $\mathscr{L}(\sigma_1) = \mathscr{L}(\sigma)$.]

Problem 2.71. Let $(S_n)_{n\ge 0}$ be a RW with positive drift in a standard model. As in the renewal case, put $N(A) := \sum_{n\ge 0} \mathbf{1}_A(S_n)$ for $A \in \mathscr{B}(\mathbb{R})$. Then

$$\sup_{t\in\mathbb{R}} \mathbb{P}_\lambda(N([t,t+h]) \ge n) \;\le\; \mathbb{P}_0(N([-h,h]) \ge n) \tag{2.59}$$

for all $h > 0$, $n \in \mathbb{N}_0$ and $\lambda \in \mathscr{P}(\mathbb{R})$. In particular,

$$\sup_{t\in\mathbb{R}} \mathbb{U}_\lambda([t,t+h]) \;\le\; \mathbb{U}_0([-h,h]) \tag{2.60}$$

and $\{N([t,t+h]) : t \in \mathbb{R}\}$ is uniformly integrable under each $\mathbb{P}_\lambda$ for all $h > 0$. [Hint: Generalize the proof of Lemma 2.14.]

Problem 2.72. Prove that the (one-sided) key renewal theorem 2.31 remains valid if g is dRi, but not necessarily vanishing on the negative halfline, and that $g * \mathbb{U}(t) \to 0$ as $t \to -\infty$ holds true in this case as well.

Chapter 3
Iterated random functions

This chapter is devoted to a rather short introduction of the general theory of iterations of iid random Lipschitz functions, also called *iterated function systems*. They may be viewed as a particular class of Markov chains on a topological state space for which stability results are usually deduced via appropriate contraction conditions on the occurring class of random functions.

3.1 The model, definitions, some basic observations and examples

In order to provide an appropriate framework for the subsequent considerations, we begin with a formal definition of the special class of Markov chains to be studied here. The relevance in connection with random recursive equations, the actual topic of this course, becomes immediately apparent by the way these chains are defined in (3.1) below.

Although all examples encountered so far have been Markov chains on $\mathbb{R}$ or $\mathbb{R}^m$, we have chosen to take a more general approach here by allowing the state space to be any complete separable metric space $(\mathbb{X},d)$ endowed with the Borel σ-field $\mathscr{B}(\mathbb{X})$. The reader will hopefully acknowledge that this appears to be quite natural and does not make our life more complicated. Nevertheless it may be useful to point out the following facts:

Convergence in distribution for random elements $X,X_1,X_2,...$ in $(\mathbb{X},\mathscr{B}(\mathbb{X}))$ is still defined in the usual manner, i.e. $X_n \stackrel{d}{\to} X$ if

$$\lim_{n\to\infty} \mathbb{E}f(X_n) \;=\; \mathbb{E}f(X) \quad \text{for all } f \in \mathscr{C}_b(\mathbb{X}),$$

where $\mathscr{C}_b(\mathbb{X})$ denotes the space of bounded continuous functions $f:\mathbb{X}\to\mathbb{R}$. Uniqueness of the limit distribution is guaranteed by the fact that this space is *measure-determining*, i.e., two bounded measures λ_1,λ_2 on $(\mathbb{X},\mathscr{B}(\mathbb{X}))$ are equal whenever

$$\int_{\mathbb{X}} f(x)\,\lambda_1(dx) \;=\; \int_{\mathbb{X}} f(x)\,\lambda_2(dx) \quad \text{for all } f \in \mathscr{C}_b(\mathbb{X}).$$

Finally, the Portmanteau theorem remains valid as well. For further information on convergence of probability measures on metric spaces we refer to the classic monograph by BILLINGSLEY [17].

3.1.1 *Definition of an iterated function system and its canonical model*

The formal definition of an iterated function system is first, followed by the discussion of some measurability aspects and the specification of a canonical model.

Definition 3.1. Let $(\mathbb{X},d)$ be a complete separable metric space with Borel-σ-field $\mathfrak{B}(\mathbb{X})$. A temporally homogeneous Markov chain $(X_n)_{n\ge 0}$ with state space $\mathbb{X}$ is called *iterated function system (IFS) of iid Lipschitz maps* if it satisfies a recursion of the form

$$X_n = \Psi(\theta_n, X_{n-1}) \tag{3.1}$$

for $n \ge 1$, where

(IFS-1) $X_0, \theta_1, \theta_2, \ldots$ are independent random elements on a common probability space $(\Omega, \mathfrak{A}, \mathbb{P})$;

(IFS-2) $\theta_1, \theta_2, \ldots$ are identically distributed with common distribution Λ and taking values in a measurable space $(\Theta, \mathscr{A})$;

(IFS-3) $\Psi : (\Theta \times \mathbb{X}, \mathscr{A} \otimes \mathscr{B}(\mathbb{X})) \to (\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is jointly measurable and Lipschitz continuous in the second argument, that is

$$d(\Psi(\theta,x), \Psi(\theta,y)) \le C_\theta\, d(x,y)$$

for all $x, y \in \mathbb{X}$, $\theta \in \Theta$ and a suitable $C_\theta \in \mathbb{R}_{\ge}$.

A natural way to generate an IFS is to first pick an iid sequence $\Psi_1, \Psi_2, \ldots$ of random elements from the space $\mathscr{C}_{Lip}(\mathbb{X})$ of Lipschitz self-maps on $\mathbb{X}$ and to then produce a Markov chain $(X_n)_{n\ge 0}$ by picking an initial value X_0 and defining

$$X_n \;=\; \Psi_n \circ \ldots \circ \Psi_1(X_0) \tag{3.2}$$

for each $n \ge 1$. In the context of the above definition, $\Psi_n = \Psi(\theta_n, \cdot)$, but it becomes a measurable object only if we endow $\mathscr{C}_{Lip}(\mathbb{X})$ with a suitable σ-field. Therefore we continue with a short description of what could be called the canonical model of an IFS which particularly meets the last requirement.

Let $\mathbb{X}_0 := \{x_1, x_2, ...\}$ be a countable dense subset of $\mathbb{X}$ and $\mathscr{L}(\mathbb{X}_0, \mathbb{X})$ the "sequence" space of all mappings from $\mathbb{X}_0$ to $\mathbb{X}$. The latter clearly forms a complete separable metric space, for instance, when choosing

$$\rho(\psi_1, \psi_2) = \sum_{n \geq 1} \frac{1}{2^n} \frac{d(\psi_1(x_n), \psi_2(x_n))}{1 + d(\psi_1(x_n), \psi_2(x_n))}$$

for $\psi_1, \psi_2 \in \mathscr{L}(\mathbb{X}_0, \mathbb{X})$ as a metric. We endow $\mathscr{L}(\mathbb{X}_0, \mathbb{X})$ with the product σ-field $\mathscr{B}(\mathbb{X})^{\mathbb{X}_0}$ generated by the product topology. Finally, we define the *Lipschitz constant of* ψ as

$$L(\psi) \ := \sup_{x,y \in \mathbb{X}, x \neq y} \frac{d(\psi(x), \psi(y))}{d(x,y)} \tag{3.3}$$

with the convention $L(\psi) := 0$ if ψ is constant. The following lemma is taken from [33].

Lemma 3.2. *Given the previous notation, the following assertions hold true:*

(a) $\mathscr{C}_{Lip}(\mathbb{X})$ *is a Borel subset of* $\mathscr{L}(\mathbb{X}_0, \mathbb{X})$.
(b) *The mapping* $\psi \mapsto L(\psi)$ *is a Borel function on* $\mathscr{C}_{Lip}(\mathbb{X})$.
(c) *The mapping* $(\psi, x) \mapsto \psi(x)$ *is a Borel function on* $\mathscr{C}_{Lip}(\mathbb{X}) \times \mathbb{X}$.

Proof. The map $L_0 : \mathscr{L}(\mathbb{X}_0, \mathbb{X}) \to [0, \infty]$, defined by

$$L_0(\psi) \ := \sup_{x,y \in \mathbb{X}_0, x \neq y} \frac{d(\psi(x), \psi(y))}{d(x,y)},$$

is clearly a Borel function, for it is the supremum of countably many continuous functions, namely

$$\mathscr{L}(\mathbb{X}_0, \mathbb{X}) \ \ni \ \psi \ \mapsto \ \frac{d(\psi(x), \psi(y))}{d(x,y)}$$

for $(x,y) \in \mathbb{X}_0^2$, $x \neq y$. Now observe that, if $L_0(\psi) < \infty$, then ψ has a unique extension to a Lipschitz function on $\mathbb{X}$ with $L(\psi) = L_0(\psi)$ because $\mathbb{X}_0$ is dense in $\mathbb{X}$. Conversely, the restriction of any Lipschitz continuous ψ to $\mathbb{X}_0$ satisfies $L_0(\psi) = L(\psi)$ whence we conclude

$$\mathscr{C}_{Lip}(\mathbb{X}) \ = \ \{\psi : L_0(\psi) < \infty\} \ \in \ \mathscr{B}(\mathbb{X})^{\mathbb{X}_0}$$

(by unique embedding) as well as the measurability of $\psi \mapsto L(\psi)$.

In oder to prove (c), let $x_1, x_2, ...$ be an enumeration of the elements of $\mathbb{X}_0$ and $B_\varepsilon(x)$ the open ε-ball with center x. For $n, k \in \mathbb{N}$, define

$$A_{n,1} := B_{1/n}(x_1) \quad \text{and} \quad A_{n,k} := B_{1/n}(x_k) \cap \bigcup_{j=1}^{k-1} B_{1/n}(x_j)^c \text{ for } k \geq 2.$$

Then each $(A_{n,k})_{k\geq 1}$ forms a measurable partition of $\mathbb{X}$. For any $\psi : \mathbb{X} \to \mathbb{X}$, put

$$\psi_n(x) \;:=\; \sum_{k\geq 1} \psi(x_k)\,\mathbf{1}_{A_{n,k}}(x) \quad \text{for } n \geq 1.$$

Then the mapping $(\psi,x) \mapsto \psi_n(x)$ is measurable from $\mathscr{L}(\mathbb{X}_0,\mathbb{X}) \times \mathbb{X}$ to $\mathbb{X}$ and its retraction to $\mathscr{C}_{Lip}(\mathbb{X}) \times \mathbb{X}$ converges pointwise to the evaluation map $(\psi,x) \mapsto \psi(x)$ which gives the desired result. □

In view of the previous lemma we can choose the following *canonical model* for an IFS of iid Lipschitz maps: Let $\mathscr{A}$ be the Borel σ-field on $\Theta := \mathscr{C}_{Lip}(\mathbb{X})$, more precisely $\mathscr{A} := \mathscr{B}(\mathbb{X})^{\mathbb{X}_0} \cap \mathscr{C}_{Lip}(\mathbb{X})$, and let $(X_0,\theta) = (X_0,\theta_1,\theta_2,...)$ be the identity map on the product space $(\mathbb{X} \times \mathscr{C}_{Lip}(\mathbb{X})^{\mathbb{N}}, \mathscr{B}(\mathbb{X}) \otimes \mathscr{A}^{\mathbb{N}})$, so that θ_n denotes the n^{th} projection for each n, taking values in $(\Theta,\mathscr{A})$. If we choose an infinite product distribution $\Lambda^{\mathbb{N}}$ on $(\mathscr{C}_{Lip}(\mathbb{X})^{\mathbb{N}}, \mathscr{A}^{\mathbb{N}})$ and $\mathbb{P}_x := \delta_x \otimes \Lambda^{\mathbb{N}}$ on $(\Omega,\mathfrak{A}) := (\mathbb{X} \times \mathscr{C}_{Lip}(\mathbb{X})^{\mathbb{N}}, \mathscr{B}(\mathbb{X}) \otimes \mathscr{A}^{\mathbb{N}})$, these projections are iid with common distribution F and independent of X_0 under any $\mathbb{P}_\lambda := \int \mathbb{P}_x\,\lambda(dx)$, $\lambda \in \mathscr{P}(\mathbb{X})$. Finally, define $\Psi : (\mathscr{C}_{Lip}(\mathbb{X}) \times \mathbb{X}, \mathscr{A} \otimes \mathscr{B}(\mathbb{X}))$ by $\Psi(\theta,x) := \theta(x)$ and $X_n := \Psi(\theta_n,X_0)$. Then

$$(\Omega,\mathfrak{A},(X_n)_{n\geq 0},(\mathbb{P}_\lambda)_{\lambda\in\mathscr{P}(\mathbb{X})}) \tag{3.4}$$

provides a canonical model for the IFS $(X_n)_{n\geq 0}$ of iid Lipschitz maps in which

$$\Psi_n := \Psi(\theta_n,\cdot), \quad n \geq 0$$

is a sequence of iid random elements in $\mathscr{C}_{Lip}(\mathbb{X})$ independent of X_0 under each $\mathbb{P}_\lambda$. As a Markov chain, $(X_n)_{n\geq 0}$ has one-step transition kernel

$$P(x,B) \;=\; \mathbb{P}_x(\Psi(\theta_1,X_0) \in B) \;=\; \Lambda(\Psi(\cdot,x) \in B), \quad B \in \mathscr{B}(\mathbb{X}), \tag{3.5}$$

which is easily seen to be Fellerian [☞ Problem 3.12]. In the following, we will always assume a standard model of the afore-mentioned type be given and write Ψ_n for $\Psi(\theta_n,\cdot)$, thus

$$X_n \;=\; \Psi_n(X_{n-1}) \;=\; \Psi_n \circ ...\Psi_1(X_0)$$

as already stated in (3.2). We further put $L_n := L(\Psi_n)$ for $n \geq 1$, by Lemma 3.2(b) a random variable taking values in $\mathbb{R}_\geq$, and note that $L_1, L_2, ...$ are iid under each $\mathbb{P}_\lambda$ with a distribution independent of λ. Therefore we use $\mathbb{P}$ for probabilities not depending on the initial distribution of the Markov chain.

3.1.2 Lipschitz constants, contraction properties and the top Liapunov exponent

In view of the fact that $(\mathscr{C}_{Lip}(\mathbb{X}),\circ)$ forms a multiplicative semigroup and thus $\Psi_{n:k} := \Psi_n \circ \ldots \circ \Psi_k \in \mathscr{C}_{Lip}(\mathbb{X})$ for any $n \geq 1$ and $1 \leq k \leq n$, it is natural to ask about how the Lipschitz constant $L(\Psi_{n:1})$ of $\Psi_{n:1}$ relates to those of its factors $\Psi_1,\ldots,\Psi_n$. The following simple lemma is basic for our analysis.

Lemma 3.3. *For any $\psi_1,\psi_2 \in \mathscr{C}_{Lip}(\mathbb{X})$,*

$$L(\psi_1 \circ \psi_2) \leq L(\psi_1)\cdot L(\psi_2).$$

Proof. Problem 3.13. □

As an immediate consequence of this lemma, we infer that

$$L(\Psi_{n:1}) \ \leq\ L(\Psi_{n:k+1})L(\Psi_{k:1}) \quad \text{for any } 1 \leq k < n \tag{3.6}$$

$$\text{and} \quad L(\Psi_{n:1}) \ \leq\ \prod_{k=1}^{n} L_k \quad \text{for any } n \geq 1. \tag{3.7}$$

An important consequence of (3.6) is that it entails the existence of the so-called *(top) Liapunov exponent* with the help of Kingman's subadditive ergodic theorem, the latter being stated without proof as Theorem A.5 in the Appendix. The following result is due to FURSTENBERG & KESTEN [56] for linear maps and to ELTON [44] for Lipschitz maps.

Theorem 3.4. [Furstenberg-Kesten, Elton] *Let $(X_n)_{n\geq 0}$ be an IFS of iid Lipschitz maps with Lipschitz constants $L_1,L_2,\ldots$ satisfying $\mathbb{E}\log^+ L_1 < \infty$. Then*

$$\lim_{n\to\infty} \frac{1}{n}\log L(\Psi_{n:1}) \ =\ \inf_{n\geq 1} \frac{\mathbb{E}\log L(\Psi_{n:1})}{n} \ =:\ \ell \quad a.s.$$

*where $\ell \in \mathbb{R}\cup\{-\infty\}$ is called **(top) Liapunov exponent** of $(X_n)_{n\geq 0}$. If ℓ is finite, then the convergence holds in L^1 as well.*

Proof. By (3.6), the triangular scheme $(Y_{k,n})_{n\geq 1}^{0\leq k<n}$, defined by

$$Y_{k,n} := \log L(\Psi_{n:k+1}),$$

is subadditive in the sense that $Y_{0,n} \leq Y_{0,k} + Y_{k,n}$ a.s. for all $0 \leq k < n$. It also satisfies all other conditions of the subadditive ergodic theorem A.5 in the Appendix as the reader can readily check, leading to the conclusion that

$$\lim_{n\to\infty}\frac{1}{n}\log L(\Psi_{n:1}) \;=\; Y_\infty \quad \text{a.s.}$$

for some random variable Y_∞ with mean ℓ if ℓ is finite, and $Y_\infty = -\infty$ otherwise. And in the first case, the convergence is also in L^1. Finally, the fact that Y_∞ actually a.s. equals its mean value ℓ follows by an appeal to the Kolmogorov zero-one law, for Y_∞ is measurable with respect to the terminal σ-field of $\Psi_1, \Psi_2, \ldots$ □

A combination of the previous result with (3.7) and the SLLN further provides us with:

Corollary 3.5. *Let $(X_n)_{n\ge 0}$ be an IFS of iid Lipschitz maps with Lipschitz constants $L_1, L_2, \ldots$ satisfying $\mathbb{E}\log^+ L_1 < \infty$. Then its Liapunov exponent ℓ satisfies*

$$\ell \;\le\; \mathbb{E}\log L_1. \tag{3.8}$$

Proof. It suffices to note that $(\sum_{k=1}^n \log L_k)_{n\ge 0}$ forms a SRW with drift $\mathbb{E}\log L_1$ and that $\log L(\Psi_{n:1}) \le \sum_{k=1}^n \log L_k$ for each $n \ge 1$. □

It should not be surprising that the Lipschitz constants $L(\Psi_{n:1})$ play an important role in the stability analysis of $(X_n)_{n\ge 0}$. This will already become quite clear in the next section when studying *strongly contractive IFS* to be defined below along with other contraction conditions. Recall that a Lipschitz map ψ is called *contractive* or a *contraction* if $L(\psi) < 1$.

Definition 3.6. An IFS $(X_n)_{n\ge 0}$ of iid Lipschitz maps is called

– *strongly contractive*	if $\log L_1 \le -l$ a.s. for some $l \in \overline{\mathbb{R}}_>$.	
– *strongly mean contractive of order p*	if $\log \mathbb{E}L_1^p < 0$.	$(p > 0)$
– *mean contractive*	if $\mathbb{E}\log L_1 < 0$.	
– *contractive*	if it has Liapunov exponent $\ell < 0$.	

It is obvious that strong contraction implies contraction and strong mean contraction of any order, while an application of Jensen's inequality shows that the latter implies mean contraction; for a converse see Problem 3.14. Moreover, strong mean contraction of order p may always be reduced to the case $p = 1$ by switching the metric [☞ Problem 3.15].

3.1.3 Forward versus backward iterations

The recursive character of an IFS naturally entails that its state X_n at any time n is obtained via *forward iteration* or *left multiplication* of the random Lipschitz functions $\Psi_1, \ldots, \Psi_n$. This means, we first apply Ψ_1 to X_0, then Ψ_2 to $\Psi_1(X_0)$, and so on until we finally apply Ψ_n to $\Psi_{n-1}\circ\ldots\circ\Psi_1(X_0)$. On the other hand, since

$$(\Psi_1,\ldots,\Psi_n) \stackrel{d}{=} (\Psi_n,\ldots,\Psi_1),$$

the distribution of the forward iteration X_n is at all times n the same as of the *backward iteration* or *right multiplication* $\widehat{X}_n := \Psi_1\circ\ldots\circ\Psi_n(X_0)$, that is

$$\widehat{X}_n \stackrel{d}{=} X_n \quad \text{for all } n\ge 0. \tag{3.9}$$

Consequently, we may also study the sequence of backward iterations $(\widehat{X}_n)_{n\ge 0}$ when trying to prove asymptotic stability of $(X_n)_{n\ge 0}$, i.e. $X_n \stackrel{d}{\to} \pi$ for some $\pi\in\mathscr{P}(\mathbb{X})$. The usefulness of this observation relies on the fact that in the stable case the $\widehat{X}_n$ exhibit a stronger pathwise convergence as we will see, which particularly shows that the joint distributions of $(X_n)_{n\ge 0}$ and $(\widehat{X}_n)_{n\ge 0}$ are generally very different. Most notably, $(\widehat{X}_n)_{n\ge 0}$ is not a Markov chain except for trivial cases.

In the following, we put $\Psi_{k:n} := \Psi_k\circ\ldots\circ\Psi_n$ for $1\le k\le n$ and note as direct counterparts of (3.6) and (3.7) that

$$L(\Psi_{1:n}) \;\le\; L(\Psi_{1:k})\,L(\Psi_{k+1:n}) \quad \text{for any } 1\le k<n \tag{3.10}$$

$$\text{and} \quad L(\Psi_{1:n}) \;\le\; \prod_{k=1}^{n} L_k \quad \text{for any } n\ge 1. \tag{3.11}$$

Also Theorem 3.4 and its corollary remain valid when replacing $L(\Psi_{n:1})$ with $L(\Psi_{1:n})$ in (3.8).

3.1.4 Examples

At the end of this section we present a collection of examples some of which we have already encountered before.

Example 3.7 (Random difference equations). Iterations of iid linear functions $\Psi_n(x) = M_n x + Q_n$ on $\mathbb{R}$, with Lipschitz constants $L_n = |M_n|$, constitute one of the basic examples of an IFS and lead back to the one-dimensional random difference equation (RDE)

$$X_n \;:=\; M_n X_{n-1} + Q_n, \quad n\ge 1, \tag{3.12}$$

discussed in Section 1.5. Recall from there that

$$X_n \;=\; \Psi_{n:1}(X_0) \;=\; \Pi_n X_0 + \sum_{k=1}^{n} \Pi_{k+1:n} Q_k \tag{3.13}$$

for each $n\in\mathbb{N}$, where $\Pi_{k:n} := M_k\cdot\ldots\cdot M_n$ for $1\le k\le n$, $\Pi_{n+1:n} := 1$ and $\Pi_n := \Pi_{1:n}$, which shows that

$$L(\Psi_{n:1}) \;=\; |\Pi_n| \;=\; \prod_{k=1}^{n} L_k$$

for each $n \in \mathbb{N}$. The distributional equality (3.9) of forward and backward iteration at any time n was also stated there in (1.32), viz.

$$X_n \;\stackrel{d}{=}\; \Pi_n X_0 + \sum_{k=1}^{n} \Pi_{k-1} Q_k \;=\; \Psi_{1:n}(X_0) \;=\; \widehat{X}_n.$$

From these facts we see that $(X_n)_{n\ge 0}$ is

– *strongly contractive*	if $\log|M_1| \le -l$ a.s. for some $l \in \overline{\mathbb{R}}_>$.	
– *strongly mean contractive of order p*	if $\log \mathbb{E}|M_1|^p < 0$.	$(p > 0)$
– *mean contractive*	if $\mathbb{E}\log|M_1| < 0$.	

In the multivariate case, the RDE (3.12) is defined on $\mathbb{R}^m$ for some $m \ge 2$ with $X_0, X_1, \dots$ and $Q_1, Q_2, \dots$ being column vectors and $M_1, M_2, \dots$ being $m \times m$ real matrices. For $x \in \mathbb{R}^m$ and a $m \times m$ matrix A, let $|x|$ be the usual Euclidean norm and

$$\|A\| \;:=\; \max_{|x|=1} |Ax|$$

the usual operator norm of A. Contraction conditions must now be stated in terms of $\|M_1\|$ but look the same as before otherwise. Hence, $(X_n)_{n\ge 0}$ is

– *strongly contractive*	if $\log\|M_1\| \le -l$ a.s. for some $l \in \overline{\mathbb{R}}_>$.	
– *strongly mean contractive of order p*	if $\log \mathbb{E}\|M_1\|^p < 0$.	$(p > 0)$
– *mean contractive*	if $\mathbb{E}\log\|M_1\| < 0$.	

Example 3.8 (Lindley processes). Lindley processes were introduced in Section 1.2 in connection with the G/G/1-queue and have the general form

$$X_n \;:=\; (X_{n-1} + \xi_n)^+, \quad n \ge 1, \tag{3.14}$$

for a sequence $(\xi_n)_{n\ge 1}$ of iid real-valued random variables which are not a.s. vanishing. This is an example of an IFS of iid Lipschitz functions on $\mathbb{X} = \mathbb{R}_\ge$ having Lipschitz constants $L_n = 1$, namely $\Psi_n(x) := (x + \xi_n)^+$ for $n \ge 1$. Denote by $(S_n)_{n\ge 0}$ the SRW associated with the ξ_n, thus $S_0 = 0$ and $S_n = \xi_1 + \dots + \xi_n$ for $n \ge 1$. Forward and backward iterations are easily computed as

$$\begin{aligned} X_n \;&=\; \Psi_{n:1}(X_0) \;=\; \max\{0, S_n - S_{n-1}, S_n - S_{n-2}, \dots, S_n - S_1, X_0 + S_n\} \\ \text{and} \quad \widehat{X}_n \;&=\; \Psi_{1:n}(X_0) \;=\; \max\{0, S_1, S_2, \dots, S_{n-1}, X_0 + S_n\}, \end{aligned}$$

and the latter sequence is obviously nondecreasing. Furthermore, it converges a.s. to a finite limit, viz. $\widehat{X}_\infty := \sup_{n\ge 0} S_n$, iff $(S_n)_{n\ge 0}$ is negatively divergent, which particularly holds if $\mathbb{E}\xi < 0$ [☞ Thm. 1.4]. In this case, X_n converges to the same limit in distribution by (3.9). Notice that despite this stability result none of the above contraction conditions is valid, for $L_1 = 1$.

Example 3.9 (AR(1)-model with ARCH(1) errors). As another recurring example of an IFS we mention the AR(1)-model with ARCH(1) errors, defined by

$$X_n \;=\; \alpha X_{n-1} + \left(\beta + \lambda X_{n-1}^2\right)^{1/2} \varepsilon_n, \quad n \ge 1, \tag{3.15}$$

for $(\alpha, \beta, \lambda) \in \mathbb{R} \times \mathbb{R}_>^2$ and a sequence $(\varepsilon_n)_{n\ge1}$ of iid symmetric random variables. Here $\Psi_n(x) := \alpha x + (\beta + \lambda x^2)^{1/2} \varepsilon_n$ for $n \ge 1$ and therefore

$$\frac{|\Psi_n(x) - \Psi_n(y)|}{|x-y|} \;=\; \left|\alpha + \frac{\lambda^{1/2}(x+y)\varepsilon_n}{(\beta/\lambda + x^2)^{1/2} + (\beta/\lambda + y^2)^{1/2}}\right| \;\le\; \alpha + \lambda^{1/2}|\varepsilon_n|$$

for all $x, y \in \mathbb{R}$. By combining this with

$$\lim_{x,y\to\pm\infty} \frac{|\Psi_n(x) - \Psi_n(y)|}{|x-y|} \;=\; |\alpha \pm \lambda^{1/2}\varepsilon_n|,$$

it follows easily that $L(\Psi_n) = \alpha + \lambda^{1/2}|\varepsilon_n|$. Hence, the IFS is mean contractive if $\mathbb{E}\log(\alpha + \lambda^{1/2}|\varepsilon|) < 0$. On the other hand, neither forward nor backward iterations are easily computed here so that stability can only be analyzed by more sophisticated tools than in the previous two examples.

Example 3.10 (Random logistic maps). The *logistic map* $x \mapsto \theta x(1-x)$ is a self-map of the unit interval $[0,1]$ if $0 \le \theta \le 4$. Therefore, we obtain an IFS of i.i.d. Lipschitz functions on $[0,1]$ by defining

$$X_n \;=\; \xi_n X_{n-1}(1 - X_{n-1}), \quad n \ge 1, \tag{3.16}$$

for a sequence $(\xi_n)_{n\ge1}$ of iid random variables taking values in $[0,4]$. Hence $\Psi_n(x) = \xi_n x(1-x)$, which has Lipschitz constant $L_n = \xi_n$ as one can easily verify. Contraction conditions as introduced before are thus to be formulated in terms of moments of ξ, but it should be noted that the Markov chain $(X_n)_{n\ge0}$ possesses a stationary distribution in any case by Lemma 1.23 because the state space is compact which trivially ensures tightness of $(X_n)_{n\ge0}$. In fact, if the chain is mean contractive, i.e. $\mathbb{E}\log\xi < 0$ holds, then X_n converges a.s. to zero under any initial distribution and geometrically fast [☞ Problem 3.15] which appears to be fairly boring and leaves us with the real challenge to find out what happens if mean contraction fails to hold and to provide conditions under which a stationary distribution $\pi \ne \delta_0$ exists. These questions have been addressed in a number of articles by DAI [30], STEINSALTZ [103], ATHREYA & DAI [8, 9], and ATHREYA & SCHUH [10]. We will return to this question in Subsection **??**.

Problems

Problem 3.11. Let $(X_n)_{n\ge 0}$ be an IFS of iid Lipschitz maps in a canonical model as stated in (3.4). Prove that, under each $\mathbb{P}_\lambda$ and for each $n \in \mathbb{N}_0$, X_n and $(\Psi_k)_{k>n}$ are independent with

$$\mathbb{P}_\lambda(X_n \in \cdot, (\Psi_k)_{k>n} \in \cdot) \;=\; \mathbb{P}_{\lambda_n}(X_0 \in \cdot, (\Psi_k)_{k\ge 1} \in \cdot),$$

where $\lambda_n := \mathbb{P}_\lambda(X_n \in \cdot)$.

Problem 3.12. Prove that the transition kernel P defined in (3.5) is Fellerian, i.e. it maps bounded continuous functions on $\mathbb{X}$ to functions of the same type [☞ Section 1.6].

Problem 3.13. Prove Lemma 3.3.

Problem 3.14. Let $(X_n)_{n\ge 0}$ be a mean contractive IFS of iid Lipschitz maps. Prove that, if $\mathbb{E}L_1^p < \infty$ for some $p > 0$, then $(X_n)_{n\ge 0}$ is also strongly mean contractive of some order $q \le p$, i.e. $\mathbb{E}L_1^q < 1$.

Problem 3.15. Let $(X_n)_{n\ge 0}$ be an IFS of iid Lipschitz maps which is strongly mean contractive of order $p \ne 1$. Prove that there exists a complete separable metric d' on $\mathbb{X}$ generating the same topology as d such that $(X_n)_{n\ge 0}$ is strongly mean contractive of order one under d', that is, when using the Lipschitz constants defined with the help of d'.

Problem 3.16. Consider the IFS $(X_n)_{n\ge 0}$ of random logistic maps introduced in Example 3.10. Prove that, if $\mathbb{E}\log\xi < 0$, then $\mu^{-n}X_n \to 0$ a.s. for any $\mu < 1$ such that $\log\mu > \mathbb{E}\log\xi$. What happens if $\mathbb{E}\log\xi = 0$?

3.2 Geometric ergodicity of strongly contractive IFS

Aiming at an ergodic theorem for mean contractive IFS, we first study the simpler strongly contractive case for which we are going to prove geometric ergodicity, i.e. convergence to a stationary distribution at a geometric rate (under a mild moment condition). The distance between probability distributions on $\mathbb{X}$ is measured by the *Prokhorov metric* associated with d (the metric on $\mathbb{X}$) and denoted by the same letter. Given $\lambda_1, \lambda_2 \in \mathscr{P}(\mathbb{X})$, it is defined as the infimum of the $\delta > 0$ such that

$$\lambda_1(A) \le \lambda_2(A^\delta) + \delta \quad \text{and} \quad \lambda_2(A) \le \lambda_1(A^\delta) + \delta$$

for all $A \in \mathscr{B}(\mathbb{X})$, where $A^\delta := \{x \in \mathbb{X} : d(x,y) < \delta \text{ for some } y \in A\}$. We note that $d(\lambda_1, \lambda_2) \le 1$ and without proof that convergence in the Prokhorov metric is equivalent to weak convergence, that is

$$d(\lambda_n, \lambda) \to 0 \quad \text{iff} \quad \lambda_n \xrightarrow{w} \lambda.$$

The following simple coupling lemma provides a useful tool to derive an estimate for $d(\lambda_1,\lambda_2)$.

Lemma 3.17. *Let X_1,X_2 be two $\mathbb{X}$-valued random elements on the same probability space $(\Omega,\mathfrak{A},\mathbb{P})$ such that $\mathscr{L}(X_1)=\lambda_2$ and $\mathscr{L}(X_2)=\lambda_2$. Then $\mathbb{P}(d(X_1,X_2)\ge\delta)<\delta$ implies $d(\lambda_1,\lambda_2)\le\delta$.*

Proof. The assertion follows from the obvious inequality

$$\max\{\mathbb{P}(X_1\in A,X_2\notin A^{\delta}),\mathbb{P}(X_1\notin A^{\delta},X_2\in A)\}\ \le\ \mathbb{P}(d(X_1,X_2)\ge\delta)$$

for all $A\in\mathscr{B}(\mathbb{X})$ and $\delta>0$. □

After these preliminary remarks we are ready to state the announced ergodic theorem.

Proposition 3.18. *Given a strongly contractive IFS $(X_n)_{n\ge0}$ of iid Lipschitz maps in a standard model such that $\log L_1\le -l$ for some $l\in\mathbb{R}_>$ and*

$$\mathbb{E}\log^+ d(x_0,\Psi_1(x_0))<\infty, \tag{3.17}$$

for some $x_0\in\mathbb{X}$, the following assertions hold true:

(a) For any $x\in\mathbb{X}$, the backward iteration $\widehat{X}_n$ converges $\mathbb{P}_x$-a.s. to a random element $\widehat{X}_\infty$ with distribution π which does not depend on x and satsifies the SFPE

$$\widehat{X}_\infty=\Psi_1(\widehat{X}'_\infty) \tag{3.18}$$

where $\widehat{X}'_\infty$ is a copy of $\widehat{X}_\infty$ independent of Ψ_1.

(b) For any $x\in\mathbb{X}$ and $\gamma\in(1,e^l)$,

$$\lim_{n\to\infty}\gamma^n d(\widehat{X}_n,\widehat{X}_\infty)=0\quad \mathbb{P}_x\text{-a.s.} \tag{3.19}$$

(c) For any $x\in\mathbb{X}$, the forward iteration X_n converges to π in distribution under $\mathbb{P}_x$, and π is the unique stationary distribution of $(X_n)_{n\ge0}$.

(d) Under $\mathbb{P}_\pi$, $(X_n)_{n\ge0}$ forms an ergodic stationary sequence, i.e.

$$\mathbf{1}_B(X_0,X_1,\ldots)=\mathbf{1}_B(X_1,X_2,\ldots)\quad \mathbb{P}_\pi\text{-a.s.}$$

implies $\mathbb{P}_\pi((X_n)_{n\ge0}\in B)\in\{0,1\}$.

(e) If (3.17) is sharpened to

$$\mathbb{E}\,d(x_0,\Psi_1(x_0))^p<\infty, \tag{3.20}$$

for some $p > 0$ and $x_0 \in \mathbb{X}$, then geometric ergodicity in the sense that

$$\lim_{n\to\infty} r^n\, d(\mathbb{P}_x(X_n \in \cdot), \pi) = 0 \tag{3.21}$$

for some $r > 1$ holds true.

Before we turn to the proof of this result, some comments are in order.

Remark 3.19. A fundamental conclusion from this result is that forward and backward iterations, despite having the same one-dimensional marginals, exhibit a drastically different behavior. While backward iterations converge a.s. at a geometric rate to a limit having distribution π, the convergence of the forward iterations, naturally to the same limit, occurs in the distributional sense only. Their trajectories, however, typically oscillate wildly in space due to the ergodicity which ensures that every π-positive subset is visited infinitely often. This is illustrated in Figure 3.1 below.

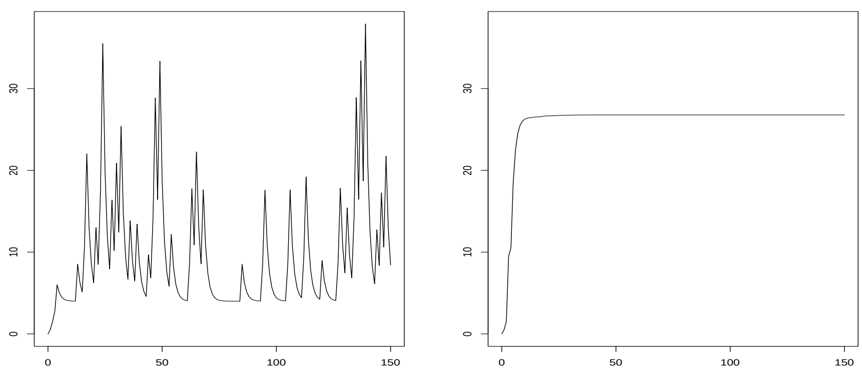


Fig. 3.1 Ergodic behavior of the forward iterations (left panel) versus pathwise convergence of the backward iterations (right panel), illustrated by a simulation of 150 iterations of the IFS which picks the function $\psi_1(x) = 0.5x + 2$ with probability 0.75 and the function $\psi_2(x) = 2x + 0.5$ with probability 0.25 at each step.

Remark 3.20. The extra moment conditions (3.17) – in which $\log^+ x$ may be replaced with the subadditive majorant $\log^* x := \log(1+x)$ – and (3.20), frequently called *jump-size conditions* hereafter, are needed beside strong contraction to ensure that the chain is not carried away too far in one step when moving in space. The reader should realize that this is indeed a property not guaranteed by contraction, which rather ensures forgetfulness of initial conditions. Let us further point out that, if any of these jump-size conditions is valid for one $x_0 \in \mathbb{X}$, then it actually holds for all $x_0 \in \mathbb{X}$. This follows from

$$\begin{aligned} d(x,\Psi_1(x)) &\le d(x,x_0)+d(x_0,\Psi_1(x_0))+d(\Psi_1(x_0),\Psi_1(x)) \\ &\le (1+L_1)\,d(x,x_0)+d(x_0,\Psi_1(x_0)) \\ &\le (1+e^{-l})\,d(x,x_0)+d(x_0,\Psi_1(x_0)) \end{aligned}$$

for all $x \in \mathbb{X}$.

In the following we will use $\mathbb{P}$ for probabilities that do not depend on the initial distribution of $(X_n)_{n\ge 0}$. For instance,

$$\mathbb{P}_x(X_n \in \cdot) \;=\; \mathbb{P}(\Psi_{n:1}(x) \in \cdot)$$

because $\Psi_1, \Psi_2, \ldots$ are independent of X_0. Note also that, for any $x, y \in \mathbb{X}$,

$$(X_n^x, X_n^y) := (\Psi_{n:1}(x), \Psi_{n:1}(y)), \quad n \ge 0$$

provides a canonical coupling of two forward chains starting at x and y. A similar coupling is naturally given by $(\widehat{X}_n^x, \widehat{X}_n^y) := (\Psi_{1:n}(x), \Psi_{1:n}(y))$, $n \ge 0$, for the backward iterations.

Proof (of Prop. 3.18). We leave it as an exercise [☞ Problem 3.21] to verify that

$$\gamma^n d(\widehat{X}_n, \widehat{X}_\infty) \to 0 \quad \mathbb{P}_x\text{-a.s.}$$

for $\gamma \ge 1$ and a random variable $\widehat{X}_\infty$ if

$$\sum_{n\ge 0} \gamma^n\, d(\widehat{X}_n, \widehat{X}_{n+1}) < \infty \quad \mathbb{P}_x\text{-a.s.} \tag{3.22}$$

Strong contraction implies for any $\gamma \in [1, e^l)$

$$\begin{aligned} \gamma^n\, d(\widehat{X}_n, \widehat{X}_{n+1}) &= \gamma^n\, d(\Psi_{1:n}(X_0), \Psi_{1:n} \circ \Psi_{n+1}(X_0)) \\ &\le \beta^n\, d(X_0, \Psi_{n+1}(X_0)) \\ &= \beta^n\, d(x, \Psi_{n+1}(x)) \quad \mathbb{P}_x\text{-a.s.} \end{aligned}$$

for any $x \in \mathbb{X}$, where $\beta := \gamma e^{-l} \in (0,1]$. Now use (3.17), by Remark 3.20 valid for any $x \in \mathbb{X}$, to infer

$$\begin{aligned} \sum_{n\ge 0} \mathbb{P}_x(\gamma^n\, d(\widehat{X}_n, \widehat{X}_{n+1}) > \varepsilon) &\le \sum_{n\ge 0} \mathbb{P}(\beta^n\, d(x, \Psi_{n+1}(x)) > \varepsilon) \\ &= \sum_{n\ge 0} \mathbb{P}(\log d(x, \Psi_1(x)) > \log \varepsilon + n \log(1/\beta)) \\ &\le \frac{\log(1/\varepsilon) + \mathbb{E} \log^+ d(x, \Psi_1(x))}{\log(1/\beta)} \;<\; \infty \end{aligned}$$

for any $\varepsilon > 0$ and thus $Z_n(\gamma) := \gamma^n\, d(\widehat{X}_n, \widehat{X}_{n+1}) \to 0$ $\mathbb{P}_x$-a.s. for any $\gamma \in [1, e^l)$ by an appeal to the Borel-Cantelli lemma. But the last conclusion further implies (3.22), because

$$\sum_{n\ge 0}\gamma^n d(\widehat{X}_n,\widehat{X}_{n+1}) \;\le\; \sum_{n\ge 0}\left(\frac{\gamma}{\beta}\right)^n Z_n(\beta) \;<\; \infty$$

for any $1\le\gamma<\beta<e^l$. This completes the proof of (b) and the first part of (a). As

$$d(\widehat{X}_n^x,\widehat{X}_n^y) \;=\; d(\Psi_{1:n}(x),\Psi_{1:n}(y)) \;\le\; e^{-nl}d(x,y) \quad \mathbb{P}\text{-a.s.}$$

we see that $\widehat{X}_\infty$ and its distribution are the same under every $\mathbb{P}_x$. Also, $(\Psi_{2:n})_{n\ge 2}$ being a copy of $(\Psi_{1:n})_{n\ge 1}$, we find that $\Psi_{2:n}(x)$ converges a.s. to some $\widehat{X}_\infty'$ not depending on x and with the same law as $\widehat{X}_\infty$. Finally, the asserted SFPE (3.18) follows from

$$\widehat{X}_\infty \;=\; \lim_{n\to\infty}\Psi_1(\Psi_{2:n}(x)) \;=\; \Psi_1\left(\lim_{n\to\infty}\Psi_{2:n}(x)\right) \;=\; \Psi_1(\widehat{X}_\infty') \quad \mathbb{P}\text{-a.s.}$$

where the continuity of Ψ_1 enters in the second equality. This completes the proof of (a).

As for part (c), the first assertion is obvious from (a) because $X_n \stackrel{d}{=} \widehat{X}_n$ under each $\mathbb{P}_x$ and therefore each $\mathbb{P}_\lambda$, $\lambda\in\mathscr{P}(\mathbb{X})$. But this also implies that π must be the unique stationary distribution of $(X_n)_{n\ge 0}$. Indeed, any stationary π' satisfies

$$\int f(x)\,\pi'(dx) \;=\; \mathbb{E}_{\pi'}f(X_n) \;\stackrel{n\to\infty}{\longrightarrow}\; \int f(x)\,\pi(dx)$$

for all $f\in\mathscr{C}_b(\mathbb{X})$ and thus $\pi'=\pi$ because the class $\mathscr{C}_b(\mathbb{X})$ is measure-determining.

The proof of (d) forces us to make an excursion into ergodic theory and follows the argument given by ELTON [44, p. 43]. First of all, we may w.l.o.g. extend $(\Psi_n)_{n\ge 0}$ to a *doubly infinite* sequence $(\Psi_n)_{n\in\mathbb{Z}}$ of iid Lipschitz maps. This sequence is ergodic [☞ Prop. A.1], which in the terminology of ergodic theory means that the shift $\mathscr{S}_1:(\ldots,\psi_{-1},\psi_0,\psi_1,\ldots)\mapsto(\ldots,\psi_0,\psi_1,\psi_2,\ldots)$ consitutes a *measure-preserving ergodic transformation* on $(\mathscr{C}_{Lip}(\mathbb{X})^{\mathbb{Z}},\mathscr{A}^{\mathbb{Z}},\Lambda^{\mathbb{Z}})$. Next, fix any $x\in\mathbb{X}$ and define the doubly infinite stationary sequence

$$Y_n := \lim_{k\to\infty}\Psi_{-k:n}(x), \quad n\in\mathbb{Z}$$

which is clearly a function φ, say, of $(\Psi_n)_{n\in\mathbb{Z}}$ and does not depend on the choice of x (by part (a)). Let Γ denote its distribution and notice that $Y_1=\widehat{X}_\infty$ as well as $\mathbb{P}((Y_n)_{n\ge 0}\in\cdot)=\mathbb{P}_\pi((X_n)_{n\ge 0}\in\cdot)$. The stationarity of $(Y_n)_{n\in\mathbb{Z}}$ means that the shift $\mathscr{S}_2:(\ldots,x_{-1},x_0,x_1,\ldots)\mapsto(\ldots,x_0,x_1,x_2,\ldots)$ is a measure-preserving transformation on $(\mathbb{X}^{\mathbb{Z}},\mathscr{B}(\mathbb{X})^{\mathbb{Z}},\Gamma)$. Now the ergodicity of $\mathscr{S}_2$, and thus of $(Y_n)_{n\in\mathbb{Z}}$, follows because $\mathscr{S}_2$ *is a factor of* $\mathscr{S}_1$, viz. $\varphi\circ\mathscr{S}_1=\mathscr{S}_2\circ\varphi$ a.s. for the measure-preserving map $\varphi:(\mathscr{C}_{Lip}(\mathbb{X})^{\mathbb{Z}},\mathscr{A}^{\mathbb{Z}},\Lambda^{\mathbb{Z}})\to(\mathbb{X}^{\mathbb{Z}},\mathscr{B}(\mathbb{X})^{\mathbb{Z}},\Gamma)$, defined by $(\psi_n)_{n\in\mathbb{Z}}\mapsto(\limsup_{k\to\infty}\psi_{-k:n}(x))_{n\in\mathbb{Z}}$ [☞ Prop. A.2 in the Appendix and before for further information].

Turning to (e), assume (3.20) for some $p>0$, w.l.o.g. $p\le 1$. Then it follows with the help of the subadditivity of $x\mapsto x^p$ on $\mathbb{R}_\ge$ that, for any $s>0$,

$$\begin{aligned}
\mathbb{P}_x(d(\widehat{X}_n,\widehat{X}_\infty)\ge s^{-n}) &\le s^{np}\,\mathbb{E}_x\,d(\widehat{X}_n,\widehat{X}_\infty)^p\\
&\le s^{np}\sum_{k\ge n}\mathbb{E}_x\,d(\widehat{X}_k,\widehat{X}_{k+1})^p\\
&= s^{np}\sum_{k\ge n}\mathbb{E}\,d(\Psi_{1:k}(x),\Psi_{1:k+1}(x))^p\\
&\le (se^{-l})^{np}\,\mathbb{E}\,d(x,\Psi_1(x))^p\sum_{k\ge 0}e^{-klp}\\
&= (se^{-l})^{np}\,\frac{\mathbb{E}\,d(x,\Psi_1(x))^p}{1-e^{-lp}}.
\end{aligned}$$

The last expression is ultimately bounded by $o(1)\,s^{-n}$ as $n\to\infty$ if $(se^{-l})^p<s^{-1}$ or, equivalently, $1<s<e^{ql}$ with $q:=p/(p+1)$. Hence, for any such s we have shown that

$$\lim_{n\to\infty}s^n\,\mathbb{P}_x(d(\widehat{X}_n,\widehat{X}_\infty)\ge s^{-n})=0,$$

and this entails (3.21) for $r\in(1,s)$ by an appeal to Lemma 3.17, for $(\widehat{X}_n,\widehat{X}_\infty)$ constitutes a coupling of $\mathbb{P}_x(X_n\in\cdot)$ and π under $\mathbb{P}_x$. ☐

Problems

Problem 3.21. Given a sequence $(X_n)_{n\ge 0}$ of random variables taking values in a complete metric space $(\mathbb{X},d)$, prove that

$$\sum_{n\ge 0}d(X_n,X_{n+1})<\infty\quad \mathbb{P}\text{-a.s.}$$

implies the a.s. convergence of X_n to a random variable X_∞. More generally, if

$$\sum_{n\ge 0}a_n\,d(X_n,X_{n+1})<\infty\quad \mathbb{P}\text{-a.s.}$$

holds true for a nondecreasing sequence $(a_n)_{n\ge 0}$ in $\mathbb{R}_>$, then

$$\lim_{n\to\infty}a_n\,d(X_n,X_\infty)=0.$$

Problem 3.22. The proof of part (e) of Prop. 3.18 has shown that (3.21) holds true for any $r\in(1,e^{pl/(p+1)})$, provided that $p\le 1$ in the jump-size condition (3.20). Show that this remains valid in the case $p>1$ as well. [Hint: Show first that

$$\left(\mathbb{E}_x d(\widehat{X}_n,\widehat{X}_\infty)^p\right)^{1/p}\ \le\ e^{-ln}\,\frac{\left(\mathbb{E}\,d(x,\Psi_1(x))^p\right)^{1/p}}{1-e^{-l}}$$

and then argue as in the afore-mentioned proof.]

Problem 3.23. Suppose that $\Psi, \Psi_1, \Psi_2, \ldots$ are iid Lipschitz maps on a complete metric space $(\mathbb{X}, d)$ such that, for

$$\alpha := \mathbb{P}(\Psi \equiv x_0) > 0$$

for some $x_0 \in \mathbb{X}$. Define $\sigma := \inf\{n \geq 1 : \Psi_n \equiv x_0\}$ and then the pre-σ-occupation measure

$$\pi(A) \;:=\; \mathbb{E}\left(\sum_{n=0}^{\sigma-1} \mathbf{1}_A\big(\Psi_{n:1}(x_0)\big)\right), \quad A \in \mathscr{A},$$

where $\Psi_{0:1}(x) = x$. Show that π is the unique stationary distribution of the strongly contractive IFS generated by $(\Psi_n)_{n\geq 1}$.

3.3 Ergodic theorem for mean contractive IFS

We will now proceed with the main result of this chapter, an ergodic theorem for mean contractive IFS of iid Lipschitz maps. The basic idea for its proof is taken from [3] and combines our previous result for strongly contractive IFS with a renewal argument based on the observation that any weakly contractive IFS contains a strongly contractive one. Before dwelling on this further, let us state the result we are going to prove in this section.

Theorem 3.24. *Given a mean contractive IFS $(X_n)_{n\geq 0}$ of iid Lipschitz maps in a standard model which also satisfies the jump-size condition* (3.17) *for some and thus all $x_0 \in \mathbb{X}$, the following assertions hold true:*

(a) *For any $x \in \mathbb{X}$, the backward iteration $\widehat{X}_n$ converges $\mathbb{P}_x$-a.s. to a random element $\widehat{X}_\infty$ with distribution π which does not depend on x and satisfies the SFPE* (3.18).

(b) *For some $\gamma > 1$ and any $x \in \mathbb{X}$,* (3.19) *holds true, that is*

$$\lim_{n\to\infty} \gamma^n d(\widehat{X}_n, \widehat{X}_\infty) = 0 \quad \mathbb{P}_x\text{-a.s.}$$

(c) *For any $x \in \mathbb{X}$, the forward iteration X_n converges to π in distribution under $\mathbb{P}_x$, and π is the unique stationary distribution of $(X_n)_{n\geq 0}$.*

(d) *Under $\mathbb{P}_\pi$, $(X_n)_{n\geq 0}$ forms an ergodic stationary sequence.*

(e) *If, for some $p > 0$, $(X_n)_{n\geq 0}$ is even strongly mean contractive of order p and satisfies the sharpened jump-size condition* (3.20)*, then geometric ergodicity in the sense of* (3.21) *for some $r > 1$ holds true.*

Remark 3.25. In view of Problem 3.14, the assumptions of (e) could be relaxed to

$$\mathbb{E}L_1^p < \infty \quad \text{and} \quad \mathbb{E}d(x_0, \Psi_1(x_0))^p < \infty \tag{3.23}$$

for some $x_0 \in \mathbb{X}$ and $p > 0$.

For the rest of this section, the assumptions of Theorem 3.24 will always be in force without further mention. We embark on the crucial observation that, given a weakly contractive IFS $(X_n)_{n\geq 0}$ of iid Lipschitz maps with Lipschitz constants $L_1, L_2, \ldots$, the SRW

$$S_0 := 0 \quad \text{and} \quad S_n := \sum_{k=1}^{n} \log L_k \quad \text{for } n \geq 1$$

has negative drift. Hence we may fix any $l > 0$ and consider the SRP of a.s. finite level $-l$ ladder epochs, defined by $\sigma_0 := 0$ and

$$\sigma_n := \inf\{k > \sigma_{n-1} : S_k - S_{\sigma_{n-1}} < -l\}$$

for $n \geq 1$. For simplicity, we choose l such that $\mathbb{P}(\sigma_1 = 1) > 0$ and thus $(\sigma_n)_{n\geq 0}$ is 1-arithmetic. The following lemma is basic.

Lemma 3.26. *The embedded sequence $(X_{\sigma_n})_{n\geq 0}$ forms a strongly contractive IFS of iid Lipschitz maps satisfying* (3.17)*, and the same holds true for the sequence $(Y_n)_{n\geq 0}$, defined by $Y_0 := X_0$ and*

$$Y_n := \Psi_{\sigma_{n-1}+1:\sigma_n} \circ \ldots \circ \Psi_{1:\sigma_1}(X_0)$$

for $n \geq 1$.

Proof. Plainly, $X_{\sigma_n} = \Phi_{n:1}(X_0)$ with $\Phi_n := \Psi_{\sigma_n:\sigma_{n-1}+1} \in \mathscr{C}_{Lip}(\mathbb{X})$ for $n \geq 1$. Since $(\sigma_n)_{n\geq 0}$ has iid increments, one can readily check that $\Phi_1, \Phi_2, \ldots$ are iid as well. Moreover, by (3.7),

$$\log L(\Phi_1) \;=\; \log L(\Psi_{\sigma_1} \circ \ldots \circ \Psi_1) \;\leq\; S_{\sigma_1} \;<\; -l$$

which a confirms the strong contraction property. In order to verify the jump-size condition (3.17), i.e.

$$\mathbb{E} \log^+ d(x_0, \Psi_{\sigma_1:1}(x_0)) < \infty$$

for some and thus all $x_0 \in \mathbb{X}$, we will use $\log^+ x \leq \log^* x := \log(1+x)$ and the subadditivity of the latter function. With $S_k^* := \sum_{j=1}^{k} \log^* L_j$ for $k \geq 1$, we infer

$$\begin{aligned}
\log d(x_0, \Psi_{\sigma_1:1}(x_0)) \;&\leq\; \log\left(d(x_0, \Psi_{\sigma_1}(x_0)) + \sum_{n=1}^{\sigma_1 - 1} d(\Psi_{\sigma_1:n+1}(x_0), \Psi_{\sigma_1:n}(x_0)) \right) \\
&\leq\; \log\left(\sum_{n=1}^{\sigma_1} e^{S_{\sigma_1} - S_n} d(x_0, \Psi_n(x_0)) \right) \\
&\leq\; \log\left(e^{S^*_{\sigma_1}} \sum_{n=1}^{\sigma_1} d(x_0, \Psi_n(x_0)) \right)
\end{aligned}$$

$$\leq\ S^*_{\sigma_1} + \sum_{n=1}^{\sigma_1} \log^* d(x_0, \Psi_n(x_0)), \tag{3.24}$$

whence, by using Wald's equation,

$$\mathbb{E} \log^+ d(x_0, \Psi_{\sigma_1:1}(x_0))\ \leq\ \Big(\mathbb{E} \log^* L_1 + \mathbb{E} \log^* d(x_0, \Psi_1(x_0))\Big)\, \mathbb{E}\sigma_1\ <\ \infty$$

as claimed.

Turning to the sequence $(Y_n)_{n\geq 0}$, it suffices to point out that it is obtained from $(X_{\sigma_n})_{n\geq 0}$ by reversing the iteration order within, and only within, the segments determined by the σ_n [☞ Fig. 3.2]. In other words, the Φ_n are substituted for $\Phi_n^{\leftarrow} := \Psi_{\sigma_{n-1}+1:\sigma_n}$, which are still iid and satisfy $L(\Phi_n^{\leftarrow}) \leq -l$ a.s. Again, the bound (3.24) is obtained, when embarking on

$$\log d(x_0, \Psi_{1:\sigma_1}(x_0))\ \leq\ \log \left(d(x_0, \Psi_1(x_0)) + \sum_{n=2}^{\sigma_1} d(\Psi_{1:n-1}(x_0), \Psi_{1:n}(x_0)) \right).$$

The rest is straightforward. □

$$\begin{array}{rl}
X_{\sigma_n}: & \Big|\Psi_{\sigma_n}\ldots\Psi_{\sigma_{n-1}+1}\Big|\ \Big|\Psi_{\sigma_{n-1}}\ldots\Psi_{\sigma_{n-2}+1}\Big|\ \ldots\ \Big|\Psi_{\sigma_1}\ldots\Psi_1\Big| \\
Y_n: & \Big|\Psi_{\sigma_{n-1}+1}\ldots\Psi_{\sigma_n}\Big|\ \Big|\Psi_{\sigma_{n-2}+1}\ldots\Psi_{\sigma_{n-1}}\Big|\ \ldots\ \Big|\Psi_1\ldots\Psi_{\sigma_1}\Big| \\
\widehat{Y}_n = \widehat{X}_{\sigma_n}: & \Big|\Psi_1\ldots\Psi_{\sigma_1}\Big|\ \ldots\ \Big|\Psi_{\sigma_{n-2}+1}\ldots\Psi_{\sigma_{n-1}}\Big|\ \Big|\Psi_{\sigma_{n-1}+1}\ldots\Psi_{\sigma_n}\Big|
\end{array}$$

Fig. 3.2 Schematic illustration of how the blocks of Ψ_k's determined by the σ_n are composed in the definition of X_{σ_n}, Y_n and $\widehat{Y}_n = \widehat{X}_{\sigma_n}$.

The reason for introducing $(Y_n)_{n\geq 0}$ becomes apparent when observing the following twist: the backward iterations of $(X_n)_{n\geq 0}$ at the ladder epochs σ_n, i.e. $(\widehat{X}_{\sigma_n})_{n\geq 0}$, are *not* given by the backward iterations of the IFS $(X_{\sigma_n})_{n\geq 0}$ but rather of $(Y_n)_{n\geq 0}$, thus $(\widehat{Y}_n)_{n\geq 0}$. By the previous lemma in combination with Prop. 3.18, we infer the $\mathbb{P}_x$-a.s. convergence of $\widehat{Y}_n$ to a limit not depending on x, which is clearly the candidate for the a.s. limit of $\widehat{X}_n$ and therefore denoted $\widehat{X}_\infty$.

Let us define $\tau(n) := \inf\{k \geq 0 : \sigma_k \geq n\}$ for $n \geq 0$, and

$$\begin{aligned}
C_n\ :=&\ d(X_0, \Phi_n^{\leftarrow}(X_0)) \vee \max_{\sigma_{n-1}<k<\sigma_n} \big\{ d(\Psi_{\sigma_{n-1}+1:k}(X_0), \Psi_{\sigma_{n-1}+1:\sigma_n}(X_0)) \big\} \\
=&\ \text{distance between } \Phi_n^{\leftarrow}(X_0) = \Psi_{\sigma_{n-1}+1:\sigma_n}(X_0) \text{ and the set} \\
&\ \{X_0, \Psi_{\sigma_{n-1}+1}(X_0), \Psi_{\sigma_{n-1}+1:\sigma_{n-1}+2}(X_0), \ldots, \Psi_{\sigma_{n-1}+1:\sigma_n-1}(X_0)\}
\end{aligned} \tag{3.25}$$

for $n \geq 1$. By the elementary renewal theorem [☞ Lemma 2.1(e) and (f)],

$$n^{-1}\tau(n) \to \mu^{-1} \quad \text{as well as} \quad n^{-1}\mathbb{E}\tau(n) \to \mu^{-1},$$

where $\mu = \mu(l) := \mathbb{E}\sigma_1$. Under each $\mathbb{P}_x$, the C_n are clearly iid, and a standard renewal argument shows that $C_{\tau(n)}$ converges in distribution to a random variable C_∞ [☞ Problem 3.30]. However, the really needed piece of information about $C_{\tau(n)}$ will be a consequence of the following lemma.

Lemma 3.27. *For any $x \in \mathbb{X}$, $\mathbb{E}_x \log^+ C_1 < \infty$ and hence $n^{-1}\log^+ C_n \to 0$ as well as $e^{-\varepsilon n} C_n \to 0$ $\mathbb{P}_x$-a.s. for each $\varepsilon > 0$.*

Indeed, as $\tau(n) \to \infty$, the last assertion particularly implies

$$\lim_{n\to\infty} e^{-\varepsilon\tau(n)} C_{\tau(n)} = 0 \quad \mathbb{P}_x\text{-a.s.} \tag{3.26}$$

for all $x \in \mathbb{X}$ and $\varepsilon > 0$.

Proof. Fix any $x \in \mathbb{X}$. By proceeding as for (3.24), we find

$$\begin{aligned}\log C_1 \ &\le\ \log\left(d(x_0,\Psi_1(x_0)) + \sum_{n=2}^{\sigma_1} d(\Psi_{1:n-1}(X_0),\Psi_{1:n}(X_0))\right)\\ &\le\ \log\left(\sum_{n=1}^{\sigma_1} e^{S_{n-1}} d(X_0,\Psi_n(X_0))\right)\\ &\le\ S_{\sigma_1}^* + \sum_{n=1}^{\sigma_1} \log^* d(X_0,\Psi_n(X_0)) \quad \mathbb{P}_x\text{-a.s.}\end{aligned}$$

and the last expression has finite expectation under $\mathbb{P}_x$ by an appeal to Wald's equation. As a consequence, $n^{-1}\log^+ C_n \to 0$ $\mathbb{P}_x$-a.s., and this is readily seen to also give $e^{-\varepsilon n} C_n \to 0$ $\mathbb{P}_x$-a.s. for all $\varepsilon > 0$. □

The crucial estimate of $d(\widehat{X}_n,\widehat{X}_\infty)$ in terms of the previously introduced variables is provided in a further lemma.

Lemma 3.28. *For all $n \ge 0$ and $x \in \mathbb{X}$, the inequality*

$$d(\widehat{X}_n,\widehat{X}_\infty)\ \le\ e^{-(\tau(n)-1)l} C_{\tau(n)} + d(\widehat{Y}_{\tau(n)},\widehat{X}_\infty) \quad \mathbb{P}_x\text{-a.s.}$$

holds, where $C_0 := 0$.

Proof. Using the strong contraction property of $(\widehat{Y}_n)_{n\ge 0}$ and $\widehat{Y}_n \to \widehat{X}_\infty$ $\mathbb{P}_x$-a.s., we obtain

$$\begin{aligned}d(\widehat{X}_n,\widehat{X}_\infty)\ &\le\ d(\widehat{X}_n,\widehat{Y}_{\tau(n)}) + d(\widehat{Y}_{\tau(n)},\widehat{X}_\infty)\\ &\le\ e^{-(\tau(n)-1)l} C_{\tau(n)} + d(\widehat{Y}_{\tau(n)},\widehat{X}_\infty) \quad \mathbb{P}_x\text{-a.s.}\end{aligned}$$

and this is the asserted inequality. □

Proof (of Theorem 3.24). (a) Recall that $\widehat{X}_\infty$ equals the $\mathbb{P}_x$-a.s. limit of $\widehat{Y}_n = \widehat{X}_{\sigma_n}$ for each $x \in \mathbb{X}$. By combining this with (3.26) and the previous lemma, the a.s. convergence $\widehat{X}_n \to \widehat{X}_\infty$ under each $\mathbb{P}_x$ follows. The SFPE for $\widehat{X}_\infty$ is obtained in the same manner as in the proof of Prop. 3.18.

(b) First note that, by Prop. 3.18(b),

$$\lim_{n\to\infty} \beta^{\tau(n)} d(\widehat{Y}_{\tau(n)}, \widehat{X}_\infty) = 0 \quad \mathbb{P}_x\text{-a.s.}$$

for some $\beta > 1$. Since $n^{-1}\tau(n) \to \mu^{-1}$ $\mathbb{P}$-a.s., we can pick $\varepsilon > 0$ and $\gamma > 1$ such that $\gamma^{n/\tau(n)} < \beta$ and $\gamma^n e^{-l(\tau(n)-1)} < e^{-\varepsilon\tau(n)}$ $\mathbb{P}$-a.s. for all sufficiently large n (depending on the realization of the $\tau(n)$). By another use of Lemma 3.28, we then infer

$$\begin{aligned}\gamma^n d(\widehat{X}_n, \widehat{X}_\infty) \ &\le\ \gamma^n e^{-l(\tau(n)-1)} C_{\tau(n)} + \left(\gamma^{n/\tau(n)}\right)^{\tau(n)} d(\widehat{Y}_{\tau(n)}, \widehat{X}_\infty)\\ &\le\ e^{-\varepsilon\tau(n)} C_{\tau(n)} + \beta^{\tau(n)} d(\widehat{Y}_{\tau(n)}, \widehat{X}_\infty) \quad \mathbb{P}_x\text{-a.s.}\end{aligned}$$

for all sufficiently large n and any $x \in \mathbb{X}$, and this yields (3.19) upon letting n tend to ∞.

(c) and (d) follow again in the same way as in the proof of Prop. 3.18.

(e) Now assume strong mean contraction of order p, w.l.o.g. $0 < p \le 1$, and (3.20). Put $\rho := (\mathbb{E}L_1^p)^{1/p} = (\mathbb{E}e^{pS_1})^{1/p}$. Using the independence of S_n and $\Psi_{n+1}(x)$ for all $n \in \mathbb{N}_0$ and $x \in \mathbb{X}$, a similar estimation as in the proof of part (e) of Prop. 3.18 leads to

$$\begin{aligned}\mathbb{P}_x(d(\widehat{X}_n, \widehat{X}_\infty) \ge s^{-n}) \ &\le\ s^{np}\, \mathbb{E}_x d(\widehat{X}_n, \widehat{X}_\infty)^p\\ &\le\ s^{np} \sum_{k\ge n} \mathbb{E}_x d(\widehat{X}_k, \widehat{X}_{k+1})^p\\ &=\ s^{np} \sum_{k\ge n} \mathbb{E}\, d(\Psi_{1:k}(x), \Psi_{1:k+1}(x))^p\\ &\le\ s^{np} \sum_{k\ge n} \mathbb{E}\, e^{pS_k} d(\Psi_{1:k}(x), \Psi_{1:k+1}(x))^p\\ &\le\ (s\rho)^{np}\, \mathbb{E}\, d(x, \Psi_1(x))^p \sum_{k\ge 0} \rho^{kp}\\ &=\ (s\rho)^{np} \frac{\mathbb{E}\, d(x, \Psi_1(x))^p}{1-\rho^p}.\end{aligned}$$

for $s > 0$. It follows that

$$\lim_{n\to\infty} s^n\, \mathbb{P}_x(d(\widehat{X}_n, \widehat{X}_\infty) \ge s^{-n}) = 0$$

for $1 < s < \rho^{-p/(p+1)}$, and once again this entails (3.21) for $r \in (1, s)$ by an appeal to Lemma 3.17. ◻

We close this section with a result on the existence of moments of π, more precisely of

$$\int_{\mathbb{X}} d(x_0,x)^p\ \pi(dx)\ =\ \mathbb{E}d(x_0,\widehat{X}_\infty)^p$$

for any fixed $x_0\in\mathbb{X}$ and $p>0$. In the case of Euclidean space $\mathbb{X}=\mathbb{R}^m$ for some $m\ge 1$ with the usual norm $d(x,y)=|x-y|$ and $x_0=0$, this means to consider

$$\int_{\mathbb{R}^m}|x|^p\ \pi(dx)\ =\ \mathbb{E}|\widehat{X}_\infty|^p.$$

The following theorem is due to BENDA [13, Prop. 2.2] for $p\ge 1$; for a weaker version see [3, Theorem 2.3(d)]. It does not only complement the main result of this section, but will also be useful in connection with the implicit renewal theory developed in Chapter 4.

Theorem 3.29. *If $(X_n)_{n\ge 0}$ is strongly mean contractive of order $p>0$ and satisfies the corresponding sharpened jump-size condition* (3.20)*, i.e. $\mathbb{E}L_1^p<1$ and $\mathbb{E}d(x_0,\Psi_1(x_0))^p<\infty$ for some/all $x_0\in\mathbb{X}$, then*

$$\mathbb{E}d(x_0,\widehat{X}_\infty)^p\ =\ \int_{\mathbb{X}} d(x_0,x)^p\ \pi(dx)\ <\ \infty$$

for some and then all $x_0\in\mathbb{X}$.

Proof. Put $\beta:=\mathbb{E}L_1^p=\mathbb{E}e^{pS_1}$. If $p\le 1$, we infer by using the subadditivity of $x\mapsto x^p$ and the model assumptions that, for any $x_0\in\mathbb{X}$,

$$\begin{aligned}\mathbb{E}d(x_0,\widehat{X}_\infty)^p\ &\le\ \mathbb{E}\left(d(x_0,\Psi_1(x_0))+\sum_{n\ge 1}d(\Psi_{n:1}(x_0),\Psi_{n+1:1}(x_0))\right)^p\\ &\le\ \mathbb{E}d(x_0,\Psi_1(x_0))^p+\sum_{n\ge 1}\mathbb{E}e^{pS_n}d(x_0,\Psi_{n+1}(x_0))^p\\ &\le\ \mathbb{E}d(x_0,\Psi_1(x_0))^p\sum_{n\ge 0}\mathbb{E}e^{pS_n}\\ &=\ \frac{\mathbb{E}d(x_0,\Psi_1(x_0))^p}{1-\beta}\ <\ \infty.\end{aligned}$$

In the case $p>1$ we can use Minkowski's inequality, which also holds for infinitely many summands, to obtain in a similar manner

$$\begin{aligned}\|d(x_0,\widehat{X}_\infty)\|_p\ &\le\ \left\|d(x_0,\Psi_1(x_0))+\sum_{n\ge 1}d(\Psi_{n:1}(x_0),\Psi_{n+1:1}(x_0))\right\|_p\\ &\le\ \|d(x_0,\Psi_1(x_0))\|_p+\sum_{n\ge 1}\|e^{S_n}d(x_0,\Psi_{n+1}(x_0))\|_p\end{aligned}$$

$$\le \|d(x_0,\Psi_1(x_0))\|_p \sum_{n\ge 0} \|e^{S_n}\|_p$$
$$= \frac{\mathbb{E}d(x_0,\Psi_1(x_0))^p}{1-\beta^{1/p}} < \infty.$$

This completes the proof. □

Problems

Problem 3.30. Use a renewal argument to prove that C_n defined in (3.25) converges in distribution to a random variable C_∞ with cdf

$$\mathbb{P}_x(C_\infty \le t) = \frac{1}{\mu}\sum_{n\ge 0}\mathbb{P}_x(\sigma_1 > n, C_1 \le t), \quad t\in\mathbb{R}_\ge.$$

[Recall that l in the definition of σ_1 was chosen such that σ_1 is 1-arithmetic.]

Problem 3.31. Following WU & SHAO [114], say that an IFS $(X_n)_{n\ge 0}$ of iid Lipschitz maps $\Psi_1,\Psi_2,\ldots$ is *geometrically moment contracting of order* $p>0$ if there exist $x_0\in\mathbb{X}$, $\kappa_p\in\mathbb{R}_>$ and $\rho_p\in(0,1)$ such that

$$\mathbb{E}d(\Psi_{n:1}(x),\Psi_{n,1}(x_0))^p \le \kappa_p\, d(x_0,x)^p\,\rho_p^n \tag{3.27}$$

for all $x\in\mathbb{X}$ and $n\in\mathbb{N}$. Prove the following assertions:

(a) If (3.27) is valid for some $x_0\in\mathbb{X}$, then it actually holds for any $x_0\in\mathbb{X}$ in the slightly weaker form

$$\mathbb{E}d(\Psi_{n:1}(x),\Psi_{n,1}(x_0))^p \le \kappa_p(x_0)\left(d(x_0,x)^p+1\right)\rho_p^n$$

for all $x\in\mathbb{X}$, $n\in\mathbb{N}$ and some $\kappa_p(x_0)\in\mathbb{R}_>$.

(b) (3.27) implies the very same condition for any $q\in(0,p)$.

(c) If $(X_n)_{n\ge 0}$ is strongly mean contractive of order p, then it is also geometrically moment contracting of order p.

(d) [114, Theorem 2] If $(X_n)_{n\ge 0}$ is geometrically moment contracting of some order p and satisfies the sharpened jump-size condition (3.20), then all assertions of Theorem 3.24 remain valid.

3.4 A central limit theorem for strongly mean contractive IFS

Let $\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})$ be the vector space of all real-valued Lipschitz functions g on $\mathbb{X}$ and

$$L(g) := \sup_{x,y\in\mathbb{X},x\ne y}\frac{|g(x)-g(y)|}{d(x,y)}$$

the Lipschitz constant of g [☞ (3.3)]. For $\pi \in \mathscr{P}(\mathbb{X})$ and $g \in L^1(\pi)$, we put

$$\pi(g) \ := \ \int_{\mathbb{X}} g(x)\ \pi(dx).$$

Notice that, if $\mu_p(x_0) := \int d(x_0,x)^p\, \pi(dx) < \infty$ for some $p \geq 1$ and $x_0 \in \mathbb{X}$, then

$$\|g\|_p \ \leq \ |g(x_0)| + \|g - g(x_0)\|_p \ \leq \ |g(x_0)| + L(g)\, \mu_p(x_0) \ < \ \infty. \tag{3.28}$$

and therefore

$$\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R}) \subset L^p(\pi).$$

The purpose of this section is to study the asymptotic behavior of *additive functionals* of an IFS $(X_n)_{n\geq 0}$, viz.

$$S_n(g) : = \ \sum_{k=1}^{n} g(X_k) \ = \ \sum_{k=1}^{n} g(\Psi_{k:1}(X_0)), \quad n \geq 1,$$

for suitable functions g. The following CLT was obtained by BENDA [13] and requires g to be Lipschitz continuous. For a corresponding SLLN under slightly weaker assumptions ☞ Problem 3.33.

Theorem 3.32. *Let $(X_n)_{n\geq 0}$ be an IFS of iid Lipschitz maps in a standard model which is strongly mean contractive of order* 2 *and satisfies the sharpened jump-size condition* (3.20) *with $p = 2$ for some and thus all $x_0 \in \mathbb{X}$. Let π be its stationary distribution and further $g \in \mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})$. Then the following assertions hold true:*

(a) *There exists $h \in \mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})$ such that $g = h - Ph$, where*

$$Ph(x) \ := \ \mathbb{E} h \circ \Psi_1(x) \ = \ \mathbb{E}_x\, h(X_1)$$

is the transition operator of $(X_n)_{n\geq 0}$.

(b) *If $\sigma^2 := \pi(h^2) - \pi((Ph)^2) > 0$, then*

$$\frac{S_n(g) - n\pi(g)}{\sigma n^{1/2}} \ \xrightarrow{d} \ Normal(0,1) \tag{3.29}$$

under each $\mathbb{P}_x$, $x \in \mathbb{X}$.

After inspection of the proof the reader can easily verify and is asked to do so in Problem 3.35 that the theorem remains valid for $g \in \mathscr{C}_{Lip}(\mathbb{X},\mathbb{R}) \cap L^2(\pi)$ if $(X_n)_{n\geq 0}$ is only strongly mean contractive of order $p > 0$ and satisfying the corresponding sharpened jump-size condition (3.20).

Proof. First of all, we point out that it suffices to prove (3.29) under $\mathbb{P}_\pi$. Namely, $\vartheta := \mathbb{E}L_1 \leq (\mathbb{E}L_1^2)^{1/2} < 1$ implies

$$r^{-n}L(\Psi_{n:1})\ \le\ r^{-n}\prod_{k=1}^{n}L_k\ \to\ 0\quad\text{a.s.}$$

for any $r\in(\vartheta,1)$ and then

$$\frac{1}{n^{1/2}}\left|\sum_{k=1}^{n}g(\Psi_{k:1}(x))-\sum_{k=1}^{n}g(\Psi_{k:1}(Z))\right|\ \le\ \frac{L(g)\,d(x,Z)}{n^{1/2}}\sum_{k=1}^{n}L(\Psi_{k:1})\ \to\ 0\quad\text{a.s.}$$

for any $x\in\mathbb{X}$ and any random variable Z with law π and independent of $\Psi_1,\Psi_2,\ldots$ Therefore, we will study $S_n(g)$ only under $\mathbb{P}=\mathbb{P}_\pi$ hereafter and thus assume that X_0 has distribution π. Next observe that, by stationarity, $P^ng=\mathbb{E}g(\Psi_{n:1}(X_0))=\mathbb{E}g(X_0)=\pi(g)$ for each $n\ge 0$, implying

$$S_n(g)-n\pi(g)\ =\ S_n(g-\pi(g))\quad\text{a.s.}$$

for each $n\ge 1$. It is therefore no loss of generality to further assume $\pi(g)=0$.

Define

$$\mathscr{L}\ :=\ \left\{f\in\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R}):\pi(f)=0\right\}$$

and, for any fixed $x_0\in\mathbb{X}$, the norm

$$\|f\|\ :=\ \sup_{x\in\mathbb{X}}\frac{|f(x)|}{1+d(x_0,x)}+L(f).$$

Then $(\mathscr{L},\|\cdot\|)$ is a Banach space and $P:\mathscr{L}\to\mathscr{L}$ a bounded linear operator with

$$\|P\|\ :=\ \sup_{\|f\|=1}\|Pf\|\ \le\ 1$$

[☞ Problem 3.34(a)]. Since, by (3.20) (with $p=2$) and Theorem 3.29,

$$\mu_2(x_0)\ =\ \int d(x_0,x)^2\ \pi(dx)\ <\ \infty,$$

we infer $\mathscr{L}\subset\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})\subset L^2(\pi)$ from (3.28).

In order to get an estimate for $\|P^nf\|$ for any $n\ge 1$, note that

$$\begin{aligned}|P^nf(x)|\ &=\ |\mathbb{E}f\circ\Psi_{n:1}(x)|\ =\ |\mathbb{E}f\circ\Psi_{1:n}(x)|\\ &=\ \left|\mathbb{E}f\circ\Psi_{1:n}(x)-\lim_{m\to\infty}\mathbb{E}f\circ\Psi_{1:m}(x)\right|\\ &\le\ \lim_{m\to\infty}\mathbb{E}\,|f\circ\Psi_{1:n}(x)-f\circ\Psi_{1:m}(x)|\\ &\le\ L(f)\,\mathbb{E}\,L(\Psi_{1:n})\lim_{m\to\infty}\mathbb{E}\,d(x,\Psi_{n+1:m}(x))\\ &=\ L(f)\,\mu(x)\,\vartheta^n,\end{aligned}\tag{3.30}$$

where $\mu(x):=\int d(x,y)\,\pi(dy)$ is finite for any x by another appeal to Theorem 3.29 and

$$\lim_{m\to\infty} \mathbb{E} f\circ\Psi_{1:m}(x) = 0 \quad \text{and} \quad \lim_{m\to\infty} \mathbb{E}\, d(x,\Psi_{n+1:m}(x) = \mu(x)$$

has been used which the reader is asked to verify in Problem 3.34(b). Furthermore,

$$\begin{aligned} L(P^n f) &= \sup_{x,y\in\mathbb{X},x\neq y} \frac{|\mathbb{E} f\circ\Psi_{n:1}(x) - \mathbb{E} f\circ\Psi_{n:1}(y)|}{d(x,y)} \\ &\le \sup_{x,y\in\mathbb{X},x\neq y} \frac{\mathbb{E}\,|f\circ\Psi_{n:1}(x) - f\circ\Psi_{n:1}(y)|}{d(x,y)} \\ &\le L(f)\,\vartheta^n \end{aligned} \tag{3.31}$$

for any $n \ge 1$. A combination of (3.30) and (3.31) then yields

$$\|P^n f\| \le \|f\|\, K\,\vartheta^n \quad \text{with} \quad K := \sup_{x\in\mathbb{X}} \frac{\mu(x)}{1+d(x_0,x)} + 1 < \infty$$

(observe that $\mu(x) \le d(x_0,x) + \mu(x_0)$ for the finiteness of K) and thereupon

$$\limsup_{n\to\infty} \|P^n\|^{1/n} < 1.$$

Consequently, $I - P : \mathscr{L} \to \mathscr{L}$ with I denoting the identity operator is invertible. This completes the proof of (a) when choosing $h := (I-P)^{-1} g = \sum_{n\ge 0} P^n g$.

Since $Ph(x) = \mathbb{E}_x h(X_1) = \mathbb{E}(h(X_k)|X_{k-1} = x)$ for any $x \in \mathbb{X}$ and $k \ge 1$, we see that, under $\mathbb{P} = \mathbb{P}_\pi$, the sequence

$$M_n := \sum_{k=1}^n \Big(h(X_k) - \mathbb{E}(h(X_k)|X_{k-1})\Big), \quad n \ge 0$$

forms a zero-mean L^2-martingale with stationary, ergodic increments having variance σ^2. By invoking an old result by BILLINGSLEY [15], we then infer that

$$n^{-1/2} M_n \xrightarrow{d} Normal(0,\sigma^2).$$

The proof of (b) is finally completed by the observation that

$$\begin{aligned} S_n(g) &= S_n(h) - S_n(Ph) \\ &= Ph(X_0) - Ph(X_n) + \sum_{k=1}^n \Big(h(X_k) - Ph(X_{k-1})\Big) \\ &= Ph(X_0) - Ph(X_n) + \sum_{k=1}^n \Big(h(X_k) - \mathbb{E}(h(X_k)|X_{k-1})\Big) \\ &= Ph(X_0) - Ph(X_n) + M_n \end{aligned}$$

for all $n \ge 1$ in combination with $n^{-1/2}(Ph(X_0) - Ph(X_n)) \xrightarrow{\mathbb{P}} 0$. □

Despite its elegant proof, the previous theorem has the obvious disadvantage that it requires Lipschitz continuity of g which excludes even indicators of simple subsets of $\mathbb{X}$ like ε-balls and does therefore not provide information on the asymptotic behavior of *relative frequencies*

$$\frac{S_n(\mathbf{1}_B)}{n} \;=\; \frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_B(S_n)$$

which are of particular interest for statistical purposes.

WU & WOODROOFE [115] have provided an alternative result which shows that, under a slightly stronger integrability condition on g, one can dispense with the continuity of g and also relax the conditions on the IFS $(X_n)_{n\ge 0}$.

A function $g:\mathbb{X}\to\mathbb{R}$ is called *Dini continuous* if

$$\int_0^1 \frac{\omega_g(t)}{t}\,dt \;<\; \infty,$$

where $\omega_g(t) := \sup\{|g(x)-g(y)| : d(x,y)\le t\}$ denotes the *modulus of continuity* of g.

Problems

Problem 3.33. Consider the situation of Theorem 3.32, but with $(X_n)_{n\ge 0}$ being only strongly mean contractive of order one and satisfying the sharpened jump-size condition (3.20) with $p=1$. Prove that

$$\frac{S_n(g)}{n} \to \pi(g) \quad \mathbb{P}_x\text{-a.s.} \tag{3.32}$$

for all $x\in\mathbb{X}$ and $g\in\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})\subset L^1(\pi)$ [☞ (3.28)].

Problem 3.34. Given the assumptions of Theorem 3.32 and the notation of its proof, prove the following assertions from there:

(a) $(\mathscr{L},\|\cdot\|)$ is a Banach space and $P:\mathscr{L}\to\mathscr{L}$ a bounded linear operator with norm $\|P\|\le 1$.

(b) Recalling that $\mu(x)=\int d(x,y)\,\pi(dy)$ and $f\in\mathscr{L}$,

$$\lim_{m\to\infty}\mathbb{E}f\circ\Psi_{1:m}(x)=0 \quad\text{and}\quad \lim_{m\to\infty}\mathbb{E}\,d(x,\Psi_{n+1:m}(x))=\mu(x).$$

Problem 3.35. Show that Theorem 3.32 remains valid for $g\in\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})\cap L^2(\pi)$ if $(X_n)_{n\ge 0}$ is strongly mean contractive of order $p>0$ and satisfying the corresponding sharpened jump-size condition (3.20). Furthermore, (3.32) then holds true for any $g\in\mathscr{C}_{Lip}(\mathbb{X},\mathbb{R})\cap L^1(\pi)$.

Problem 3.36.

Chapter 4
Power law behavior of stochastic fixed points and implicit renewal theory

The previous chapter has shown that any mean contractive IFS $(X_n)_{n\ge0}$ of iid Lipschitz maps $\Psi_1,\Psi_2,\ldots$ converges in distribution to a unique stationary limit π which is characterized as the unique solution to the SFPE

$$X \stackrel{d}{=} \Psi(X) \tag{4.1}$$

where Ψ denotes a generic copy of the Ψ_n independent of X. In the following, we will deal with the problem of gaining information about the tail behavior of $\pi = \mathbb{P}(X\in\cdot)$, more precisely, the behavior of

$$\mathbb{P}(X>t) \quad \text{and/or} \quad \mathbb{P}(X<-t) \quad \text{as } t\to\infty.$$

If they are asymptotically equal to a nonzero constant times a power $|t|^{\vartheta}$ for some $\vartheta>0$, we say that X (or π) exhibits a *power law behavior*.[1] For the case when $\Psi(x)$, for x large, is approximately Mx for a random variable M, GOLDIE [58] developed a method he called *implicit renewal theory*, which allows to establish power law behavior of X under appropriate moment conditions on M. The present chapter is devoted to a presentation of his main results.

4.1 Goldie's implicit renewal theorem

Given an SFPE of type (4.1) with $\Psi(x)\approx Mx$ for a random variable M if x is large, Goldie's basic idea to study the asymptotics of $\mathbb{P}(X>t)$ and $\mathbb{P}(X<-t)$ as $t\to\infty$ is to consider the differences

$$\mathbb{P}(X>t)-\mathbb{P}(MX>t) \quad \text{and} \quad \mathbb{P}(X<-t)-\mathbb{P}(MX<-t)$$

[1] In some papers like [33] it is alternatively said that X has *algebraic tails*.

as t tends to ∞. Additionally assuming that M is nonnegative, the renewal-theoretic character of this approach becomes immediately apparent after a logarithmic transform. Put $Y := \log X^+$, $\xi := \log M$ and $G(t) := \mathbb{P}(Y > t)$ for $t \in \mathbb{R}$. Since M and X are independent, we infer for $\Delta(t) := \mathbb{P}(X > e^t) - \mathbb{P}(MX > e^t)$ that

$$\Delta(t) \ = \ G(t) - \int G(t-x)\, \mathbb{P}(\xi \in dx), \quad t \in \mathbb{R}, \tag{4.2}$$

or, equivalently, the renewal equation $G = \Delta + G * Q$ with $Q := \mathbb{P}(\xi \in \cdot)$ holds. However, unlike the usual situation [☞ Section 2.7], the function Δ is also unknown here and indeed an integral involving G. That renewal-theoretic arguments still work to draw conclusions about G is the key feature of the approach and the following result in particular. It will be made more precise in the next section.

Theorem 4.1. [Implicit renewal theorem] *Let M, X be independent random variables such that, for some $\vartheta > 0$,*

(IRT-1) $\mathbb{E}|M|^{\vartheta} = 1$.
(IRT-2) $\mathbb{E}|M|^{\vartheta} \log^+ |M| < \infty$.
(IRT-3) *The conditional law $\mathbb{P}(\log|M| \in \cdot | M \neq 0)$ of $\log|M|$ given $M \neq 0$ is nonarithmetic, in particular, $\mathbb{P}(|M| = 1) < 1$.*

Then $-\infty \le \mathbb{E}\log|M| < 0$, $0 < \mu_\vartheta := \mathbb{E}|M|^{\vartheta} \log|M| < \infty$, and the following assertions hold true:

(a) Suppose M is a.s. nonnegative. If

$$\int_0^\infty \big|\mathbb{P}(X > t) - \mathbb{P}(MX > t)\big|\, t^{\vartheta-1}\, dt \ < \ \infty \tag{4.3}$$

or, respectively,

$$\int_0^\infty \big|\mathbb{P}(X < -t) - \mathbb{P}(MX < -t)\big|\, t^{\vartheta-1}\, dt \ < \ \infty, \tag{4.4}$$

then

$$\lim_{t\to\infty} t^{\vartheta}\, \mathbb{P}(X > t) \ = \ C_+, \tag{4.5}$$

respectively

$$\lim_{t\to\infty} t^{\vartheta}\, \mathbb{P}(X < -t) \ = \ C_-, \tag{4.6}$$

where C_+ and C_- are given by the equations

$$C_+ \ := \ \frac{1}{\mu_\vartheta} \int_0^\infty \big(\mathbb{P}(X > t) - \mathbb{P}(MX > t)\big)\, t^{\vartheta-1}\, dt, \tag{4.7}$$

$$C_- \ := \ \frac{1}{\mu_\vartheta} \int_0^\infty \big(\mathbb{P}(X < -t) - \mathbb{P}(MX < -t)\big)\, t^{\vartheta-1}\, dt. \tag{4.8}$$

(b) *If* $\mathbb{P}(M<0)>0$ *and* (4.3), (4.4) *are both satisfied, then* (4.5) *and* (4.6) *hold with* $C_+=C_-=C/2$, *where*

$$C \;:=\; \frac{1}{\mu_\vartheta}\int_0^\infty \big(\mathbb{P}(|X|>t)-\mathbb{P}(|MX|>t)\big)\,t^{\vartheta-1}\,dt. \tag{4.9}$$

The proof of this result naturally requires some work which will be carried out in the Section 4.3. Let us rather point here as in [58] that the theorem has real content only if $\mathbb{E}|X|^\vartheta=\infty$, because otherwise, by the independence of M and X and condition (IRT-1),

$$C \;=\; \frac{1}{\vartheta\mu_\vartheta}\Big(\mathbb{E}|X|^\vartheta-\mathbb{E}|MX|^\vartheta\Big) \;=\; \frac{1}{\vartheta\mu_\vartheta}\,\mathbb{E}|X|^\vartheta\Big(1-\mathbb{E}|M|^\vartheta\Big) \;=\; 0$$

in which case (4.5) and (4.6) take the form

$$\lim_{t\to\infty} t^\vartheta\,\mathbb{P}(|X|>t) \;=\; 0,$$

Naturally, this follows also directly from $t^\vartheta\,\mathbb{P}(|X|>t)\le\mathbb{E}\mathbf{1}_{\{|X|>t\}}|X|^\vartheta\to 0$. We thus see that the "right" choice of M and ϑ is crucial.

The next corollary specializes to the situation where X additionally satisfies the SFPE (4.1) for a Borel-measurable random function Ψ.

Corollary 4.2. *Let* $(\Omega,\mathfrak{A},\mathbb{P})$ *be any probability space,* $\Psi:\Omega\times\mathbb{R}\to\mathbb{R}$ *a* $\mathfrak{A}\otimes\mathscr{B}(\mathbb{R})$*-measurable function and* X,M *further random variables on* Ω *such that* X *solves* (4.1) *and is independent of* (Ψ,M)*. Suppose also that* M *satisfies (IRT-1)-(IRT-3). Then, in Theorem 4.1, conditions* (4.3) *and* (4.4) *may be replaced by (the generally stronger)*

$$\mathbb{E}\Big|(\Psi(X)^+)^\vartheta-((MX)^+)^\vartheta\Big|<\infty \tag{4.10}$$

and

$$\mathbb{E}\Big|(\Psi(X)^-)^\vartheta-((MX)^-)^\vartheta\Big|<\infty \tag{4.11}$$

respectively, and the formulae in (4.7), (4.8) *and* (4.9) *by*

$$C_+ \;=\; \frac{1}{\vartheta\mu_\vartheta}\,\mathbb{E}\Big((\Psi(X)^+)^\vartheta-((MX)^+)^\vartheta\Big), \tag{4.12}$$

$$C_- \;=\; \frac{1}{\vartheta\mu_\vartheta}\,\mathbb{E}\Big((\Psi(X)^-)^\vartheta-((MX)^-)^\vartheta\Big), \tag{4.13}$$

and

$$C = \frac{1}{\vartheta\mu_\vartheta}\,\mathbb{E}\Big(|\Psi(X)|^\vartheta - |MX|^\vartheta\Big), \tag{4.14}$$

respectively.

The proof with the help of Theorem 4.1 is quite simple and provided after the following lemma.

Lemma 4.3. *Let X,Y be two real-valued random variables and $\vartheta > 0$. Then*

$$\int_0^\infty \big|\mathbb{P}(X>t) - \mathbb{P}(Y>t)\big|\,t^{\vartheta-1}\,dt \;\le\; \frac{1}{\vartheta}\,\mathbb{E}\Big|(X^+)^\vartheta - (Y^+)^\vartheta\Big|, \tag{4.15}$$

finite or infinite. If finite, absolute value signs may be removed to give

$$\int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(Y>t)\big)\,t^{\vartheta-1}\,dt \;=\; \frac{1}{\vartheta}\,\mathbb{E}\Big((X^+)^\vartheta - (Y^+)^\vartheta\Big). \tag{4.16}$$

Proof. Problem 4.6. □

Remark 4.4. The previous lemma bears a subtlety that is easily overlooked at first reading (and has actually been done so also in [58]). If F,G denote the df's of X,Y, then (4.15) may be restated as

$$\int_0^\infty \big|F(t) - G(t)\big|\,t^{\vartheta-1}\,dt \;\le\; \frac{1}{\vartheta}\,\mathbb{E}\Big|(X^+)^\vartheta - (Y^+)^\vartheta\Big| \tag{4.17}$$

and holds true for *every* (F,G)*-coupling* (X,Y) on some probability space $(\Omega,\mathfrak{A},\mathbb{P})$, which means that $\mathscr{L}(X) = F$ and $\mathscr{L}(Y) = G$. Moreover, any such coupling having $\mathbb{E}\big|(X^+)^\vartheta - (Y^+)^\vartheta\big| < \infty$ leads to the same value when removing the absolute value signs, namely the left-hand integral in (4.16), i.e.

$$\int_0^\infty \big(F(t) - G(t)\big)\,t^{\vartheta-1}\,dt.$$

This is trivial if $\mathbb{E}(X^+)^\vartheta$ and $\mathbb{E}(Y^+)^\vartheta$ are finite individually, but requires a proof otherwise. Finally, it should be noted that the above integral (with or without absolute value signs) actually depends only on the restrictions F_+,G_+ of F,G to $\mathbb{R}_\ge$, i.e., the cdf of X^+,Y^+, respectively.

Concerning inequality (4.17), it is natural to ask whether equality can be achieved by choosing a special (F,G)-coupling of (X,Y) or actually, by the previous remark, a special (F_+,G_+)-coupling (X^+,Y^+). This is indeed the case when $X^+ = F_+^{-1}(U)$ and $Y^+ = G_+^{-1}(U)$, where U is a *Unif*$(0,1)$ random variable and H^{-1} denotes the *pseudo-inverse* of a cdf H, defined by $H^{-1}(u) := \inf\{x \in \mathbb{R} : H(x) \ge u\}$ for $u \in (0,1)$. The reader is asked for a proof in Problem 4.6.

Proof (of Corollary 4.2). By combining the SFPE (4.1) with the previous lemma, we see that (4.3) turns into

$$\begin{aligned}\infty &> \int_0^\infty \big|\mathbb{P}(\Psi(X)>t)-\mathbb{P}(MX>t)\big|\, t^{\vartheta-1}\, dt \\ &= \mathbb{E}\big|(\Psi(X)^+)^\vartheta - ((MX)^+)^\vartheta\big|\end{aligned}$$

which is condition (4.10). Since all other asserted replacements follow in the same manner, the result is proved. □

Problems

Problem 4.5. Prove that, if at least one of (4.3) and (4.4) is valid, then these conditions hold together iff

$$\int_0^\infty \big|\mathbb{P}(|X|>t)-\mathbb{P}(|MX|>t)\big|\, t^{\vartheta-1}\, dt \;<\; \infty. \tag{4.18}$$

Problem 4.6. Prove Lemma 4.3 and, furthermore, that

$$\int_0^\infty \big|F(t)-G(t)\big|\, t^{\vartheta-1}\, dt \;=\; \frac{1}{\vartheta}\,\mathbb{E}\Big|F_+^{-1}(U)^\vartheta - G_+^{-1}(U)^\vartheta\Big| \tag{4.19}$$

where F_+^{-1}, G_+^{-1} and U are as stated in Remark 4.4.

4.2 Making explicit the implicit

Let us take as a starting point a *two-sided renewal equation* of the form

$$G(t) \;=\; \Delta(t) \;+\; \int G(t-x)\, Q(dx), \quad t\in\mathbb{R},$$

as in (4.2), where $G,\Delta:\mathbb{R}\to\mathbb{R}$ are unknown bounded functions vanishing at ∞, i.e.

$$\lim_{t\to\infty} G(t)=0 \quad\text{and}\quad \lim_{t\to\infty}\Delta(t)=0,$$

and Q is a given probability measure with mean $\mu<0$. Let $\mathbb{U}:=\sum_{n\ge 0} Q^{*n}$ denote its renewal measure. The goal is to determine the precise asymptotic behavior of $G(t)$ as $t\to\infty$. Although we have studied only standard renewal equations in Section 2.7, where G,Δ and Q vanish on $\mathbb{R}_>$, it is reasonable to believe and sustained by an iteration argument that $G=\Delta*\mathbb{U}$. In fact, if $\Delta*\mathbb{U}$ exists, this only takes to verify that $\lim_{n\to\infty} G*Q^{*n}(t)=0$ for all $t\ge 0$, as

$$G(t) \;=\; \sum_{k=0}^{n-1} \Delta * Q^{*k}(t) \;+\; G * Q^{*n}(t), \quad t \in \mathbb{R}$$

for each $n \in \mathbb{N}$. But with $(S_n)_{n\ge 0}$ denoting a SRW with increment distribution Q and thus negative drift, it follows indeed that

$$\lim_{n\to\infty} G * Q^{*n}(t) \;=\; \lim_{n\to\infty} \mathbb{E}G(t - S_n) \;=\; 0$$

by an appeal to the dominated convergence theorem and $G(t) \to 0$ as $t \to \infty$.

So far we have not really gained any new insight because an application of the key renewal theorem 2.67 to $G(t) = \Delta * \mathbb{U}(t)$, if possible, only reconfirms what we already know, namely that $G(t) \to 0$ as $t \to \infty$. On the other hand, defining $G_\theta(t) := e^{\theta t} G(t)$, $\Delta_\theta(t) := e^{\theta t}\Delta(t)$ and $Q_\theta(dx) := e^{\theta x} Q(dx)$ for $\theta \in \mathbb{R}$, we find as in Lemma 2.43 that G_θ solves a renewal equation as well, viz.

$$G_\theta(t) \;=\; \Delta_\theta(t) + \int G_\theta(t-x)\; Q_\theta(dx), \quad t \in \mathbb{R}.$$

Hence, if Q possesses a *characteristic exponent* ϑ, defined by the unique (if it exists) value $\ne 0$ such that $\phi_Q(\vartheta) = \int e^{\vartheta x} Q(dx) = 1$[2], it appears to be natural to use this renewal equation with $\theta = \vartheta$ which should lead to $G_\vartheta = \Delta_\vartheta * \mathbb{U}_\vartheta$, $\mathbb{U}_\vartheta := \sum_{n\ge 0} Q_\vartheta^{*n}$, and then to the conclusion that $G_\vartheta(t)$ converges to a constant as $t \to \infty$ by an appeal to the key renewal theorem, thus

$$\lim_{t\to\infty} e^{\vartheta t} G(t) = C$$

for some $C \in \mathbb{R}$ which in the best case is $\ne 0$. Naturally, further conditions must be imposed to make this work for us. They are stated in the following proposition together with the expected conclusion. Let us mention that ϑ, if it exists, is necessarily positive and that Q_ϑ has positive, possibly infinite mean. This follows from the fact that the mgf of Q, i.e. $\phi_Q(\theta) = \int e^{\theta x} Q(dx)$, is convex on its natural domain $\mathbb{D}_Q$ and that $\phi'(0) = \int x\,Q(dx) < 0$.

Proposition 4.7. *In addition to the assumptions on G, Δ and Q stated at the beginning of this section suppose that Q is nonarithmetic and possesses a characteristic exponent $\vartheta > 0$ and let $\mu_\vartheta := \int x e^{\vartheta x}\, Q(dx)$ denote the (positive) mean of Q_ϑ. Also assume that Δ is dRi and*

$$\int_{-\infty}^{\infty} e^{\vartheta x}\, |\Delta(x)|\, dx \;<\; \infty. \tag{4.20}$$

Then $G = \Delta + G * Q$ implies $G = \Delta * \mathbb{U}$ and

[2] In Subsection 2.7.1, a slightly different definition has been used for bounded measures on $\mathbb{R}_\ge$

$$\lim_{t\to\infty} e^{\vartheta t} G(t) \;=\; \frac{1}{\mu_\vartheta}\int_{-\infty}^{\infty} e^{\vartheta x}\Delta(x)\,dx, \qquad (4.21)$$

which, by our usual convention, equals 0 *if* $\mu_\vartheta = \infty$.

Proof. First note that, since Δ is dRi and (4.20) holds, the function Δ_ϑ is dRi as well [☞ Problem 4.8]. This in combination with the uniform local boundedness of $\mathbb{U}_\vartheta$ [☞ Lemma 2.64] implies that $\Delta_\vartheta * \mathbb{U}_\vartheta$ is everywhere finite. We have already argued above that $G = \Delta * \mathbb{U}$ so that $G_\vartheta = \Delta_\vartheta * \mathbb{U}_\vartheta$. Therefore, assertion (4.21) follows by an appeal to the (nonarithmetic version of the) key renewal theorem 2.67. □

The previous result should be kept in mind as a kind of general version of what is actually derived for special triples G,Δ and Q in the proof of the implicit renewal theorem we are now going to prove.

Problems

Problem 4.8. [☞ also Lemma 2.28] Let $g:\mathbb{R}\to\mathbb{R}$ be a dRi function and $\theta\in\mathbb{R}$ be such that $g_\theta(x) = e^{\theta x} g(x)$ is $\lambda\!\!\lambda_0$-integrable. Prove that g_θ is then dRi as well.

Problem 4.9. [Two-sided renewal equation] Prove that, given a two-sided renewal equation $G = g + G*Q$ with a dRi function g and a probability measure Q on $\mathbb{R}$, the set of solutions equals

$$\big\{a + g*\mathbb{U} : a\in\mathbb{R}\big\},$$

where $\mathbb{U}$ denotes the renewal measure of Q.

4.3 Proof of the implicit renewal theorem

It suffices to show (4.5) and the formula for C_+ in the respective parts (a) and (b) because the other assertions follow by considering $-X$ instead of X. The proof will be carried out for the three cases

$$M\ge 0 \text{ a.s.},\quad \mathbb{P}(M>0)\wedge\mathbb{P}(M<0)>0 \quad\text{and}\quad M\le 0 \text{ a.s.}$$

separately and frequently make use of the following notation most of which has already been used earlier. Let $X, M, M_1, M_2, \ldots$ be independent random variables on a common probability space $(\Omega,\mathfrak{A},\mathbb{P})$ such that $M, M_1, M_2, \ldots$ are further identically distributed. Then

$$\Pi_0 := 1 \quad\text{and}\quad \Pi_n := \prod_{k=1}^{n} M_k \quad\text{for } n\ge 1,$$

$$\xi := \log|M|, \quad Q := \mathscr{L}(\xi),$$

$$\xi_n := \log|M_n|, \quad S_0 := 0 \quad \text{and} \quad S_n := \log|\Pi_n| = \sum_{k=1}^{n} \xi_k \quad \text{for } n \geq 1,$$

$$\mathbb{U} := \sum_{n\geq 0} Q^{*n} = \sum_{n\geq 0} \mathbb{P}(S_n \in \cdot), \quad \mathbb{U}_\theta(dx) := e^{\theta x}\,\mathbb{U}(dx),$$

$$G(t) := \mathbb{P}(X > e^t), \quad G_\theta(t) := e^{\theta t}\,G(t),$$

$$\Delta(t) := \mathbb{P}(X > e^t) - \mathbb{P}(MX > e^t), \quad \Delta_\theta(t) := e^{\theta t}\Delta(t) \quad \text{for } \theta, t \in \mathbb{R},$$

$$\overline{f}(t) := \int_{(-\infty,t]} e^{-(t-x)} f(x)\,\lambda_0(dx) \;=\; \mathbb{E}f(t-Z) \quad \text{for suitable } f:\mathbb{R}\to\mathbb{R},$$

where Z is a standard exponential random variable. The function $\overline{f}$ has already been introduced in Subsection 2.6.1 and called *exponential smoothing of* f. Recall from Lemma 2.30 there that $\overline{f}$ is dRi whenever $f \in L^1$. The next simple lemma further shows that exponential smoothing is preserved under convolutions with measures.

Lemma 4.10. *Let $f:\mathbb{R}\to\mathbb{R}$ be a measurable function and V a finite measure on $\mathbb{R}$ such that $\overline{f}$ as well as $f*V$ exist as real-valued functions on $\mathbb{R}$. Then*

$$\overline{f}*V = \overline{f*V}. \tag{4.22}$$

and $f = g + f*V$ for a measurable function g with exponential smoothing $\overline{g}$ implies $\overline{f} = \overline{g} + \overline{f}*V$.

Proof. W.l.o.g. suppose that $\|V\| = 1$. Let Y, Z be independent random variables such that $\mathscr{L}(Y) = V$ and $\mathscr{L}(Z) = Exp(1)$. For (4.22), it then suffices to note that

$$\overline{f}*V(t) \;=\; \mathbb{E}f(t-Y-Z) \;=\; \overline{f*V}(t)$$

for all $t \in \mathbb{R}$, while the last assertion then follows from

$$\overline{f} \;=\; \overline{g+f*V} \;=\; \overline{g} + \overline{f*V} \;=\; \overline{g} + \overline{f}*V.$$

having used that exponential smoothing is a linear operation. □

By combing this lemma with next one, we will be able to use exponential smoothing when studying the asymptotic properties of the function G_ϑ in the proof of the implicit renewal theorem.

Lemma 4.11. [Smoothing lemma] *If $\mathbb{P}(X > t)$ satisfies*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t x^\vartheta\,\mathbb{P}(X > x)\,dx \;=\; C_+$$

for some $\vartheta > 0$ *and* $C_+ \in \mathbb{R}_\geq$, *then* (4.5) *holds true as well.*

In other words, if the Césaro smoothing of $e^{\vartheta t}\,\mathbb{P}(X > t)$ converges to some C_+, then so does the function $e^{\vartheta t}\,\mathbb{P}(X > t)$ itself as $t \to \infty$.

Proof. Fixing any $b > 1$, we infer

$$\begin{aligned} C_+(b-1)t \;&\simeq\; \int_0^{bt} x^\vartheta\,\mathbb{P}(X > x)\,dx \;-\; \int_0^t x^\vartheta\,\mathbb{P}(X > x)\,dx \\ &=\; \int_t^{bt} x^\vartheta\,\mathbb{P}(X > x)\,dx \\ &\leq\; \mathbb{P}(X > t)\int_t^{bt} x^\vartheta\,dx \\ &=\; \frac{b^{\vartheta+1}-1}{\vartheta+1}\,t^{\vartheta+1}\,\mathbb{P}(X > t) \end{aligned}$$

and thereby

$$\liminf_{t\to\infty} t^\vartheta\,\mathbb{P}(X > t) \;\geq\; C_+(\vartheta+1)\frac{b-1}{b^{\vartheta+1}-1}.$$

Now let b tend to 1 and use

$$\lim_{b\downarrow 1}\frac{b-1}{b^{\vartheta+1}-1} \;=\; \frac{1}{\vartheta+1}$$

to conclude $\liminf_{t\to\infty} t^\vartheta\,\mathbb{P}(X > t) \geq C_+$.

By an analogous argument for $0 < b < 1$, one finds that

$$C_+(1-b)t \;\simeq\; \int_{bt}^t x^\vartheta\,\mathbb{P}(X > x)\,dx \;\geq\; \frac{1-b^{\vartheta+1}}{\vartheta+1}\,t^{\vartheta+1}\,\mathbb{P}(X > t)$$

which upon letting b again tend to 1 yields $\limsup_{t\to\infty} t^\vartheta\,\mathbb{P}(X > t) \leq C_+$. ☐

In view of this lemma, it suffices to verify $t^{-1}\int_0^t x^\vartheta\,\mathbb{P}(X > x)\,dx \to C_+$ as $t \to \infty$ instead of (4.5), and since

$$\begin{aligned} \frac{1}{t}\int_0^t x^\vartheta\,\mathbb{P}(X > x)\,dx \;&=\; \frac{1}{t}\int_{-\infty}^{\log t} e^{(\vartheta+1)x}\,\mathbb{P}(Y > e^x)\,dx \\ &=\; \frac{1}{t}\int_{-\infty}^{\log t} e^x\,G_\vartheta(x)\,dx \\ &=\; \int_{-\infty}^{\log t} e^{-(\log t - x)}\,G_\vartheta(x)\,dx \\ &=\; \overline{G_\vartheta}(\log t) \end{aligned}$$

this means to show that

$$\lim_{t\to\infty}\overline{G_\vartheta}(t) = C_+.$$

4.3.1 The case when $M \geq 0$ a.s.

Rewriting (IRT-1)-(IRT3) in terms of $\xi = \log M$, we have

(IRT-1) $\|Q_\vartheta\| = \mathbb{E}e^{\vartheta\xi} = 1$.
(IRT-2) $\int_{\mathbb{R}_\geq} x\,Q_\vartheta(dx) = \mathbb{E}e^{\vartheta\xi}\xi^+ < \infty$
[thus $0 < \mu_\vartheta = \int x\,Q_\vartheta(dx) = \mathbb{E}e^{\vartheta\xi}\xi < \infty$ as explained before Prop. 4.7].
(IRT-3) Q_ϑ is nonarithmetic.

We have already argued for this case that $G = \Delta + G * Q$, $G = \Delta * \mathbb{U}$ and thus $G_\vartheta = \Delta_\vartheta * \mathbb{U}_\vartheta$. As

$$\int_{-\infty}^{\infty} |\Delta_\vartheta(x)|\,dt \;=\; \int_{-\infty}^{\infty} \big|\mathbb{P}(X > e^x) - \mathbb{P}(MX > e^x)\big|\,e^{\vartheta x}\,dx$$
$$=\; \int_{-\infty}^{\infty} \big|\mathbb{P}(X > t) - \mathbb{P}(MX > t)\big|\,t^{\vartheta-1}\,dt,$$

we see that $\Delta_\vartheta \in L^1$ by (4.3), and also (when removing absolute values)

$$\frac{1}{\mu_\vartheta}\int_{-\infty}^{\infty} \Delta_\vartheta(x)\,dx \;=\; C_+.$$

By Lemma 2.30 and (2.23), $\overline{\Delta_\vartheta}$ is dRi and $\mu_\vartheta^{-1}\int_{-\infty}^{\infty}\overline{\Delta_\vartheta}(x)\,dx = C_+$ as well. Now use Lemma 4.10, the smoothing lemma 4.11 and the key renewal theorem 2.67 to conclude

$$\lim_{t\to\infty} G_\vartheta(t) \;=\; \lim_{t\to\infty}\overline{G_\vartheta}(t) \;=\; \lim_{t\to\infty}\overline{\Delta_\vartheta} * \mathbb{U}_\vartheta(t) \;=\; \frac{1}{\mu_\vartheta}\int_{-\infty}^{\infty}\overline{\Delta_\vartheta}(x)\,dx \;=\; C_+$$

as claimed.

4.3.2 The case when $\mathbb{P}(M > 0) \wedge \mathbb{P}(M < 0) > 0$

The main idea for the proof of the remaining two cases is to reduce it to the first case by comparison of X with $\Pi_\sigma X$, where

$$\sigma \;:=\; \inf\{n \geq 1 : \Pi_n \geq 0\} \;=\; \begin{cases} 1, & \text{if } M_1 \geq 0, \\ \inf\{n \geq 2 : M_n \leq 0\}, & \text{otherwise.} \end{cases}$$

Plainly, σ is a.s. finite, and we may thus hope to be successful in our endeavor if Π_σ satisfies (IRT-1)-(IRT-3).

The reader should keep in mind that, besides (IRT-1)-(IRT-3) for M, we are now *always* assuming (4.3) *and* (4.4), or, equivalently [☞ Problem 4.5],

$$\int_{-\infty}^{\infty} \Delta_\vartheta^*(x)\,dx \;=\; \int_0^{\infty} \big|\mathbb{P}(|X| > t) - \mathbb{P}(|MX| > t)\big|\,t^{\vartheta-1}\,dt \;<\; \infty,$$

where $\Delta^*(x) := e^{\vartheta x}\Delta^*(x)$ as usual and

$$\Delta^*(x) := \mathbb{P}(|X| > e^x) - \mathbb{P}(|MX| > e^x), \quad x \in \mathbb{R}.$$

We begin with a lemma that verifies (IRT-1)-(IRT-3) for Π_σ.

Lemma 4.12. *The stopped product $\Pi_\sigma = e^{S_\sigma}$ satisfies the conditions (IRT-1)-(IRT-3), i.e.*

$$\mathbb{E}\Pi_\sigma^\vartheta = \mathbb{E}e^{\vartheta S_\sigma} = 1, \quad \mathbb{E}\Pi_\sigma^\vartheta \log^+ \Pi_\sigma = \mathbb{E}e^{\vartheta S_\sigma} S_\sigma^+ < \infty,$$

and the law of $\log \Pi_\sigma$ *given* $\Pi_\sigma \neq 0$ *is nonarithmetic. Moreover,*

$$\mathbb{E}\Pi_\sigma^\vartheta \log \Pi_\sigma = \mathbb{E}e^{\vartheta S_\sigma} S_\sigma = 2\mu_\vartheta.$$

Proof. First note that $(|\Pi_n|^\vartheta)_{n\ge 0}$ constitutes a nonnegative mean one product martingale with respect to the filtration $\mathscr{F}_n := \sigma(\Pi_0, M_1, ..., M_n)$ for $n \ge 0$. As usual, put $\mathscr{F}_\infty := \sigma(\bigcup_{n\ge 0} \mathscr{F}_n)$. As in the proof of Theorem 2.68, define a new probability measure $\widehat{\mathbb{P}}$ on $(\Omega, \mathscr{F}_\infty)$ by

$$\widehat{\mathbb{P}}(A) := \mathbb{E}|\Pi_n|^\vartheta \mathbf{1}_A = \mathbb{E}e^{\vartheta S_n}\mathbf{1}_A \quad \text{for } A \in \mathscr{F}_n \text{ and } n \ge 0.$$

Then $M_1, M_2, ...$ are still iid under $\widehat{\mathbb{P}}$ with common distribution

$$\widehat{\mathbb{P}}(M_1 \in B) = \mathbb{E}|M_1|^\vartheta \mathbf{1}_B(M_1), \quad B \in \mathscr{B}(\mathbb{R}).$$

Equivalently, $(S_n)_{n\ge 0}$ remains a SRW under $\widehat{\mathbb{P}}$ with increment distribution

$$\widehat{\mathbb{P}}(\xi_1 \in B) = \mathbb{E}e^{\vartheta \xi_1}\mathbf{1}_B(\xi_1) \quad , B \in \mathscr{B}(\mathbb{R}),$$

and drift $\widehat{\mathbb{E}}\xi_1 = \mathbb{E}e^{\vartheta\xi_1}\xi_1 = \mu_\vartheta$. It is shown in Problem 4.15 that σ is a.s. finite and has finite moments of any order under $\widehat{\mathbb{P}}$. The almost sure finiteness ensures that, for any $A \in \mathscr{F}_\sigma$,

$$\widehat{\mathbb{P}}(A) = \sum_{n\ge 1} \widehat{\mathbb{P}}(A \cap \{\sigma = n\}) = \sum_{n\ge 1} \mathbb{E}|\Pi_n|^\vartheta \mathbf{1}_{A\cap\{\sigma=n\}} = \mathbb{E}|\Pi_\sigma|^\vartheta \mathbf{1}_A,$$

for $A \cap \{\sigma = n\} \in \mathscr{F}_n$ for each $n \ge 1$. Choosing $A = \Omega$, we particularly find that $\mathbb{E}|\Pi_\sigma|^\vartheta = 1$. Next, use the $\mathscr{F}_\sigma$-measurability of S_σ and Wald's equation to infer

$$\mathbb{E}e^{\vartheta S_\sigma} S_\sigma^+ \;=\; \widehat{\mathbb{E}}S_\sigma^+ \;\le\; \widehat{\mathbb{E}}\left(\sum_{k=1}^{\sigma} \xi_k^+\right) \;=\; \widehat{\mathbb{E}}\xi_1^+ \, \widehat{\mathbb{E}}\sigma$$

which is finite because $\widehat{\mathbb{E}}\sigma < \infty$ and $\widehat{\mathbb{E}}\xi_1^+ = \mathbb{E}|M|^\vartheta \log^+ |M| < \infty$ by (IRT-2). But then, by another use of Wald's equation,

$$\mathbb{E}e^{\vartheta S_\sigma}S_\sigma \;=\; \widehat{\mathbb{E}}S_\sigma \;=\; \widehat{\mathbb{E}}\xi_1\,\widehat{\mathbb{E}}\sigma \;=\; \mu_\vartheta\,\widehat{\mathbb{E}}\sigma \;=\; 2\mu_\vartheta,$$

where $\widehat{\mathbb{E}}\sigma = 2$ is again shown as a part of Problem 4.15.

Finally, use [and prove as part (c) of Problem 4.15] that

$$\mathbb{P}(S_\sigma \in \cdot, \Pi_\sigma \neq 0) = p\,Q_> + (1-p)^2 \sum_{n\geq 0} p^n\, Q_>^{*n} * Q_<^{*2}, \tag{4.23}$$

where $p := \mathbb{P}(M>0)$, $Q_> := \mathbb{P}(\xi \in \cdot|M>0)$ and $Q_< := \mathbb{P}(\xi \in \cdot|M<0)$. Since, by (IRT-3), at least one of $Q_<$ or $Q_>$ is nonarithmetic, the same must hold for the conditional law $\mathbb{P}(S_\sigma \in \cdot|\Pi_\sigma \neq 0)$ as one may easily deduce with the help of FT's [☞ again Problem 4.15]. □

Lemma 4.13. *If* (4.3), (4.4)*, and thus* (4.18) *are valid, then*

$$\int_0^\infty \big|\mathbb{P}(|X|>t) - \mathbb{P}(|\Pi_\sigma X|>t)\big|\, t^{\vartheta-1}\,dt \;<\; \infty. \tag{4.24}$$

holds true as well and, furthermore,

$$\frac{1}{2\mu_\vartheta}\int_0^\infty \big(\mathbb{P}(|X|>t) - \mathbb{P}(|\Pi_\sigma X|>t)\big)\, t^{\vartheta-1}\,dt \;=\; C \tag{4.25}$$

for C as defined in (4.9).

Proof. First observe that, for all $t \geq 0$,

$$\begin{aligned}
&\big|\mathbb{P}(|X|>t) - \mathbb{P}(|\Pi_\sigma X|>t)\big| \\
&\quad= \lim_{m\to\infty}\big|\mathbb{P}(|X|>t) - \mathbb{P}(|\Pi_{\sigma\wedge m} X|>t)\big| \\
&\quad= \lim_{m\to\infty}\left|\mathbb{E}\left(\sum_{n=1}^{\sigma\wedge m} \mathbf{1}_{(t,\infty)}(|\Pi_{n-1}X|) - \mathbf{1}_{(t,\infty)}(|\Pi_n X|)\right)\right| \\
&\quad= \lim_{m\to\infty}\left|\mathbb{E}\left(\sum_{n=1}^{m} \mathbf{1}_{\{\sigma\geq n\}}\Big(\mathbf{1}_{(t,\infty)}(|\Pi_{n-1}X|) - \mathbf{1}_{(t,\infty)}(|\Pi_n X|)\Big)\right)\right| \\
&\quad\leq \sum_{n\geq 1}\Big|\mathbb{P}(\sigma\geq n, |\Pi_{n-1}X|>t) - \mathbb{P}(\sigma\geq n, |\Pi_n X|>t)\Big|.
\end{aligned}$$

Consequently, defining $P_n(ds) := \mathbb{P}(\sigma \geq n, \Pi_{n-1} \in ds)$ for $n \geq 1$, we obtain

$$\begin{aligned}
&\int_0^\infty \big|\mathbb{P}(|X|>t) - \mathbb{P}(|\Pi_\sigma X|>t)\big|\, t^{\vartheta-1}\,dt \qquad (4.26)\\
&\quad\leq \sum_{n\geq 1}\int_0^\infty \Big|\mathbb{P}(\sigma\geq n, |\Pi_{n-1}X|>t) - \mathbb{P}(\sigma\geq n, |\Pi_n X|>t)\Big|\, t^{\vartheta-1}\,dt \\
&\quad= \sum_{n\geq 1}\int_0^\infty \left|\int_{\mathbb{R}_>} \mathbb{P}\left(|X|>\frac{t}{s}\right) - \mathbb{P}\left(|MX|>\frac{t}{s}\right)\, P_n(ds)\right|\, t^{\vartheta-1}\,dt
\end{aligned}$$

$$\le \sum_{n\ge 1}\int_{\mathbb{R}_>}\int_0^\infty \left|\mathbb{P}\left(|X|>\frac{t}{s}\right)-\mathbb{P}\left(|MX|>\frac{t}{s}\right)\right|\left(\frac{t}{s}\right)^{\vartheta-1} dt\ s^{\vartheta-1}P_n(ds)$$
$$= \left(\int_0^\infty \big|\mathbb{P}(|X|>t)-\mathbb{P}(|MX|>t)\big|\,t^{\vartheta-1}\,dt\right)\sum_{n\ge 1}\mathbb{E}\mathbf{1}_{\{\sigma\ge n\}}|\Pi_{n-1}|^{\vartheta},$$

where the independence of $(\{\sigma\ge n\},\Pi_{n-1})\in\sigma(\Pi_0,M_1,...,M_{n-1})$ and (M_n,X) has been utilized for the third line and the change of variables $t/s\rightsquigarrow t$ for the last one. In view of the fact that (4.18) holds true, it remains to verify for (4.24) that the last series, which may also be written as $\mathbb{E}(\sum_{n=1}^{\sigma}|\Pi_{n-1}|^{\vartheta})$, is finite. To this end, let $\widehat{\mathbb{P}}$ be defined as in the proof of Lemma 4.12. Then

$$a_n := \mathbb{E}\mathbf{1}_{\{\sigma\ge n\}}|\Pi_{n-1}|^{\vartheta} = \widehat{\mathbb{P}}(\sigma\ge n)$$

for each $n\ge 1$, because $\{\sigma\ge n\}\in\mathscr{F}_{n-1}$ and therefore [☞ Problem 4.15(b)]

$$\sum_{n\ge 1}a_n = \widehat{\mathbb{E}}\sigma = 2,$$

which completes the proof of (4.24). But by now repeating the calculation in (4.26) without absolute value signs, all inequalities turn into equalities, giving

$$\int_0^\infty \big(\mathbb{P}(|X|>t)-\mathbb{P}(|\Pi_\sigma X|>t)\big)\,t^{\vartheta-1}\,dt$$
$$= \left(\int_0^\infty \big(\mathbb{P}(|X|>t)-\mathbb{P}(|MX|>t)\big)\,t^{\vartheta-1}\,du\right)\sum_{n\ge 1}\mathbb{E}\mathbf{1}_{\{\sigma\ge n\}}|\Pi_{n-1}|^{\vartheta}$$
$$= 2\int_0^\infty \big(\mathbb{P}(|X|>t)-\mathbb{P}(|MX|>t)\big)\,t^{\vartheta-1}\,dt$$

and so (4.25) upon multiplication with $(2\mu_\vartheta)^{-1}$. □

Finally, we must verify condition (4.3) when substituting M for Π_σ and provide the formula that replaces (4.7) in this case.

Lemma 4.14. *Under the same assumptions as in the previous lemma,*

$$\int_0^\infty \big|\mathbb{P}(X>t)-\mathbb{P}(\Pi_\sigma X>t)\big|\,t^{\vartheta-1}\,dt \;<\; \infty \tag{4.27}$$

as well as

$$\frac{1}{2\mu_\vartheta}\int_0^\infty \big(\mathbb{P}(X>t)-\mathbb{P}(\Pi_\sigma X>t)\big)\,t^{\vartheta-1}\,dt \;=\; \frac{C}{2}. \tag{4.28}$$

Proof. We leave it as an exercise to first verify (4.27) along similar lines as in (4.26) [☞ Problem 4.16]. Keeping the notation of the proof of the previous lemma, we then obtain (using $\Pi_{n-1} = -|\Pi_{n-1}|$ on $\{\sigma \geq n\}$ for any $n \geq 2$ and $\sum_{n\geq 2} a_n = 1$)

$$\begin{aligned}
&\int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(\Pi_\sigma X > t)\big)\, t^{\vartheta-1}\, dt \\
&\quad = \sum_{n\geq 1} \int_0^\infty \Big(\mathbb{P}(\sigma \geq n, \Pi_{n-1}X > t) - \mathbb{P}(\sigma \geq n, \Pi_n X > t)\Big)\, t^{\vartheta-1}\, dt \\
&\quad = \int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(MX>t)\big)\, t^{\vartheta-1}\, dt \\
&\quad + \sum_{n\geq 2} \int_0^\infty \Big(\mathbb{P}(\sigma \geq n, |\Pi_{n-1}|X < -t) - \mathbb{P}(\sigma \geq n, |\Pi_{n-1}|M_n X < -t)\Big)\, t^{\vartheta-1}\, dt \\
&\quad = \int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(MX>t)\big)\, t^{\vartheta-1}\, dt \\
&\quad + \sum_{n\geq 2} \int_{\mathbb{R}_>} \int_0^\infty \left(\mathbb{P}\left(X < -\frac{t}{s}\right) - \mathbb{P}\left(MX > -\frac{t}{s}\right)\right) \left(\frac{t}{s}\right)^{\vartheta-1} dt\; s^{\vartheta-1}\, P_n(ds) \\
&\quad = \int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(MX>t)\big)\, t^{\vartheta-1}\, dt \\
&\quad + \left(\int_0^\infty \big(\mathbb{P}(X<-t) - \mathbb{P}(MX<-t)\big)\, t^{\vartheta-1}\, dt\right) \sum_{n\geq 2} a_n \\
&\quad = \int_0^\infty \big(\mathbb{P}(|X|>t) - \mathbb{P}(|MX|>t)\big)\, t^{\vartheta-1}\, dt
\end{aligned}$$

and therefore

$$\frac{1}{2\mu_\vartheta} \int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(\Pi_\sigma X > t)\big)\, t^{\vartheta-1}\, dt \;=\; \frac{C}{2}$$

as claimed. □

Taking a deep breath, we are finally able to settle the present case by using part (a) of the theorem upon replacing M with $\Pi_\sigma \geq 0$. Lemma 4.12 ensures validity of (IRT-1)-(IRT-3) under this replacement and also that $2\mu_\vartheta$ takes the place of μ_ϑ. Condition (4.3) now turns into (4.27), which has been verified as part of Lemma 4.14. Therefore, we conclude

$$C_+ \;=\; \lim_{t\to\infty} t^\vartheta\, \mathbb{P}(X>t) \;=\; \frac{1}{2\mu_\vartheta} \int_0^\infty \big(\mathbb{P}(X>t) - \mathbb{P}(\Pi_\sigma > t)\big)\, t^{\vartheta-1}\, dt,$$

and, by (4.28), the last expression equals $C/2$ as asserted.

4.3.3 The case when $M \leq 0$ *a.s.*

This case is handled by the same reduction argument as the previous one, but is considerably simpler because of the obvious fact that $\sigma \equiv 2$ holds true here. We leave it to the reader to check all necessary conditions as well as to show that (4.25) and (4.28) remain valid [☞ Problem 4.17].

Problems

Problem 4.15. Given the assumptions of Subsection 4.3.2, prove the following assertions:

(a) $\mathbb{P}(\sigma-2\in\cdot|\sigma\ge 2)=Geom(\theta)$ with $\theta=\mathbb{P}(M\le 0)$.
(b) $\widehat{\mathbb{P}}(\sigma-2\in\cdot|\sigma\ge 2)=Geom(\hat\theta)$ with $\hat\theta=\mathbb{E}|M|^{\vartheta}\mathbf{1}_{\{M<0\}}$, and $\widehat{\mathbb{E}}\sigma=2$.
(c) The conditional law under $\mathbb{P}$ of S_σ given $\Pi_\sigma\ne 0$ satisfies (4.23).
(d) Compute the FT of $\mathbb{P}(S_\sigma\in\cdot|\Pi_\sigma\ne 0)$ in terms of those of $Q_<,Q_>$ and use it to show that this law is nonarithmetic.

Problem 4.16. Give a proof of (4.27) under the assumptions of Lemma 4.14.

Problem 4.17. Give a proof of the implicit renewal theorem for the case $M\le 0$ a.s.

Problem 4.18. [Tail behavior at 0] Prove the following version of the implicit renewal theorem:
Let M,X be independent random variables taking values in $\overline{\mathbb{R}}\backslash\{0\}$ such that, for some $\vartheta>0$,

(IRT2-1) $\mathbb{E}|M|^{-\vartheta}=1$.
(IRT2-2) $\mathbb{E}|M|^{-\vartheta}\log^-|M|<\infty$.
(IRT2-3) The conditional law $\mathbb{P}(\log|M|\in\cdot||M|<\infty)$ of $\log|M|$ given $|M|<\infty$ is nonarithmetic, in particular, $\mathbb{P}(|M|=1)<1$.

Then $0<\mathbb{E}\log|M|\le\infty$, $0<\mu_\vartheta:=-\mathbb{E}|M|^{-\vartheta}\log|M|<\infty$, and the following assertions hold true:

(a) Suppose M is a.s. positive. If

$$\int_0^\infty \big|\mathbb{P}(X\le t)-\mathbb{P}(MX\le t)\big|\,t^{-1-\vartheta}\,dt\;<\;\infty \tag{4.29}$$

or, respectively,

$$\int_0^\infty \big|\mathbb{P}(X\ge -t)-\mathbb{P}(MX\ge -t)\big|\,t^{-1-\vartheta}\,dt\;<\;\infty, \tag{4.30}$$

then

$$\lim_{t\to 0+} t^{-\vartheta}\,\mathbb{P}(0<X\le t)\;=\;C_+, \tag{4.31}$$

respectively

$$\lim_{t\to 0+} t^{\vartheta}\,\mathbb{P}(-t\le X<0)\;=\;C_-, \tag{4.32}$$

where C_+ and C_- are given by the equations

$$C_+ := \frac{1}{\mu_\vartheta}\int_0^\infty \big(\mathbb{P}(X\le t)-\mathbb{P}(MX\le t)\big)\,t^{-1-\vartheta}\,dt, \tag{4.33}$$

$$C_- := \frac{1}{\mu_\vartheta}\int_0^\infty \big(\mathbb{P}(X\ge -t)-\mathbb{P}(MX\ge -t)\big)\,t^{-1-\vartheta}\,dt. \tag{4.34}$$

(b) If $\mathbb{P}(M<0)>0$ and (4.29), (4.30) are both satisfied, then (4.31) and (4.32) hold with $C_+=C_-=C/2$, where

$$C := \frac{1}{\mu_\vartheta}\int_0^\infty \big(\mathbb{P}(|X|\le t)-\mathbb{P}(|MX|\le t)\big)\,t^{-1-\vartheta}\,dt. \tag{4.35}$$

4.4 Applications

We will proceed with an application of the previously developed results to a number of examples some of which are also discussed in [58]. In view of the fact that implicit renewal theory deals with the tail behavior of solutions to SFPE's of the form (4.1) and embarks on linear approximation of the random function Ψ involved, the simplest and most natural example that comes to mind is a RDE and therefore studied first.

4.4.1 Random difference equations and perpetuities

Returning to the situation described in Section 1.5, let $(M,Q),(M_1,Q_1),(M_2,Q_2),\ldots$ be iid two-dimensional random variables and $(X_n)_{n\ge 0}$ recursively defined by the (one-dimensional) RDE

$$X_n := M_nX_{n-1}+Q_n,\quad n\ge 1.$$

As usual, let $\Pi_0 := 0$ and $\Pi_n := M_1\cdot\ldots\cdot M_n$ for $n\ge 1$. If $\mathbb{E}\log|M|<0$ and $\mathbb{E}\log^+|Q|<\infty$, then $(X_n)_{n\ge 0}$ is a mean contractive IFS on $\mathbb{R}$ satisfying the jump-size condition (3.17) (with $x_0=0$ and $d(x,y)=|x-y|$). By Theorem 3.24, it is then convergent in distribution (under any initial distribution) to the unique solution of the SFPE

$$X \stackrel{d}{=} MX+Q,\qquad X \text{ independent of } (M,Q), \tag{4.36}$$

which is (the law of) the perpetuity $X:=\sum_{n\ge 1}\Pi_{n-1}Q_n$ and in turn is obtained as the a.s. limit (under any initial distribution) of the backward iterations. By applying the same arguments to the RDE

$$Y_n = |M_n|Y_{n-1}+|Q_n|,\quad n\ge 1,$$

we see that $Y:=\sum_{n\ge 1}|\Pi_{n-1}Q_n|$ is a.s. finite as well and its law the unique solution to the SFPE

$$Y = |M|Y + |Q|, \qquad Y \text{ independent of } (M,Q).$$

An application of the implicit renewal theorem provides us with the following result about the tail behavior of X under appropriate conditions on M and Q. Its far more difficult extension to the multidimensional situation is a famous result due to KESTEN [69].

Theorem 4.19. *Suppose that M satisfies (IRT-1)-(IRT-3) and that $\mathbb{E}|Q|^\vartheta < \infty$. Then there exists a unique solution to the SFPE* (4.36)*, given by the law of the perpetuity $X := \sum_{n\ge1} \Pi_{n-1}Q_n$. This law satisfies* (4.5) *as well as* (4.6)*, where*

$$C_\pm = \frac{\mathbb{E}\big(((MX+Q)^\pm)^\vartheta - ((MX)^\pm)^\vartheta\big)}{\vartheta\mu_\vartheta} \tag{4.37}$$

if $M \ge 0$ a.s., while

$$C_+ = C_- = \frac{\mathbb{E}\big(|MX+Q|^\vartheta - |MX|^\vartheta\big)}{2\vartheta\mu_\vartheta} \tag{4.38}$$

if $\mathbb{P}(M<0)>0$. Furthermore,

$$C_+ + C_- > 0 \quad \textit{iff} \quad \mathbb{P}(Q = c(1-M)) < 1 \quad \textit{for all } c \in \mathbb{R}. \tag{4.39}$$

A crucial ingredient to the proof of this theorem is the following moment result that will enable us to verify validity of (4.3) and (4.4) of the implicit renewal theorem. For a.s. nonnegative M,Q and $\kappa > 1$, it was obtained by VERVAAT [111]; for a stronger version see [4] and Problem 4.33.

Proposition 4.20. *Suppose that $\mathbb{E}|M|^\kappa \le 1$ and $\mathbb{E}|Q|^\kappa < \infty$ for some $\kappa > 0$. Then $Y = \sum_{n\ge1} |\Pi_{n-1}Q_n|$ satisfies $\mathbb{E}Y^p < \infty$ for any $p \in (0,\kappa)$.*

Proof. This is actually a direct consequence of the more general Theorem 3.29, but we repeat the argument for the present situation because it is short and simple.

As argued earlier, $\mathbb{E}|M|^p < 1$ for any $p \in (0,\kappa)$. If $p \le 1$, the subadditivity of $x \mapsto x^p$ implies that

$$\mathbb{E}Y^p \;\le\; \sum_{n\ge1} \mathbb{E}|\Pi_{n-1}Q_n|^p \;\le\; \mathbb{E}|Q|^p \sum_{n\ge1} (\mathbb{E}|M|^p)^{n-1} \;\le\; \frac{\mathbb{E}|Q|^p}{1-\mathbb{E}|M|^p} \;<\; \infty,$$

whereas in the case $p > 1$ a similar estimation with the help of Minkowski's inequality yields

$$\|Y\|_p \;\le\; \sum_{n\ge1}\|\Pi_{n-1}Q_n\|_p \;=\; \|Q\|_p\sum_{n\ge1}\|M\|_p^{n-1} \;=\; \frac{\|Q\|_p}{1-\|M\|_p} \;<\; \infty.$$

This completes the proof. □

GRINCEVIČIUS [60] provided the following extension of Lévy's symmetrization inequalities that will be utilized in the proof of (4.39). Under the assumptions of Theorem 4.19, define $m_0 := med(X)$,

$$\begin{aligned}
\Pi_{k:n} &:= \prod_{j=k}^{n} M_j \quad \text{for } 1\le k\le n,\\
\widehat{X}_n &:= \sum_{k=1}^{n}\Pi_{k-1}Q_k, \quad \widehat{X}_{k:n} := \sum_{j=k}^{n}\Pi_{k:j-1}Q_j \quad \text{for } 1\le k\le n,\\
\widehat{X}_0^* &:= m_0, \quad \widehat{X}_n^* := \widehat{X}_n + \Pi_n m_0 \quad \text{for } n\ge1,\\
R_k &:= \widehat{X}_k + \Pi_k\, med\big(\widehat{X}_{k+1:n} + \Pi_{k+1:n}y\big) \quad \text{for } 1\le k\le n,\ y\in\mathbb{R}\\
U_n &:= \Pi_{n-1}\big(Q_n - m_0(1-M_n)\big) \quad \text{for } n\ge1.
\end{aligned}$$

where $\widehat{X}_{n+1:n} := 0$ and $\Pi_{n+1:n} := 1$ in the definition of R_n. The $\widehat{X}_n$ are obviously the backward iterations when $X_0 = 0$ and hence a.s. convergent to $X = \sum_{n\ge1}\Pi_{n-1}Q_n$.

Lemma 4.21. [Grincevičius] *With the given notation,*

$$\mathbb{P}\left(\max_{1\le k\le n} R_k > x\right) \;\le\; 2\,\mathbb{P}(\widehat{X}_n + \Pi_n y > x)$$

for all $x,y\in\mathbb{R}$.

Specializing to $y=0$, we obtain

$$\mathbb{P}\left(\max_{1\le k\le n}\Big(\widehat{X}_k + \Pi_k\, med\big(\widehat{X}_{k+1:n}\big)\Big) > x\right) \;\le\; 2\,\mathbb{P}(\widehat{X}_n > x)$$

for any $x\in\mathbb{R}$ and then upon letting $n\to\infty$

$$\mathbb{P}\left(\sup_{n\ge1}\widehat{X}_n^* > x\right) \;=\; \mathbb{P}\left(\sup_{n\ge1}\Big(\widehat{X}_n + \Pi_n m_0\Big) > x\right) \;\le\; 2\,\mathbb{P}(X > x), \tag{4.40}$$

because $\lim_{n\to\infty}\widehat{X}_{k+1:n} \stackrel{d}{=} X$ for any $k\ge1$. The same inequality holds, of course, with $-\widehat{X}_n, -X$ instead of $\widehat{X}_n, X$ whence

$$\mathbb{P}\left(\sup_{n\ge1}|\widehat{X}_n^*| > x\right) \;\le\; 2\,\mathbb{P}(|X| > x) \tag{4.41}$$

for all $x\in\mathbb{R}_\ge$.

Proof. Fixing any $x,y\in\mathbb{R}$, define

$$A_k := \{R_1\le x,\dots,R_{k-1}\le x, R_k > x\},$$
$$B_k := \left\{\widehat{X}_{k+1:n}+\Pi_{k+1:n}y \ge med\big(\widehat{X}_{k+1:n}+\Pi_{k+1:n}y\big)\right\}$$

for $k=1,\dots,n$. Observe that A_k and B_k are independent events with $\mathbb{P}(B_k)\ge 1/2$ for each k and that

$$\left\{\max_{1\le k\le n}\Big(\widehat{X}_k+\Pi_k\, med\big(\widehat{X}_{k+1:n}+\Pi_{k+1:n}y\big)\Big)>x\right\} \;=\; \sum_{k=1}^n A_k,$$

$$\left\{\widehat{X}_n+\Pi_n y > x\right\} \;\supset\; \sum_{k=1}^n A_k\cap B_k.$$

For the last inclusion we have used that, on $A_k\cap B_k$,

$$x \;<\; R_k \;\le\; \widehat{X}_k+\Pi_k\left(\widehat{X}_{k+1:n}+\Pi_{k+1:n}y\right) \;=\; \widehat{X}_n+\Pi_n y$$

for each $k=1,\dots,n$. Now

$$\mathbb{P}(\widehat{X}_n+\Pi_n y>x) \;\ge\; \sum_{k=1}^n \mathbb{P}(A_k)\,\mathbb{P}(B_k) \;\ge\; \frac{1}{2}\sum_{k=1}^n\mathbb{P}(A_k) \;=\; \frac{1}{2}\mathbb{P}\left(\max_{1\le k\le n} R_k > x\right)$$

proves the assertion. □

Proof (of Theorem 4.19). We first prove (4.5) and (4.6) for which, by Corollary 4.2, it suffices to verify (4.10) and (4.11). But since $-X$ satisfies the same SFPE as X when replacing (M,Q) with $(M,-Q)$, it is further enough to consider only the first of these two conditions, viz. $\mathbb{E}\big|((MX+Q)^+)^\vartheta-((MX)^+)^\vartheta\big|<\infty$.

By making use of the inequality

$$(x+y)^p \le x^p + p\,2^{p-1}\big(x^{p-1}y+xy^{p-1}\big)+y^p, \tag{4.42}$$

valid for all $x,y\in\mathbb{R}_\ge$ and $p>1$, we find that

$$\begin{aligned}&((MX+Q)^+)^\vartheta-((MX)^+)^\vartheta \;\le\; ((MX)^+ + Q^+)^\vartheta-((MX)^+)^\vartheta\\ &\le\; \begin{cases}(Q^+)^\vartheta, & \text{if } \vartheta\in(0,1],\\ (Q^+)^\vartheta + c_\vartheta\big(((MX)^+)^{\vartheta-1}Q^+ + (MX)^+(Q^+)^{\vartheta-1}\big), & \text{if } \vartheta>1,\end{cases}\end{aligned}$$

where $c_\vartheta := \vartheta 2^{\vartheta-1}$. A combination of $(MX)^+\le|MX|$, the independence of X and (M,Q), $\mathbb{E}|M|^\vartheta<\infty$, $\mathbb{E}|Q|^\vartheta<\infty$, and of $\mathbb{E}|X|^{\vartheta-1}<\infty$ if $\vartheta>1$ [by Prop. 4.20] hence implies that

$$\mathbb{E}\left(((MX+Q)^+)^\vartheta-((MX)^+)^\vartheta\right)^+<\infty.$$

Indeed, if $\vartheta>1$, the obtained bound is

$$\mathbb{E}(Q^+)^\vartheta + c_\vartheta\,\mathbb{E}(|M|^{\vartheta-1}Q^+)\,\mathbb{E}|X|^{\vartheta-1} + c_\vartheta\,\mathbb{E}(|M|(Q^+)^{\vartheta-1})\,\mathbb{E}|X|$$

and the finiteness of $\mathbb{E}(|M|^{\vartheta-1}Q^+)$, $\mathbb{E}(|M|(Q^+)^{\vartheta-1})$ follows by an appeal to Hölder's inequality.

In order to get

$$\mathbb{E}\left(((MX+Q)^+)^\vartheta - ((MX)^+)^\vartheta\right)^- < \infty,$$

one can argue in a similar manner when using the estimate

$$\begin{aligned}
&((MX)^+)^\vartheta - ((MX+Q)^+)^\vartheta \;\le\; ((MX+Q)^+ + Q^-)^\vartheta - ((MX+Q)^+)^\vartheta\\
&\le \begin{cases}(Q^-)^\vartheta, & \text{if } \vartheta\in(0,1],\\ (Q^+)^\vartheta + c_\vartheta\big(((MX+Q)^+)^{\vartheta-1}Q^- + (MX+Q)^+(Q^-)^{\vartheta-1}\big), & \text{if } \vartheta>1.\end{cases}
\end{aligned}$$

The straightforward details are again left as an exercise [☞ Problem 4.34].

Turning to the proof of (4.39), suppose that $Q = c(1-M)$ a.s. for some $c\in\mathbb{R}$. Then $X = c$ a.s. forms the unique solution to (4.36) so that $C_+ = C_- = 0$. For the converse, let $\mathbb{P}(Q = c(1-M)) < 1$ for all $c\in\mathbb{R}$. Since

$$C_+ + C_- \;=\; \lim_{t\to\infty} t^\vartheta\,\mathbb{P}(|X|>t),$$

we must verify that the limit on the right-hand side is positive. To this end, we start by noting that, by our assumption, we can pick $\varepsilon>0$ such that

$$p := \mathbb{P}(|Q - m_0(1-M)| > \varepsilon) > 0.$$

Next, observe that

$$\widehat{X}^*_{n-1} + U_n = \widehat{X}_{n-1} + \Pi_{n-1}m_0 + \Pi_{n-1}(Q_n - m_0(1-M_n)) \;=\; \widehat{X}^*_n$$

for each $n\ge 1$, which implies

$$\sup_{n\ge0}|\widehat{X}^*_n| \;\ge\; \sup_{n\ge1}|\widehat{X}^*_n| \;\ge\; \sup_{n\ge1}|U_n| - \sup_{n\ge0}|\widehat{X}^*_n|$$

and therefore

$$\sup_{n\ge0}|\widehat{X}^*_n| \;\ge\; \frac{1}{2}\sup_{n\ge1}|U_n|.$$

Since $\widehat{X}^*_0 = m_0$, we thus find for any $t>|m_0|$ that

$$\left\{\sup_{n\ge1}|\widehat{X}^*_n| > t\right\} \;=\; \left\{\sup_{n\ge0}|\widehat{X}^*_n| > t\right\} \;\supset\; \left\{\sup_{n\ge1}|U_n| > 2t\right\}$$

and then in combination with Grincevičius' inequality (4.41)

$$\begin{aligned}\mathbb{P}(|X|>t) &\ge \frac{1}{2}\,\mathbb{P}\left(\sup_{n\ge 1}|\widehat{X}_n^*|>t\right)\\ &\ge \frac{1}{2}\,\mathbb{P}\left(\sup_{n\ge 1}|U_n|>2t\right)\\ &\ge \frac{1}{2}\sum_{n\ge 1}\mathbb{P}\left(\tau=n-1,|Q_n-m_0(1-M_n)|>\varepsilon\right)\\ &= \frac{p}{2}\,\mathbb{P}\left(\sup_{n\ge 0}|\Pi_n|>\frac{2t}{\varepsilon}\right),\end{aligned}$$

where $\tau := \inf\{n \ge 0 : |\Pi_n| > 2t/\varepsilon\}$ and

$$|U_n| = |\Pi_{n-1}(Q_n - m_0(1-M_n))| > 2t \quad \text{on } \{\tau = n-1, |Q_n - m_0(1-M_n)| > \varepsilon\}$$

for each $n \ge 1$ has been utilized for the penultimate line. Finally,

$$\lim_{t\to\infty} t^{\vartheta}\,\mathbb{P}\left(\sup_{n\ge 0}|\Pi_n|>t\right) = \lim_{t\to\infty} e^{\vartheta \log t}\,\mathbb{P}\left(\sup_{n\ge 0} S_n > \log t\right) = K_+ > 0,$$

by Theorem 2.68 leads to the desired conclusion. □

Formula (4.38) may be used to derive further information on $C_\pm$ or $C_+ + C_-$ like upper bounds or alternative formulae. We refrain from dwelling on this further and refer to Problems 4.36-4.38.

It is usually impossible to determine the law of a perpetuity explicitly, but there are exceptions. One such class is described in Proposition 4.23 below. Recall that a *beta distribution* with parameters $a, b > 0$ has $\lambda\!\!\lambda$-density

$$g_{a,b}(x) := \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}\mathbf{1}_{(0,1)}(x),$$

where the normalizing constant

$$B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

equals the so-called *complete beta integral* at (a,b). The substitution $\frac{y}{1+y}$ for x provides us with the equivalent formula

$$B(a,b) = \int_0^\infty y^{a-1}(1+y)^{-a-b}\,dx$$

for all $a, b \in \mathbb{R}_>$. As a consequence,

$$g^*_{a,b}(x) := \frac{1}{B(a,b)}\, x^{a-1}(1+x)^{-a-b}\,\mathbf{1}_{\mathbb{R}_>}(x)$$

for $a,b>0$ defines the $\lambda\!\!\lambda$-density of another distribution $\beta^*(a,b)$, say, called *beta distribution of the second kind*. Here is a useful (multiplicative) convolution property of these distributions that is crucial for the proof of the announced proposition.

Lemma 4.22. *If X and Y are two independent random variables with $\mathscr{L}(X)=\beta^*(a,b)$ and $\mathscr{L}(Y)=\beta^*(c,a+b)$ for $a,b,c\in\mathbb{R}_>$, then*

$$\mathscr{L}((1+X)Y)=\beta^*(c,b).$$

Proof. For $s\in(-c,b)$, we obtain

$$\begin{aligned}\mathbb{E}Y^s &= \frac{\Gamma(a+b+c)}{\Gamma(c)\Gamma(a+b)}\int_0^\infty y^{c+s-1}(1+y)^{-((c+s)+(a+b-s))}\,dy\\ &= \frac{\Gamma(c+s)\Gamma(a+b-s)}{\Gamma(c)\Gamma(a+b)}\end{aligned}$$

and further in a similar manner

$$\begin{aligned}\mathbb{E}(1+X)^s &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^\infty x^{a-1}(1+x)^{-(a+b-s)}\,dx\\ &= \frac{\Gamma(a+b)\Gamma(b-s)}{\Gamma(b)\Gamma(a+b-s)}.\end{aligned}$$

Consequently, the independence of X,Y implies

$$\mathbb{E}((1+X)Y)^s = \mathbb{E}(1+X)^s\,\mathbb{E}Y^s = \frac{\Gamma(c+s)\Gamma(b-s)}{\Gamma(c)\Gamma(b)}.$$

But the last expression also equals $\phi(s)=\mathbb{E}Z^s$ if $\mathscr{L}(Z)=\beta^*(c,b)$. The function ϕ is called the *Mellin transform* of Z and is the same as the mgf of $\log Z$ (as Z is positive). But the mgf, if not only defined at 0, determines the distribution of $\log Z$ and thus of Z uniquely, giving $\beta^*(c,b)=\mathscr{L}(Z)=\mathscr{L}((1+X)Y)$ as claimed. □

Here is the announced result, again taken from [58]. The cases $m=1,2$ are due to CHAMAYOU & LETAC [25].

Proposition 4.23. *For $m\in\mathbb{N}$ and positive reals $a_1,\dots,a_m,b$, let $X,Y_1,\dots,Y_m$ be independent random variables such that $\mathscr{L}(X)=\beta^*(a_1,b)$ and $\mathscr{L}(Y_k)=\beta^*(a_{k+1},a_k+b)$ for $k=1,\dots,m$, where $a_{m+1}:=a_1$. Then X satisfies the SFPE* (4.36)*, i.e. $X\stackrel{d}{=}MX+Q$, for the pair (M,Q) defined by*

$$M := \prod_{k=0}^{m-1} Y_{m-k} \quad\text{and}\quad Q := \sum_{k=0}^{m-1}\prod_{j=0}^{k} Y_{m-j}.$$

Furthermore,

$$\lim_{t\to\infty} t^b\,\mathbb{P}(X>t) \;=\; \frac{1}{b\,B(a_1,b)}. \tag{4.43}$$

Proof. We put $X_1 := X$ and $X_n := (1+X_{n-1})Y_{n-1}$ for $n=2,\ldots,m$. A simple induction in combination with the previous lemma shows that $\mathscr{L}(X_n) = \beta^*(a_n,b)$ for $n = 1,\ldots,m$. Moreover,

$$(1+X_n)Y_n \;=\; Y_n + X_nY_n \;=\; (1+X_{n-1})Y_{n-1}Y_n \;=\ldots=\; Q+MX$$

and $\mathscr{L}((1+X_n)Y_n) = \beta^*(a_{m+1},b) = \beta^*(a_1,b)$ by another appeal to Lemma 4.22.

Left with the proof of (4.43) and using $\mathscr{L}(X) = \beta^*(a_1,b)$, we infer

$$\begin{aligned}\mathbb{P}(X>t) \;&=\; \frac{1}{B(a_1,b)}\int_t^\infty x^{a_1-1}(1+x)^{-a_1-b}\,dx\\ &=\; \frac{1}{B(a_1,b)}\int_t^\infty \left(\frac{x}{1+x}\right)^{a_1-1}\left(\frac{x}{1+x}\right)^{b+1}\,dx.\end{aligned}$$

Since the last integral is easily seen to behave like $\int_t^\infty x^{-b-1}\,dx = bt^{-b}$ as $t\to\infty$, we arrive at the desired conclusion. ☐

We leave it as an exercise [☞ Problem 4.39] to verify that (M,Q) satisfies the assumptions of the implicit renewal theorem 4.1 with $\vartheta = b$ and so, by this result,

$$\frac{1}{b\,B(a_1,b)} \;=\; \frac{\mathbb{E}\big((MX+Q)^b-(MX)^b\big)}{b\mu_b}$$

holds true.

4.4.2 Lindley's equation and a related max-type equation

If we replace addition in (4.36) by the max-operation, we get a new SFPE, namely

$$X \;\stackrel{d}{=}\; MX\vee Q, \tag{4.44}$$

where X and (M,Q) are as usual independent. We make the additional assumption that $M\ge 0$ a.s. If $\mathbb{E}\log M<0$ and $\mathbb{E}\log^+ Q^+<\infty$, it has a unique solution which is the unique stationary distribution of the mean contractive IFS of iid Lipschitz maps with generic copy $\Psi(x) := MX\vee Q$ [☞ Problem 4.40] and the law of

$$X := \sup_{n\ge 1}\Pi_{n-1}Q_n.$$

Notice that $\Pi_n \to 0$ a.s. in combination with the stationarity of the Q_n entails $\Pi_{n-1}Q_n \to 0$ in probability and thus $X \geq 0$ a.s. In other words, only the right tail of X needs to be studied hereafter. Theorem 4.24 below constitutes the exact counterpart of Theorem 4.19 for (4.44), but before stating it we want to point out the direct relation of this SFPE with Lindley's equation, which is revealed after a transformation. Namely, if $Q \equiv 1$, then taking logarithms in (4.44) yields

$$Y \stackrel{d}{=} (Y + \xi) \vee 0 \;=\; (Y + \xi)^+$$

for $Y = \log X$, where $\xi := \log M$, and the unique solution is given by the law of

$$\log X = \sup_{n \geq 0} \log \Pi_n \;=\; \sup_{n \geq 0} S_n,$$

a fact already known from Problem 1.6.

Theorem 4.24. *Suppose M satisfies (IRT-1)-(IRT-3) and $\mathbb{E}(Q^+)^\vartheta < \infty$. Then there exists a unique solution to the SFPE* (4.44)*, given by the law of $X = \sup_{n \geq 1} \Pi_{n-1}Q_n$. This law satisfies* (4.5) *with*

$$C_+ = \frac{\mathbb{E}\big(((MX \vee Q)^+)^\vartheta - ((MX)^+)^\vartheta\big)}{\vartheta \mu_\vartheta}. \tag{4.45}$$

Moreover, C_+ is positive iff $\mathbb{P}(Q > 0) > 0$.

Proof. Problem 4.40 shows that (the law of) $X = \sup_{n \geq 1} \Pi_{n-1}Q_n$ provides the unique solution to the SFPE (4.44) under the assumptions stated here. (4.5) with C_+ given by (4.45) is now directly inferred from Corollary 4.2 because

$$\begin{aligned}
&\mathbb{E}\big|((MX \vee Q)^+)^\vartheta - ((MX)^+)^\vartheta\big| \\
&\qquad = \mathbb{E}\big|Q^\vartheta - ((MX)^+)^\vartheta\big| \mathbf{1}_{\{MX < Q, Q > 0\}} \\
&\qquad \leq \mathbb{E}(Q^+)^\vartheta \;<\; \infty
\end{aligned}$$

[which is (4.10) in that corollary] holds true.

Turning to the asserted equivalence, one implication is trivial, for $Q \leq 0$ a.s. entails $X = 0$ a.s. and thus $C_+ = 0$. Hence, suppose $\mathbb{P}(Q > 0) > 0$ and fix any $c > 0$ such that $\mathbb{P}(Q > c) > 0$. Defining the predictable first passage time

$$\tau(t) \;:=\; \inf\{n \geq 1 : \Pi_{n-1} > t/c\}, \quad t \geq 0,$$

we note that

$$\mathbb{P}\left(\sup_{n \geq 1} \Pi_{n-1} > \frac{t}{c}\right) \;=\; \mathbb{P}(\tau(t) < \infty) \tag{4.46}$$

and claim that

$$\mathbb{P}\left(\sup_{n\geq 1}\Pi_{n-1}Q_n>t\right) \;\geq\; \mathbb{P}(Q>c)\,\mathbb{P}(\tau(t)<\infty). \tag{4.47}$$

For a proof of the latter claim, just note that

$$\begin{aligned}\mathbb{P}\left(\sup_{n\geq 1}\Pi_{n-1}Q_n>t\right) \;&\geq\; \sum_{n\geq 1}\mathbb{P}\left(\tau(t)=n,\,\Pi_{n-1}Q_n>t\right)\\ &\geq\; \sum_{n\geq 1}\mathbb{P}\left(\tau(t)=n,\,Q_n>c\right)\\ &=\; \mathbb{P}(Q>c)\,\mathbb{P}(\tau(t)<\infty),\end{aligned}$$

where the last line follows by the independence of $\{\tau(t)=n\}\in\sigma(\Pi_0,...,\Pi_{n-1})$ and Q_n. Now we infer upon using (4.46) and (4.47) that

$$\mathbb{P}(X>t) \;=\; \mathbb{P}\left(\sup_{n\geq 1}\Pi_{n-1}Q_n>t\right) \;\geq\; \mathbb{P}(Q>c)\,\mathbb{P}\left(\sup_{n\geq 1}\Pi_{n-1}>\frac{t}{c}\right)$$

and thereby the desired result $C_+>0$, for $\mathbb{P}(Q>c)>0$ and

$$\lim_{t\to\infty}\left(\frac{t}{c}\right)^{\vartheta}\mathbb{P}\left(\sup_{n\geq 1}\Pi_{n-1}>\frac{t}{c}\right) \;=\; \lim_{t\to\infty}e^{\vartheta t}\,\mathbb{P}\left(\sup_{n\geq 0}S_n>t\right) \;>\; 0$$

by invoking once again Theorem 2.68. □

4.4.3 Letac's max-type equation $X\overset{d}{=}M(N\vee X)+Q$

A more general example of a max-type SFPE studied by GOLDIE in [58] was first introduced by LETAC [73, Example E], namely

$$X \;\overset{d}{=}\; M(N\vee X)+Q \tag{4.48}$$

for a random triple (M,N,Q) independent of X such that $M\geq 0$ a.s. As usual, this equation characterizes the unique stationary law of the pertinent IFS of iid Lipschitz maps, defined by

$$X_n \;=\; M_n(N_n\vee X_{n-1})+Q_n,\quad n\geq 1,$$

provided that mean contractivity and the jump-size condition (3.17) hold. The (M_n,N_n,Q_n), $n\geq 1$, are of course independent copies of (M,N,Q). We leave it as an exercise [☞ Problem 4.42] to verify that mean contraction holds if $\mathbb{E}\log M<0$ and (3.17) holds if, furthermore, $\mathbb{E}\log^+N^+<\infty$ $\mathbb{E}\log^+|Q|<\infty$. By computing the backward iterations, one then finds as in [58, Prop. 6.1] that

$$X \;:=\; \max\left\{\sum_{n\geq 1}\Pi_{n-1}Q_n,\sup_{n\geq 1}\left(\sum_{k=1}^{n}\Pi_{k-1}Q_k+\Pi_nN_n\right)\right\}$$

is a.s. finite and its law the unique solution to (4.48). But $\Pi_n N_n \to 0$ in probability in combination with

$$X \;\ge\; \lim_{n\to\infty}\left(\sum_{k=1}^{n}\Pi_{k-1}Q_k+\Pi_n N_n\right) \;=\; \sum_{n\ge 1}\Pi_{n-1}Q_n$$

obviously implies that

$$X \;=\; \sup_{n\ge 1}\left(\sum_{k=1}^{n}\Pi_{k-1}Q_k+\Pi_n N_n\right). \tag{4.49}$$

The following result determines the right tail behavior of X with the help the implicit renewal theorem. As for the left tail behavior see Remark 4.27 below.

Theorem 4.25. *Suppose that M satisfies (IRT-1)-(IRT-3) and*

$$\mathbb{E}(N^+)^{\vartheta}<\infty,\quad \mathbb{E}|Q|^{\vartheta}<\infty \quad \textit{and} \quad \mathbb{E}(MN^+)^{\vartheta}<\infty. \tag{4.50}$$

Then the SFPE (4.48) *has a unique solution given by the law of X in* (4.49)*. This law satisfies* (4.5) *with*

$$C_+=\frac{\mathbb{E}\big(((M(N\vee X)+Q)^+)^{\vartheta}-((MX)^+)^{\vartheta}\big)}{\vartheta\mu_{\vartheta}}. \tag{4.51}$$

Furthermore, C_+ is positive iff $\mathbb{P}(Q=c(1-M))<1$ for all $c\in\mathbb{R}$, or $Q=c(1-M)$ a.s. and $\mathbb{P}(M(N-c)>0)>0$ for some $c\in\mathbb{R}$.

Remark 4.26. In [58, Theorem 6.2] only a sufficient condition for $C_+>0$ was given, namely that $Q-c(1-M)\ge 0$ a.s. and

$$\mathbb{P}(Q-c(1-M)>0)+\mathbb{P}(M(N-c)>0)>0 \tag{4.52}$$

for some constant $c\in\mathbb{R}$.

Remark 4.27. Concerning the left tail of X in Theorem 4.25, let us point out the following: Since $M(N\vee X)+Q\ge MN+Q\mathbf{1}_{\{Q<0\}}$, we have

$$\mathbb{E}(X^-)^{\vartheta} \;=\; \mathbb{E}((M(N\vee X)^-)^{\vartheta} \;\le\; \mathbb{E}(MN^-)^{\vartheta}+\mathbb{E}(Q^-)^{\vartheta}$$

and thus $C_-=\lim_{t\to\infty}t^{\vartheta}\,\mathbb{P}(X<-t)=0$ if the last moment assumption in (4.50) is sharpened to $\mathbb{E}M|N|^{\vartheta}<\infty$.

Proof (of Theorem 4.25). By what has been stated before the theorem and is shown in Problem 4.42(a), (4.50) ensures that the IFS pertaining to the SFPE (4.48) is mean contractive and satisfies (3.17). Therefore the law of X, defined in (4.49), forms the unique solution to (4.48). In order to infer (4.5) for its right tails by Corollary 4.2, we must verify

$$\mathbb{E}\big|((M(N\vee X)+Q)^+)^\vartheta-((MX)^+)^\vartheta\big|<\infty$$

or, a fortiori,

$$\mathbb{E}\big|((M(N\vee X)+Q)^+)^\vartheta-((M(N\vee X)^+)^\vartheta\big|<\infty$$
$$\text{and}\quad \mathbb{E}\big|(M(N\vee X)^+)^\vartheta-((MX)^+)^\vartheta\big|<\infty.$$

If $\vartheta\in(0,1]$, the desired conclusion is obtained by the usual subadditivity argument [☞ proof of Theorem 4.19], namely

$$\mathbb{E}\big|((M(N\vee X)+Q)^+)^\vartheta-((MX)^+)^\vartheta\big|\ \le\ \mathbb{E}|Q|^\vartheta\ <\ \infty.$$

Left with the case $\vartheta>1$, we first point out that the sharpened jump-size condition (3.20) (with $p=\vartheta$, $x_0=0$ and $d(x,y)=|x-y|$) holds, namely

$$\mathbb{E}\big(MN^++Q^+\big)^\vartheta<\infty.$$

This is an obvious consequence of (4.50). Therefore $\mathbb{E}|X|^p<\infty$ for any $p\in(0,\vartheta)$ by Theorem 3.29. By another use of inquality (4.42), we find that

$$\begin{aligned}
&((M(N\vee X)+Q)^+)^\vartheta-((M(N\vee X)^+)^\vartheta\\
&\quad\le\ (Q^+)^\vartheta+\vartheta\,2^{\vartheta-1}\Big((M(N\vee X)^+)^{\vartheta-1}Q^++M(N\vee X)^+(Q^+)^{\vartheta-1}\Big)\\
&\quad\le\ (Q^+)^\vartheta+\vartheta\,2^{\vartheta-1}\Big((MN^+\vee X)^+)^{\vartheta-1}Q^++M(N\vee X)^+(Q^+)^{\vartheta-1}\Big)\\
&\quad\le\ (Q^+)^\vartheta+\vartheta\,2^{\vartheta-1}\Big(((MN^+)^{\vartheta-1}+(X^+)^{\vartheta-1})Q^++((MN^+)^{\vartheta-1}\\
&\qquad\qquad\qquad\qquad\qquad\qquad+(MX^+)^{\vartheta-1})(Q^+)^{\vartheta-1}\Big).
\end{aligned}$$

But the last expression has finite expectation as one can see by using $\mathbb{E}|X|^{\vartheta-1}<\infty$, our moment assumptions (4.50) and the independence of X and (M,N,Q). We refer to the proof of Theorem 4.19 for a very similar argument spelled out in greater detail. Having thus shown

$$\mathbb{E}\left(((M(N\vee X)+Q)^+)^\vartheta-((M(N\vee X)^+)^\vartheta\right)^+<\infty$$

we leave it to the reader to show by similar arguments that the corresponding negative part has finite expectation, too, and that $\mathbb{E}\big|(M(N\vee X)^+)^\vartheta-((MX)^+)^\vartheta\big|<\infty$ [☞ Problem 4.43].

Turning to the equivalence assertion, recall from above that

$$X\ \ge\ Y:=\sum_{n\ge1}\Pi_{n-1}Q_n$$

so that $\mathbb{P}(X>t) \ge \mathbb{P}(Y>t)$. But Theorem 4.19 tells us that, under the given assumptions, $t^{\vartheta}\,\mathbb{P}(Y>t)$ converges to a positive limit if $\mathbb{P}(Q=c(1-M))<1$ for all $c\in\mathbb{R}$, whence the same must then hold for $t^{\vartheta}\,\mathbb{P}(X>t)$. On the other hand, if $Q=c(1-M)$ a.s. for some c, then a simple calculation shows that

$$X \;=\; \sup_{n\ge 1}\left(c(1-\Pi_n)+\Pi_n N_n\right) \;=\; c+\sup_{n\ge 1}\left(\Pi_{n-1}M_n(N_n-c)\right).$$

By Theorem 4.24, $X-c$ then forms the unique solution to the SFPE (4.44) when choosing $Q=M(N-c)$ there. Consequently,

$$C_+ \;=\; \lim_{t\to\infty} t^{\vartheta}\,\mathbb{P}(X>t) \;=\; \lim_{t\to\infty} t^{\vartheta}\,\mathbb{P}(X-c>t) \;>\; 0$$

iff $\mathbb{P}(M(N-c)>0)>0$ as claimed. □

Writing (4.48) in the form $X \stackrel{d}{=} (MN)\vee MX+Q$, we see that an SFPE of type (4.44), which has been discussed in the previous subsection, yields as a special case when choosing $Q=0$. However, as $\mathbb{P}(MN=0)\ge\mathbb{P}(M=0)$, the statement fails to hold for those equations $X \stackrel{d}{=} MX\vee Q'$ with $\mathbb{P}(Q'=0)<\mathbb{P}(M=0)$ [☞ Problem 4.44 for further information].

Further specializing to the situation when $Q=0$ and $N>0$ a.s., Letac's equation (4.48) after taking logarithms turns into

$$Y \stackrel{d}{=} \zeta\vee Y+\xi, \tag{4.53}$$

where $Y:=\log X$, $\xi:=\log M$ and $\zeta:=\log N$. Upon choosing the usual notation for the associated IFS $(Y_n)_{n\ge 0}$, say, of iid Lipschitz maps with generic copy $\Psi(x):=\zeta\vee x+\xi$, backward iterations can be shown to satisfy [☞ Problem 4.45(a)]

$$\widehat{Y}_n \;=\; \max\left\{S_n+x,\ \max_{1\le k\le n}(S_k+\zeta_k)\right\} \tag{4.54}$$

if $\widehat{Y}_0=x$ and $(S_n)_{n\ge 0}$ denotes the SRW associated with the $\xi_1,\xi_2,\ldots$ Provided that ξ has negative mean and thus $(S_n)_{n\ge 0}$ negative drift, we infer a.s. convergence of the $\widehat{Y}_n$ to

$$Y \;:=\; \sup_{n\ge 1}(S_n+\zeta_n)$$

the law of which then constitutes the unique solution to (4.53). However, $\mathbb{P}(Y=\infty)$ may be positive. In order to rule out this possibility, it is sufficient to additionally assume $\mathbb{E}\log^+\zeta<\infty$ [☞ Problem 4.45(c)].

HELLAND & NILSEN [64] have studied a random recursive equation leading to a special case of (4.53), namely

$$Y_n \;=\; (Y_{n-1}-D_n)\vee U_n \;=\; \big(Y_{n-1}\vee(U_n+D_n)\big)-D_n,\quad n\ge 1,$$

for independent sequences $(D_n)_{n\geq 1}$ and $(U_n)_{n\geq 1}$ of iid random variables which are also independent of Y_0. The model had been suggested earlier by GADE [57] (with constant D_n) and HELLAND [63] in an attempt to describe the deep water exchanges in a sill fjord, i.e., an inlet containing a relatively deep basin with a shallower sill at the mouth. The water exchanges are described by the following simple mechanism: If, in year n, U_n denotes the density of coastal water adjacent to the fjord and Y_n the density of resident water in the basin, then fresh water running into the fjord causes the resident water density to decrease by an amount D_n from year $n-1$ to n. Nothing happens if this water is still heavier than the coastal water, but resident water is completely replaced with water of density U_n otherwise [☞ [57] and [64] for further information]. Obviously, the distributional limit of Y_n, if it exists, satisfies (4.53) with $\zeta := U+D$ and $\xi := -D$, where as usual (D,U) denotes a generic copy of the (D_n,U_n) independent of Y.

4.4.4 The AR(1)-model with ARCH(1) errors

We return to the nonlinear time series model first introduced in Section 1.6 and briefly discussed further in Example 3.9, namely the AR(1)-model with ARCH(1) errors

$$X_n \ = \ \alpha X_{n-1} + (\beta + \lambda X_{n-1}^2)^{1/2}\varepsilon_n, \quad n \geq 1, \tag{4.55}$$

where $\varepsilon, \varepsilon_1, \varepsilon_2, ...$, called innovations, are iid symmetric random variables independent of X_0 and $(\alpha, \beta, \lambda) \in \mathbb{R}\times\mathbb{R}^2_>$. This is an IFS of iid Lipschitz maps of generic form $\Psi(x) := \alpha x + (\beta + \lambda x^2)^{1/2}\varepsilon$. As pointed out in 3.9, Ψ has Lipschitz constant $L(\Psi) = |\alpha| + \lambda^{1/2}|\varepsilon|$. The following result is therefore immediate when using Theorem 3.24.

Proposition 4.28. *The IFS $(X_n)_{n\geq 0}$ stated above is mean contractive and satisfies the jump-size condition* (3.17) *if* $\mathbb{E}\log(|\alpha| + \lambda^{1/2}|\varepsilon|) < 0$. *In this case it possesses a unique stationary distribution, which is symmetric and the unique solution to the SFPE*

$$X \ \stackrel{d}{=} \ \alpha X + (\beta + \lambda X^2)^{1/2}\varepsilon, \tag{4.56}$$

where X, ε are independent.

Proof. We only mention that, if (4.56) holds, then

$$-X \ \stackrel{d}{=} \ \alpha(-X) + (\beta + \lambda X^2)^{1/2}(-\varepsilon) \ \stackrel{d}{=} \ \alpha(-X) + (\beta + \lambda(-X)^2)^{1/2}\varepsilon,$$

the second equality by the symmetry of ε. Hence the law of $-X$ also solves (4.56) implying $\mathscr{L}(X) = \mathscr{L}(-X)$ because this SFPE has only one solution. □

Remark 4.29. [☞ also [21, Remark 2]] Let us point out that, if X_n is given by (4.55), then $X_n^\star := (-1)^n X_n$ satisfies the same type of random recursive equation, viz.

$$X_n^\star \;=\; -\alpha X_{n-1}^\star + (\beta + \lambda (X_{n-1}^\star)^2)^{1/2} \varepsilon_n^\star$$

with $\varepsilon_n^\star := (-1)^n \varepsilon_n$ for $n \geq 1$. But the $\varepsilon_n^\star$ are again independent copies of ε_1, for ε_1 is symmetric. In terms of distributions, the IFS $(X_n)_{n\geq 0}$ and $(X_n^\star)_{n\geq 0}$ thus differ merely by a sign change for the parameter α, and under the assumptions of the previous result we further have that $X_n^\star$ converges in distribution to the same limit X. It is therefore no loss of generality to assume $\alpha > 0$.

The symmetry of the law of X, giving

$$\mathbb{P}(X > t) \;=\; \mathbb{P}(X < -t) \;=\; \frac{1}{2}\mathbb{P}(|X| > t) \;=\; \frac{1}{2}\mathbb{P}(X^2 > t^2)$$

for all $t \in \mathbb{R}_\geq$, allows us to subsequently focus on $Y = X^2$, which satisfies the distributional equation

$$Y \;\stackrel{d}{=}\; (\alpha^2 + \lambda \varepsilon^2) Y + 2\alpha\varepsilon X (\beta + \lambda Y)^{1/2} + \beta \varepsilon^2.$$

This is *not* a SFPE as X is not a function of Y, but when observing that $\varepsilon X \stackrel{d}{=} \eta |X|$, where η is a copy of ε independent of $|X| = Y^{1/2}$ and satisfying $\eta^2 = \varepsilon^2$, we are led to

$$\begin{aligned} Y \;&\stackrel{d}{=}\; (\alpha + \lambda^{1/2}\eta)^2 Y + 2\alpha\eta Y^{1/2}\Big((\beta + \lambda Y)^{1/2} - (\lambda Y)^{1/2}\Big) + \beta\eta^2 \\ &=\; (\alpha + \lambda^{1/2}\eta)^2 Y + \frac{2\alpha\beta\eta Y^{1/2}}{(\beta + \lambda Y)^{1/2} + (\lambda Y)^{1/2}} + \beta\eta^2 \;=:\; \Phi(Y). \end{aligned} \tag{4.57}$$

with Y and η being independent. Observe that

$$\Phi_*(y) \;:=\; My - c|\eta| + \beta\eta^2 \;\leq\; \Phi(y) \;\leq\; My + c|\eta| + \beta\eta^2 \;=:\; \Phi^*(y) \tag{4.58}$$

for all $y \in \mathbb{R}_\geq$, where $M := (\alpha + \lambda^{1/2}\eta)^2$ and $c := \alpha\beta\lambda^{-1/2}$. As a consequence, we obtain the following lemma which will be useful to prove our main result, Theorem 4.31 below.

Lemma 4.30. *Suppose that $\mathbb{E}\log M < 0$ and let $(\Phi_{*,n}, \Phi_n^*, \Phi_n)_{n\geq 1}$ be a sequence of iid copies of (Φ_*, Φ^*, Φ), defined on the same probability space as Y. Then the following assertions hold true:*

(a) $\Phi_{*,1:n} \leq \Phi_{1:n} \leq \Phi_{1:n}^*$ *for all* $n \geq 1$.
(b) $\Phi_{*,1:n}(Y) \to Y_*$ *and* $\Phi_{1:n}^*(Y) \to Y^*$ *a.s. for random variables* $Y_* \leq Y^*$ *satisfying*

$$Y_* \stackrel{d}{=} \Phi_*(Y_*) \quad \text{and} \quad Y^* \stackrel{d}{=} \Phi^*(Y^*). \tag{4.59}$$

(c) $Y_* \leq_{st} Y \leq_{st} Y^*$, *i.e.*

$$\mathbb{P}(Y_* > t) \ \leq \ \mathbb{P}(Y > t) \ \leq \ \mathbb{P}(Y^* > t)$$

for all $t \in \mathbb{R}$.

Proof. Part (a) follows directly from (4.58). Since $\mathbb{E}\log M < 0$ implies $\mathbb{E}\log^+|\eta| < \infty$, we see that the IFS generated by the $(\Phi_{*,n})_{n\geq 1}$ and $(\Phi_n^*)_{n\geq 1}$ are mean contractive and satisfying (3.17). Therefore, their backward iterations are a.s. convergent under any initial condition to limiting variables solving the SFPE's stated in (4.59). This proves (b). Finally, as Y satisfies the SFPE (4.57), we infer

$$\Phi_{1:n}(Y) \stackrel{d}{=} Y$$

and thus $\Phi_{*,1:n}(Y) \leq_{st} Y \leq_{st} \Phi_{1:n}^*(Y)$ for all $n \geq 1$. Taking the limit $n \to \infty$ yields the assertion. □

The quintessential outcome of the previous lemma is that Y can be sandwiched in the sense of stochastic majorization $(\leq_{st})$ by two perpetuities, Y_* and Y^*. The following result is now derived very easily with the help of the implicit renewal theorem.

Theorem 4.31. *Suppose that* $M = (\alpha + \lambda^{1/2}\eta)^2$ *satisfies (IRT-1)-(IRT-3) and let* Y *be a nonnegative solution to the SFPE* (4.57)*. Then the law of* Y *satisfies* (4.5) *with*

$$C_+ = \frac{\mathbb{E}\big(\Phi(Y)^\vartheta - (MY)^\vartheta\big)}{\vartheta\mu_\vartheta} \tag{4.60}$$

which is positive if $\vartheta \geq 2$.

Remark 4.32. In all previous applications, the random variable M that appeared in the respective tail result happened to be also the Lipschitz constant of the generic Lipschitz function Ψ in the SFPE under consideration, i.e. $M = L(\Psi)$. In the present situation, however, this is no longer true. We have $L(\Psi) = |\alpha| + \lambda^{1/2}|\eta|$ which, after squaring, would suggest $M' = (|\alpha| + \lambda^{1/2}|\eta|)^2$ in the above theorem. But $M' > M = (\alpha + \lambda^{1/2}\eta)^2$ a.s. even if α is positive. What this essentially tells us is that mean contraction with respect to global Lipschitz constants, albeit constituting a sufficient condition for the distributional convergence of a given IFS of iid Lipschitz maps to a unique limit law, fails to be necessary in general. For the AR(1)-model with ARCH(1) errors, BORKOVEC & KLÜPPELBERG [21, Theorem 1] show that $\mathbb{E}\log M < 0$ in combination with some additional conditions on $\mathscr{L}(\eta)$ (beyond symmetry) already ensures convergence to a unique symmetric stationary distribution. Earlier results in this direction under the second moment condition

$\alpha^2 + \lambda\,\mathbb{E}\eta^2 = \mathbb{E}M^2 < 1$ were obtained by GUÉGAN & DIEBOLT [61] and MAERCKER [79].

Proof. Note that (IRT-1) for $(\alpha + \lambda^{1/2}\eta)^2$ ensures $\mathbb{E}|\eta|^{2\vartheta} < \infty$. Under the assumptions of the theorem und with the notation of the previous lemma, the IFS generated by $(\Phi_{*,n})_{n\ge 1}$ and $(\Phi^*)_{n\ge 1}$ are strongly contractive of order ϑ and satisfy the sharpened jump-size condition (3.20) for $p = \vartheta$. Hence, the perpetuities Y_* and Y^*, have moments of all orders $p \in (0,\vartheta)$ [☞ Prop. 4.20]. Using Lemma 4.30(c), $\mathbb{E}|Y|^p \le \mathbb{E}|Y_*|^p + \mathbb{E}|Y^*|^p < \infty$ for all $p \in (0,\vartheta)$. Now it follows in a meanwhile routine manner that

$$\mathbb{E}\left|\Phi(Y)^{\vartheta} - (MY)^{\vartheta}\right| \;\le\; \mathbb{E}\left|c|\eta| + \beta\eta^2\right|^{\vartheta} \;<\; \infty$$

if $\vartheta \le 1$, and [use again (4.42) and put $c_\vartheta := \vartheta 2^{\vartheta-1}$]

$$\begin{aligned}\mathbb{E}\left|\Phi(Y)^{\vartheta} - (MY)^{\vartheta}\right| \;&\le\; \mathbb{E}\left|\Phi^*(Y)^{\vartheta} - (MY)^{\vartheta}\right|\\ &\le\; \mathbb{E}\left|c|\eta| + \beta\eta^2\right|^{\vartheta}\\ &\quad + c_\vartheta\left(\mathbb{E}Y^{\vartheta-1}\,\mathbb{E}\left|c|\eta| + \beta\eta^2\right| + \mathbb{E}Y\,\mathbb{E}\left|c|\eta| + \beta\eta^2\right|^{\vartheta-1}\right) \;<\; \infty\end{aligned}$$

if $\vartheta > 1$. Hence, by Corollary 4.2, the right tails of Y satisfy (4.5) with C_+ as stated.

Left with the proof of $C_+ > 0$ if $\vartheta \ge 2$, a Taylor expansion of $\Phi(Y)^{\vartheta}$ about MY yields

$$\begin{aligned}\Phi(Y)^{\vartheta} \;&=\; (MY)^{\vartheta} + \vartheta(MY)^{\vartheta-1}h(Y,\eta) + \vartheta(\vartheta-1)Z^{\vartheta-2}h(Y,\eta)^2\\ &\ge\; (MY)^{\vartheta} + \vartheta(MY)^{\vartheta-1}h(Y,\eta),\end{aligned}$$

where

$$h(Y,\eta) \;:=\; \frac{2\alpha\beta\eta Y^{1/2}}{(\beta+\lambda Y)^{1/2} + (\lambda Y)^{1/2}} + \beta\eta^2$$

and Z is an intermediate (random) point between MY and $\Phi(Y) = MY + h(Y,\eta)$ and thus ≥ 0. As a consequence,

$$\mathbb{E}\Big(\Phi(Y)^{\vartheta} - (MY)^{\vartheta}\Big) \;\ge\; \vartheta\,\mathbb{E}(MY)^{\vartheta-1}h(Y,\eta) \;=\; \vartheta\beta\,\mathbb{E}\eta^2 \;>\; 0,$$

having utilized $\mathbb{E}\eta = 0$ and the independence of Y and η. □

Problems

Problem 4.33. Prove the following converse of Proposition 4.20: If $\mathbb{E}Y^p < \infty$ for some $p \ge 1$ and $\mathbb{P}(|Q| > 0) > 0$, then $\mathbb{E}|M|^p < 1$ and $\mathbb{E}|Q|^p < \infty$. [Hint: Use that $\sum_{n\ge 1} |\Pi_{n-1}Q_n|^p \le Y^p$.]

Problem 4.34. Complete the proof of Theorem 4.19.

Problem 4.35. Given the assumptions of Theorem 4.19, suppose additionally that $\vartheta > 1$ and $\mathbb{E}|Q|^{\kappa} < \infty$ for some $\kappa \in [1, \vartheta)$. Prove that

$$\mathbb{E}X^n \;=\; \sum_{k=0}^{n} \binom{n}{k} \mathbb{E}\big(M^k Q^{n-k}\big)\, \mathbb{E}X^k \tag{4.61}$$

for all integers $n \le \kappa$. [This was first shown by VERVAAT [111].]

Problem 4.36. Given the assumptions of Theorem 4.19 and $M \ge 0$ a.s., prove that, if $0 < \vartheta \le 1$,

$$C_+ + C_- \;\le\; \frac{1}{\vartheta \mu_\vartheta}\, \mathbb{E}|Q|^{\vartheta},$$

while, if $\vartheta > 1$,

$$\begin{aligned} C_+ + C_- \;&\le\; \frac{2^{\vartheta-1}}{\vartheta\mu_\vartheta} \Big(\mathbb{E}|Q|^{\vartheta} + \mathbb{E}\big(M^{\vartheta-1}|Q|\big)\,\mathbb{E}|X|^{\vartheta-1}\Big) \\ &\le\; \frac{2^{\vartheta-1}}{\vartheta\mu_\vartheta} \left(\mathbb{E}|Q|^{\vartheta} + \frac{\mathbb{E}\big(M^{\vartheta-1}|Q|\big)\,\mathbb{E}|Q|^{\vartheta-1}}{1 - \|M\|_{\vartheta-1})^{\vartheta-1}}\right). \end{aligned}$$

If $\mathbb{P}(M < 0) > 0$ and thus $C_+ = C_-$, the same bounds with an additional factor $1/2$ and M replaced by $|M|$ hold for C_+ and C_-.

Problem 4.37. Given the assumptions of Theorem 4.19, $M, Q \ge 0$ a.s. and $\vartheta \in \mathbb{N}$, prove that $C_- = 0$ and

$$C_+ \;=\; \frac{1}{\vartheta\mu_\vartheta} \sum_{k=0}^{\vartheta-1} \binom{\vartheta}{k} \mathbb{E}\big(M^k Q^{\vartheta-k}\big)\, \mathbb{E}X^k \tag{4.62}$$

with $\mathbb{E}X^k$ being determined by (4.61) for $k = 1, ..., \vartheta - 1$. Show further that

$$C_+ \;=\; \begin{cases} \dfrac{\mathbb{E}Q}{\mathbb{E}M\log M}, & \text{if } \vartheta = 1, \\[2ex] \dfrac{1}{\mu_\vartheta}\left(\dfrac{1}{2}\mathbb{E}Q^2 + \dfrac{\mathbb{E}Q\,\mathbb{E}MQ}{1 - \mathbb{E}M}\right), & \text{if } \vartheta = 2. \end{cases} \tag{4.63}$$

Problem 4.38. Given the assumptions of Theorem 4.19 and $\vartheta \in 2\mathbb{N}$, prove the following assertions:

(a) If $M \ge 0$ a.s., then (4.62) holds for $C_+ + C_-$ instead of C_+.

(b) If $\mathbb{P}(M < 0) > 0$, then

$$C_+ \;=\; C_- \;=\; \frac{1}{2\vartheta\mu_\vartheta} \sum_{k=0}^{\vartheta-1} \binom{\vartheta}{k} \mathbb{E}\big(M^k Q^{\vartheta-k}\big)\, \mathbb{E}X^k$$

with $\mathbb{E}X^k$ being determined by (4.61) for $k = 1, ..., \vartheta - 1$.

(c) If $\vartheta = 2$, then the respective formula in (4.63) holds for $C_+ + C_-$ instead of C_+.

Problem 4.39. Prove that (M,Q) defined in Proposition 4.23 satisfies the conditions of the implicit renewal theorem 4.1 with $\vartheta = b$.

Problem 4.40. Let $(X_n)_{n\ge 0}$ be an IFS generated by iid Lipschitz maps of generic form $\Psi(x) := Mx \vee Q$ and suppose that $\mathbb{E}\log|M| < 0$ and $\mathbb{E}\log^+|Q| < \infty$. Prove that $(X_n)_{n\ge 0}$ has a unique stationary distribution π which forms the unique solution to the SFPE (4.44) and is the distribution of $\sup_{n\ge 1}\Pi_{n-1}Q_n$ (in the usual notation).

Problem 4.41. As a variation of (4.44), consider the SFPE

$$X \stackrel{d}{=} MX \curlyvee Q, \tag{4.64}$$

where as usual X and (M,Q) are independent and

$$x \curlyvee y \ := \ \begin{cases} x, & \text{if } |x| > |y|, \\ y, & \text{otherwise.} \end{cases}$$

Prove the following counterpart of Theorem 4.24:
If M satisfies (IRT-1)-(IRT-3) and $\mathbb{E}|Q|^\vartheta < \infty$, then (4.64) has a unique solution X (in terms of its law) and (4.5), (4.6) hold true. If $M \ge 0$ a.s., then

$$C_\pm = \frac{\mathbb{E}\big(((MX \curlyvee Q)^\pm)^\vartheta - ((MX)^\pm)^\vartheta\big)}{\vartheta\mu_\vartheta},$$

and if $\mathbb{P}(M < 0) > 0$, then

$$C_+ = C_- = \frac{\mathbb{E}\big((|Q|^\vartheta - |MX|^\vartheta)^+\big)}{\vartheta\mu_\vartheta}.$$

Moreover, $C_+ + C_-$ is positive iff $\mathbb{P}(Q \ne 0) > 0$.

Problem 4.42. (Letac's example E in [73]) Consider the IFS, defined by the random recursive equation

$$X_n \ = \ M_n(N_n \vee X_{n-1}) + Q_n, \quad n \ge 1,$$

for iid random triples (M_n,N_n,Q_n), $n \ge 1$, in $\mathbb{R}_\ge \times \mathbb{R}^2$ with generic copy (M,N,Q). Show that

(a) $(X_n)_{n\ge 0}$ is mean contractive if $\mathbb{E}\log M < 0$ and satisfies the jump-size condition (3.17) if, furthermore, $\mathbb{E}\log^+ N^+ < \infty$ and $\mathbb{E}\log^+|Q| < \infty$.

(b) If the previous conditions hold, the a.s. limit of the associated backward iterations $\widehat{X}_n$ is given by

$$X \ := \ \max\left\{\sum_{n\ge 1}\Pi_{n-1}Q_n, \sup_{n\ge 1}\left(\sum_{k=1}^n \Pi_{k-1}Q_k + \Pi_n N_n\right)\right\},$$

where Π_n has the usual meaning. [Hint: Use induction over n to verify that

$$\Psi_{n:1}(t) \;=\; \max\left\{\sum_{k=1}^{n}\Pi_{k-1}Q_k+\Pi_n t,\, \max_{1\le m\ge n}\left(\sum_{k=1}^{m}\Pi_{k-1}Q_k+\Pi_m N_m\right)\right\}$$

for any $\in\mathbb{R}$, where $\Psi_n(t) := M_n(N_n\vee t)+Q_n$ for $n\in\mathbb{N}$.]

Problem 4.43. Complete the proof of Theorem 4.25 by showing along similar lines as in the proof of Theorem 4.19 that, for the case $\vartheta > 1$,

$$\mathbb{E}\left(((M(N\vee X)+Q)^+)^{\vartheta}-((M(N\vee X)^+)^{\vartheta}\right)^- < \infty$$

as well as

$$\mathbb{E}\left|(M(N\vee X)^+)^{\vartheta}-((MX)^+)^{\vartheta}\right| < \infty.$$

Problem 4.44. Prove that, if M,Q are real-valued random variables such that $M\ge 0$ a.s. and $\mathbb{P}(Q=0)\ge\mathbb{P}(M=0)$, then there exist random variables M',N' (on a suitable probability space) such that $(M',M'N')$ forms a copy of (M,Q).

Problem 4.45. Consider an IFS $(Y_n)_{n\ge 0}$ of iid Lipschitz maps with generic copy $\Psi(x)=\zeta\vee x+\xi$ and let $(S_n)_{n\ge 0}$ denote the SRW with increments $\xi_1,\xi_2,\ldots$ (in the usual notation). Prove the following assertions:

(a) The backward iterations $\widehat{Y}_n$, when starting at $\widehat{Y}_0 = x$, are given by (4.54).
(b) If $\mathbb{E}\xi < 0$, then $\widehat{Y}_n \to Y$ a.s., where $Y=\sup_{n\ge 1}(S_n+\zeta_n)$.
(c) If, furthermore, $\mathbb{E}\zeta^+ < \infty$, then Y is a.s. finite.

Chapter 5
The smoothing transform: a stochastic linear recursion with branching
Part I: Contraction properties

Except for the introductory Chapter 1, we have dealt so far only with random recursive equations with *no branching*, viz.

$$X_n = \Psi_n(X_{n-1}) \tag{5.1}$$

for $n \geq 1$ and iid random functions $\Psi_1, \Psi_2, \ldots$ independent of X_0. The branching case to be considered next occurs if the right-hand side of (5.1) involves multiple copies of X_{n-1}, i.e.

$$X_n = \Psi_n(X_{n-1,1} X_{n-1,2}, \ldots), \tag{5.2}$$

for $n \geq 1$, where $(X_{n-1,k})_{k \geq 1}$ is a sequence of iid copies of X_{n-1} and further independent of Ψ_n. The present chapter will focus on the situation when the Ψ_n are random linear functions with generic copy

$$\Psi(x_1, x_2, \ldots) = \sum_{k \geq 1} T_k x_k + C$$

for a sequence of real-valued random variables $(C, T_1, T_2, \ldots)$. This leads to the so-called *smoothing transform(ation)*

$$\mathscr{S}: \quad F \;\mapsto\; \mathscr{L}\left(\sum_{k \geq 1} T_k X_k + C\right) \tag{5.3}$$

which maps a distribution $F \in \mathscr{P}(\mathbb{R})$ to the law of $\sum_{k \geq 1} T_k X_k + C$, where $X_1, X_2, \ldots$ are independent of $(C, T_1, T_2, \ldots)$ with common distribution F. On the event where

$$N := \sum_{k \geq 1} \mathbf{1}_{\{T_k \neq 0\}}$$

is infinite, we understand $\sum_{k \geq 1} T_k X_k$ as the almost sure limit of the finite partial sums $\sum_{k=1}^{n} T_k X_k$. Then it can be shown [☞ Problem 5.42] that $\mathscr{S}(F)$ is indeed defined *for all* $F \in \mathscr{P}(\mathbb{R})$ if

$$\mathbb{P}(N < \infty) = 1. \tag{A0}$$

is a.s. finite, but exists only for F from a subset of $\mathscr{P}(\mathbb{R})$ (always containing δ_0) otherwise. Subsets of interest here are typically characterized by the existence of moments of certain order, viz.

$$\mathscr{P}^p(\mathbb{R}) := \left\{F \in \mathscr{P}(\mathbb{R}) : \int |x|^p \, F(dx) < \infty\right\}, \tag{5.4}$$

for any $p > 0$ or, more specifically, the set of all centered and standardized distributions on $\mathbb{R}$, that is

$$\mathscr{P}^2_{0,1}(\mathbb{R}) := \left\{F \in \mathscr{P}(\mathbb{R}) : \int x \, F(dx) = 0 \text{ and } \int x^2 \, F(dx) = 1\right\}. \tag{5.5}$$

Section 5.4 will provide conditions ensuring that $\mathscr{S}$ is a self-map on $\mathscr{P}^p(\mathbb{R})$, and these do not necessarily entail (A0). A standing assumption throughout this chapter is that

$$\mathbb{P}(N \geq 2) > 0 \tag{A1}$$

because $N \leq 1$ a.s. obviously leads back to RDE's studied at length in the previous chapter. Our primary goals are,

1. for subsets of $\mathscr{P}(\mathbb{R})$ as just mentioned, to provide conditions under which $\mathscr{S}$ is contractive with respect to a suitable metric.
2. in $\mathscr{P}(\mathbb{R})$ or again subsets thereof, to study existence and uniqueness of fixed points of $\mathscr{S}$, characterized by the SFPE

$$X \stackrel{d}{=} \sum_{k \geq 1} T_k X_k + C \tag{5.6}$$

   when stated in terms of random variables, where as usual $X_1, X_2, \ldots$ are iid copies of X and independent of $(C, T_1, T_2, \ldots)$.

By going back to Chapter 1, the reader can find a number of examples where distributions solving an SFPE of type (5.6) appear:

- ☞ Section 1.1, notably (1.6) giving a characterization of limit laws of suitably normalized sums of iid random variables.
- ☞ Section 1.3, notably (1.14) characterizing the law of the total size of a Galton-Watson population.
- ☞ Section 1.4 on the `Quicksort` algorithm, notably (1.26) characterizing its limit law after suitable normalization.

A key tool in the analysis of $\mathscr{S}$ is the study of its iterations $\mathscr{S}^n$ when described in terms of random variables. This leads to the so-called *weighted branching model* that will be formally introduced in the next section.

5.1 Setting up the stage: the weighted branching model

In order to motivate the subsequent definitions, consider the *homogeneous smoothing transform*, defined by (5.3) with $C=0$. Then it is easily seen that

$$\begin{aligned} \mathscr{S}^2(F) &= \mathscr{L}\left(\sum_{i\ge 1}\sum_{j\ge 1} T_iT_j(\mathsf{i})X_{ij}\right) \\ \text{and}\quad \mathscr{S}^3(F) &= \mathscr{L}\left(\sum_{i\ge 1}\sum_{j\ge 1}\sum_{k\ge 1} T_iT_j(\mathsf{i})T_k(\mathsf{ij})X_{ijk}\right) \end{aligned}$$

where $(T_n)_{n\ge 1}$ and the $(T_n(\mathsf{i}))_{n\ge 1}$, $(T_n(\mathsf{ij}))_{n\ge 1}$ for $i,j\ge 1$ are iid and independent of the iid X_{ij},X_{ijk} with common distribution F. This reveals the branching nature behind the recursion defined by $\mathscr{S}$ which will now be formalized.

Consider the infinite *Ulam-Harris tree*

$$\mathbb{T} := \bigcup_{n\ge 0}\mathbb{N}^n, \quad \mathbb{N}^0 := \{\varnothing\},$$

of finite integer words having the empty word $\varnothing$ as its root. As already explained in Section 1.7, we write $\mathsf{v}_1...\mathsf{v}_n$ as shorthand for $(\mathsf{v}_1,...,\mathsf{v}_n)$, $|\mathsf{v}|$ for the length of v, and uv for the concatenation of u and v. If $\mathsf{v}=\mathsf{v}_1...\mathsf{v}_n$, put further $v|0:=\varnothing$ and $\mathsf{v}|k:=\mathsf{v}_1...\mathsf{v}_k$ for $1\le k\le n$. The unique shortest path (geodesic) from the root $\varnothing$ to v, or the ancestral line of v when using a genealogical interpretation, is then given by

$$\mathsf{v}|0=\varnothing \;\to\; \mathsf{v}|1 \;\to ... \to\; \mathsf{v}|n-1 \;\to\; \mathsf{v}|n=\mathsf{v}.$$

The tree $\mathbb{T}$ is now turned into a *weighted (branching) tree* by attaching a *random weight* to each of its edges. Let $T_i(\mathsf{v})$ denote the weight attached to the edge (v,vi) and assume that the $T(\mathsf{v}) := (T_i(\mathsf{v}))_{i\ge 1}$ for $\mathsf{v}\in\mathbb{T}$ form a family of iid copies of $T=(T_i)_{i\ge 1}$. The number of nonzero weights $T_i(\mathsf{v})$ is denoted $N(\mathsf{v})$, thus

$$N(\mathsf{v}) \;:=\; \sum_{i\ge 1}\mathbf{1}_{\{T_i(\mathsf{v})\ne 0\}} \;\stackrel{d}{=}\; N.$$

Put further $L(\varnothing):=1$ and then recursively

$$L(\mathsf{vi}) \;:=\; L(\mathsf{v})T_i(\mathsf{v}) \tag{5.7}$$

for any $\mathsf{v}\in\mathbb{T}$ and $\mathsf{i}\in\mathbb{N}$, which is equivalent to

$$L(\mathsf{v}) \;=\; T_{\mathsf{v}_1}(\varnothing)T_{\mathsf{v}_2}(\mathsf{v}|1)\cdot\ldots\cdot T_{\mathsf{v}_n}(\mathsf{v}|n-1) \tag{5.8}$$

for any $\mathsf{v}=\mathsf{v}_1...\mathsf{v}_n\in\mathbb{T}$. Hence, $L(\mathsf{v})$ equals the total weight of the minimal path from $\varnothing$ to v obtained upon multiplication of the edge weights along this path.

With the help of a weighted branching model as just introduced, we are now able to describe the iterations of the homogeneous smoothing transform in a convenient way.

Lemma 5.1. *Given the previous notation and $\mathscr{S}$ defined by* (5.3) *with $C=0$, let $\boldsymbol{X}:=\{X(\mathsf{v}):\mathsf{v}\in\mathbb{T}\}$ be a family of iid random variables independent of $\boldsymbol{T}:=(T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ with common distribution F and define*

$$Y_n\ :=\ \sum_{|\mathsf{v}|=n}L(\mathsf{v})X(\mathsf{v}),\quad n\ge 0, \tag{5.9}$$

called ***weighted branching process (WBP) associated with***

$$\boldsymbol{T}\otimes\boldsymbol{X}:=(T(\mathsf{v}),X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}.$$

Then $\mathscr{S}^n(F)=\mathscr{L}(Y_n)$ holds true for each $n\ge 0$.

Proof. The easy proof is left as an exercise [☞ Problem 5.6]. □

In the special case when $X(\mathsf{v})=1$ for $\mathsf{v}\in\mathbb{T}$, the sequence $\boldsymbol{Y}$ defined by (5.9) is simply called *weighted branching process associated with $\boldsymbol{T}$*.

It is not difficult to extend the previous weighted branching model so as to describe the iterations of $\mathscr{S}$ in the nonhomogeneous case when $\mathbb{P}(C=0)<1$. To this end, let

$$\boldsymbol{C}\otimes\boldsymbol{T}\ :=\ (C(\mathsf{v}),T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$$

denote a family of iid copies of (C,T), $T:=(T_i)_{i\ge 1}$. The following lemma provides the extension of the previous one by allowing a general C in the definition of $\mathscr{S}$.

Lemma 5.2. *Given the previous notation and $\mathscr{S}$ defined by* (5.3)*, let $\boldsymbol{X}:=\{X(\mathsf{v}):\mathsf{v}\in\mathbb{T}\}$ be a family of iid random variables independent of $\boldsymbol{C}\otimes\boldsymbol{T}$ with common distribution F. Define $Y(\varnothing)=X(\varnothing)$ and*

$$Y_n\ :=\ \sum_{k=0}^{n-1}\sum_{|\mathsf{v}|=k}L(\mathsf{v})C(\mathsf{v})+\sum_{|\mathsf{v}|=n}L(\mathsf{v})X(\mathsf{v}),\quad n\ge 1, \tag{5.10}$$

called ***weighted branching process associated with***

$$\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}:=(C(\mathsf{v}),T(\mathsf{v}),X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}.$$

Then $\mathscr{S}^n(F)=\mathscr{L}(Y_n)$ holds true for each $n\ge 0$.

Proof. The proof is again left as an exercise [☞ Problem 5.6]. □

We proceed to a description of the recursive structure of WBPs after the following useful definition of the *shift operators* $[\cdot]_{\mathsf{v}}$, $\mathsf{v}\in\mathbb{T}$. Given any function Ψ of $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}$ and any $\mathsf{v}\in\mathbb{T}$, put

$$[\Psi(\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X})]_{\mathsf{v}} \;:=\; \Psi\left((C(\mathsf{vw}),T(\mathsf{vw}),X(\mathsf{vw}))_{\mathsf{w}\in\mathbb{T}}\right), \tag{5.11}$$

which particularly implies

$$[\Psi(\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X})]_{\mathsf{v}} \;=\; \Psi\left([\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}]_{\mathsf{v}}\right). \tag{5.12}$$

If we think of $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}$ as the family of random variables associated with $\mathbb{T}$, then $[\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}]_{\mathsf{v}}$ equals its subfamily and copy associated with the subtree $\mathbb{T}(\mathsf{v})$ rooted at v which is isomorphic to $\mathbb{T}$. Obviously, $\boldsymbol{L}:=(L(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ is a function of $\boldsymbol{T}$, and one can easily verify that $[\boldsymbol{L}]_{\mathsf{v}}=([L(\mathsf{w})]_{\mathsf{v}})_{\mathsf{w}\in\mathbb{T}}$ with

$$[L(\mathsf{w})]_{\mathsf{v}} \;:=\; T_{\mathsf{w}_1}(\mathsf{v})T_{\mathsf{w}_2}(\mathsf{vw}_1)\cdot\ldots\cdot T_{\mathsf{w}_n}(\mathsf{vw}_1\ldots\mathsf{w}_{n-1}) \tag{5.13}$$

if $\mathsf{w}=\mathsf{w}_1\ldots\mathsf{w}_n$. Hence, $[L(\mathsf{w})]_{\mathsf{v}}$ gives the total weight of the minimal path from v to vw. Notice that, for all $\mathsf{v},\mathsf{w}\in\mathbb{T}$,

$$L(\mathsf{vw}) \;=\; L(\mathsf{v})\cdot[L(\mathsf{w})]_{\mathsf{v}} \tag{5.14}$$

and therefore

$$[L(\mathsf{w})]_{\mathsf{v}} \;=\; \frac{L(\mathsf{vw})}{L(\mathsf{v})} \tag{5.15}$$

for all $\mathsf{w}\in\mathbb{T}$ if $L(\mathsf{v})\neq 0$.

Returning to the WBP $\boldsymbol{Y}=(Y_n)_{n\geq 0}$, clearly a function of $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}$, put $\boldsymbol{Y}(\mathsf{v})=(Y_n(\mathsf{v}))_{n\geq 0}:=[\boldsymbol{Y}]_{\mathsf{v}}$ for $\mathsf{v}\in\mathbb{T}$, hence $Y_0(\mathsf{v}):=X(\mathsf{v})$ and

$$Y_n(\mathsf{v}) \;:=\; \sum_{k=0}^{n-1}\sum_{|\mathsf{w}|=k}[L(\mathsf{w})]_{\mathsf{v}}C(\mathsf{vw})+\sum_{|\mathsf{w}|=n}[L(\mathsf{w})]_{\mathsf{v}}X(\mathsf{vw}),\quad n\geq 1.$$

Some straightforward facts about the $\boldsymbol{Y}(\mathsf{v})$ are collected in the following lemma.

Lemma 5.3.

(a) *For each* $\mathsf{v}\in\mathbb{T}$, $\boldsymbol{Y}(\mathsf{v})$ *is the WBP associated with* $[\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}]_{\mathsf{v}}$ *and a copy of* $\boldsymbol{Y}=\boldsymbol{Y}(\varnothing)$.

(b) *For each* $n\geq 1$, *the processes* $\boldsymbol{Y}(\mathsf{v})$, $\mathsf{v}\in\mathbb{T}$ *with* $|\mathsf{v}|=n$, *are mutually independent and also independent of* $(C(\mathsf{u}),T(\mathsf{u}),X(\mathsf{u}))_{|\mathsf{u}|\leq n-1}$.

(c) *For each* $\mathsf{v}\in\mathbb{T}$ *and* $n\geq 1$, *the* ***backward equation***

$$Y_n(\mathsf{v}) \;=\; C(\mathsf{v})+\sum_{i\geq 1}T_i(\mathsf{v})Y_{n-1}(\mathsf{v}i) \tag{5.16}$$

holds true.

Proof. Since (a) and (b) are obvious, we confine ourselves to the proof of (c). W.l.o.g. choosing $\mathsf{v} = \varnothing$, it follows from (5.10) and (5.14) that

$$\begin{aligned} Y_n &= C(\varnothing) + \sum_{i\ge 1}\sum_{k=1}^{n-2}\left(\sum_{|\mathsf{w}|=k} L(\mathsf{iw})C(\mathsf{iw}) + \sum_{|\mathsf{w}|=n-1} L(\mathsf{iw})X(\mathsf{iw})\right) \\ &= C(\varnothing) + \sum_{i\ge 1} T_i(\varnothing)\left(C(\mathsf{i}) + \sum_{k=1}^{n-2}\sum_{|\mathsf{w}|=k}[L(\mathsf{w})]_{\mathsf{i}}\,C(\mathsf{iw}) + \sum_{|\mathsf{w}|=n-1}[L(\mathsf{w})]_{\mathsf{i}}\,X(\mathsf{iw})\right) \\ &= C(\varnothing) + \sum_{i\ge 1} T_i(\varnothing) Y_{n-1}(\mathsf{i}) \end{aligned}$$

for each $n \ge 1$ where, as usual, empty sums are defined as zero. □

Let us return to the homogeneous case when $C = 0$ and further assume $X(\mathsf{v}) = 1$ for all $\mathsf{v} \in \mathbb{T}$. Then the following useful martingale result is easily obtained. We put

$$\mathscr{F}_n := \sigma\left(T(\mathsf{v}) : |\mathsf{v}| \le n-1\right) \tag{5.17}$$

for $n \ge 1$ and let $\mathscr{F}_0$ be the trivial σ-field. Observe that

$$\mathscr{F}_n \supset \sigma(L(\mathsf{v}) : |\mathsf{v}| \le n)$$

for each $n \ge 0$.

Lemma 5.4. *If $C = 0$ and $X(\mathsf{v}) = 1$ for $\mathsf{v} \in \mathbb{T}$, then the WBP $\boldsymbol{Y}$ associated with $\boldsymbol{T}$ constitutes a (mean one) martingale with respect to $(\mathscr{F}_n)_{n\ge 0}$ if*

$$\mathbb{E}\left(\sum_{i\ge 1}|T_i|\right) < \infty \quad \text{and} \quad \mathbb{E}\left(\sum_{i\ge 1} T_i\right) = 1.$$

Proof. Under the given assumptions, $Y_n = \sum_{|\mathsf{v}|=n} L(\mathsf{v})$ for $n \ge 0$,

$$\mathbb{E}|Y_1| \;\le\; \mathbb{E}\left(\sum_{i\ge 1}|T_i|\right) \;<\; \infty \quad \text{and} \quad \mathbb{E}(Y_1|\mathscr{F}_0) \;=\; \mathbb{E}Y_1 \;=\; 1 \;=\; Y_0.$$

For general n, we use an inductive argument and assume that $\mathbb{E}(Y_k|\mathscr{F}_{k-1}) = Y_{k-1}$ a.s. and $\mathbb{E}(\sum_{|\mathsf{v}|=k}|L(\mathsf{v})|) < \infty$ for $k = 1,\dots,n-1$ (inductive hypothesis). Then, by further using the independence of $\mathscr{F}_{n-1}$ and $(T(\mathsf{v}))_{|\mathsf{v}|=n-1}$, we infer

$$\mathbb{E}|Y_n| \;=\; \mathbb{E}\left|\sum_{|\mathsf{v}|=n-1} L(\mathsf{v})\sum_{i\ge 1} T_i(\mathsf{v})\right| \;\le\; \mathbb{E}\left(\sum_{|\mathsf{v}|=n-1} |L(\mathsf{v})|\right)\mathbb{E}\left|\sum_{i\ge 1} T_i\right| \;<\; \infty$$

as well as

$$\begin{aligned}\mathbb{E}(Y_n|\mathscr{F}_{n-1}) \;&=\; \mathbb{E}\left(\sum_{|\mathsf{v}|=n-1} L(\mathsf{v})\sum_{i\ge 1} T_i(\mathsf{v})\middle|\mathscr{F}_{n-1}\right)\\ &=\; \sum_{|\mathsf{v}|=n-1} L(\mathsf{v})\,\mathbb{E}\left(\sum_{i\ge 1} T_i(\mathsf{v})\middle|\mathscr{F}_{n-1}\right)\\ &=\; \sum_{|\mathsf{v}|=n-1} L(\mathsf{v}) \quad \text{a.s.}\end{aligned}$$

which proves the assertion because the last sum equals Y_{n-1}. □

The reader is asked in Problem 5.8 to show by providing examples that Lemma 5.4 may or may not hold if $\mathbb{E}(\sum_{i\ge 1}|T_i|)<\infty$ is replaced with the weaker condition $\mathbb{E}|\sum_{i\ge 1}T_i|<\infty$. If T is nonnegative, i.e. $T_i\ge 0$ for all $i\ge 1$, then these two conditions are evidently identical, and in combination with $\mathbb{E}(\sum_{i\ge 1}T_i)=1$ actually equivalent to the martingale property of $\boldsymbol{Y}$ as one can easily check.

If all $T(\mathsf{v})$ are nonnegative, which is the situation encountered in most applications, the following result is easily derived from the previous one.

Lemma 5.5. *In the situation of the Lemma 5.4, suppose further that $T_i\ge 0$ for all $i\ge 1$ and*

$$\mathfrak{m}(\theta) \;:=\; \mathbb{E}\left(\sum_{i\ge 1} T_i^{\theta}\right) \;<\; \infty,$$

for some $\theta\ge 0$, where $T_i^0:=\mathbf{1}_{\{T_i>0\}}$. Then

$$W_{\theta,n} \;:=\; \sum_{|\mathsf{v}|=n}\frac{L(\mathsf{v})^{\theta}}{\mathfrak{m}(\theta)^n},\quad n\ge 0$$

is the WBP associated with $(T(\mathsf{v})^{\theta}/\mathfrak{m}(\theta))_{\mathsf{v}\in\mathbb{T}}$ and constitutes a martingale with respect to $(\mathscr{F}_n)_{n\ge 0}$.

Proof. Problem 5.10. □

We leave it to the reader to verify that

$$Y_{0,n} \;:=\; \sum_{|v|=n} L(v)^0 \;=\; \sum_{|v|=n}\mathbf{1}_{\{L(v)>0\}},\quad n\ge 0$$

forms a simple GWP with offspring distribution $(p_n)_{n\ge 0}$, where $p_n:=\mathbb{P}(N=n)$, and offspring mean $\mathbb{E}N=\mathfrak{m}(0)$ [☞ Problem 5.11].

Problems

Problem 5.6. Prove Lemmata 5.1 and 5.2.

Problem 5.7. Prove the following generalization of the backward equation (5.16):

$$Y_n(\mathsf{v}) \;=\; C(\mathsf{v})+\sum_{j=1}^{k-1}\sum_{|\mathsf{w}|=j}[L(\mathsf{w})]_{\mathsf{v}}\,C(\mathsf{vw})+\sum_{|\mathsf{w}|=k}[L(\mathsf{w})]_{\mathsf{v}}\,Y_{n-k}(\mathsf{vw}) \tag{5.18}$$

for all $\mathsf{v}\in\mathbb{T}$ and $n\ge k\ge 1$.

Problem 5.8. Consider the situation of Lemma 5.4 and give two examples of weight sequences $T=(T_i)_{i\ge 1}$ satisfying

$$\mathbb{E}\left(\sum_{i\ge 1}|T_i|\right)=\infty,\quad \mathbb{E}\left|\sum_{i\ge 1}T_i\right|<\infty\quad\text{and}\quad \mathbb{E}\left(\sum_{i\ge 1}T_i\right)=1,$$

such that the WBP $\boldsymbol{Y}$ associated with $\boldsymbol{T}$ is a martingale with respect to $(\mathscr{F}_n)_{n\ge 0}$ in the first case, while failing to be integrable (i.e. $\mathbb{E}|Y_n|=\infty$ for some $n\ge 1$) in the second one.

Problem 5.9. Suppose that $T_1,T_2,\ldots$ are nonnegative and consider the function

$$\mathfrak{m}(\theta)\;:=\;\mathbb{E}\left(\sum_{i\ge 1}T_i^{\theta}\right)$$

on its canonical domain $\mathbb{D}_{\mathfrak{m}}:=\{\theta\ge 0:\mathfrak{m}(\theta)<\infty\}$. Prove the following assertions:

(a) If $\mathbb{D}_{\mathfrak{m}}\ne\emptyset$, then $\mathbb{D}_{\mathfrak{m}}$ is an interval (possibly containing only one element).

(b) The function m is convex on $\mathbb{D}_m$ and further infinitely often differentiable on the interior of $\mathbb{D}_m$ if the latter is nonempty. In this case, the k^{th} derivative $m^{(k)}$ is given by

$$\mathfrak{m}^{(k)}(\theta)\;=\;\mathbb{E}\left(\sum_{i\ge 1}T_i^{\theta}\log^k T_i\right)$$

for any $k\in\mathbb{N}$ with the usual convention that $x^{\theta}\log^k x:=0$ if $x=0$.

Problem 5.10. Prove Lemma 5.5.

Problem 5.11. Suppose that $\boldsymbol{T}$ has nonnegative entries. Show that the WBP $(Y_{0,n})_{n\ge 0}$ associated with $\boldsymbol{T}^0:=(T(\mathsf{v})^0)_{\mathsf{v}\in\mathbb{T}}$ is a simple GWP with offspring distribution given by the law of $N=\sum_{i\ge 1}\mathbf{1}_{\{T_i>0\}}$.

5.2 A digression: weighted branching and random fractals

Fractal geometry provides another area where the weighted branching model introduced before enters in a natural way when dealing with so-called *random recursive constructions*, a notion coined by MAULDIN & WILLIAMS in [85].

A *Cantor set* is usually defined as a compact and perfect (= closed with no isolated points) subset of $\mathbb{R}^d$ having topological dimension zero. The most prominent example is the *Cantor ternary set*, which is constructed by initially removing the open middle third from the unit interval $[0,1]$ and by then indefinitely doing the same with the remaining subintervals, the latter being $[0,1/3]$ and $[2/3,1]$ after the first round, $[0,1/9]$, $[2/9,1/3]$, $[2/3,7/9]$ and $[8/9,1]$ after the second round, and so on. Here is a generalization: Fix positive numbers a,b, $0<a<b<1$, and remove (a,b) from $[0,1]$ in the first step. Apply the same procedure to the remaining subintervals $[0,a]$ and $[b,1]$ in Step 2, then to $[0,a^2]$, $[ab,a]$, $[b,b+a(1-b)]$ and $[b(2-b),1]$ in Step 3, and so on. The thus constructed Cantor set is defined by

$$\mathfrak{C}_{a,b} \;=\; \bigcap_{n\ge 0}\bigcup_{k=1}^{2^n} I_{n,k}, \tag{5.19}$$

where $I_{0,1} := [0,1]$ and $I_{n,1},...,I_{n,2^n}$ for $n\ge 1$ are the remaining subintervals after n steps.

A *random recursive construction* of a subset of $\mathbb{R}^d$ as defined in [85] generalizes this procedure even further. Let $|\cdot|$ denote Euclidean metric on $\mathbb{R}^d$ and, for any $A\subset\mathbb{R}^d$,

$$\mathit{diam}(A) := \sup\{|x-y| : x,y\in A\}, \quad \mathit{int}(A) \quad \text{and} \quad \mathit{cl}(A)$$

the diameter, interior and closure of A, respectively. Two sets A,B are called *nonoverlapping* if they have no common interior points $(\mathit{int}(A)\cap\mathit{int}(B)=\emptyset)$ and *geometrically similar* if A may be obtained from B via translation, rotation, reflection and dilation. A sequence $(A_n)_{n\ge 1}$ of pairwise nonoverlapping subsets of $\mathbb{R}^d$ such that $A=\bigcup_{n\ge 1}A_n$ is called a *quasi-partition* of A hereafter.

Now fix any compact $\emptyset\ne J\subset\mathbb{R}^d$ with the additional property that $\mathit{cl}(\mathit{int}(J))=J$, which guarantees $\lambda\!\!\lambda^d(J)>0$ as well as $\lambda\!\!\lambda^d(J)=\lambda\!\!\lambda^d(\mathit{int}(J))$. Defined on a probability space $(\Omega,\mathfrak{A},\mathbb{P})$, let $\boldsymbol{J}=\{J_{\mathsf{v}}:\mathsf{v}\in\mathbb{T}\}$ be a family of random subsets of J with the following properties:

(RC-1) $J_\varnothing = J$ a.s., and $J_{\mathsf{v}}(\omega)$, if nonempty, is geometrically similar to J for any $\mathsf{v}\in\mathbb{T}$ and $\omega\in\Omega$.

(RC-2) Defining $J_{\mathsf{v}0}:=J_{\mathsf{v}}\setminus\bigcup_{i\ge 1}J_{\mathsf{v}i}$, we have a.s. that $\lambda\!\!\lambda^d(\mathit{int}(J_{\mathsf{v}0}))>0$ and $(J_{\mathsf{v}i})_{i\ge 0}$ forms a quasi-partition of J_{v} for each $\mathsf{v}\in\mathbb{T}$.

(RC-3) The random sequences $T(\mathsf{v})=(T_i(\mathsf{v}))_{i\ge 1}$, $\mathsf{v}\in\mathbb{T}$, defined by

$$T_i(\mathsf{v}) \;:=\; \frac{\mathit{diam}(J_{\mathsf{v}i})}{\mathit{diam}(J_{\mathsf{v}})}\,\mathbf{1}_{\{J_{\mathsf{v}}\ne\emptyset\}} \tag{5.20}$$

are iid with $\mathbb{P}(N(\mathsf{v}) \geq 1) > 0$.

The *d-dimensional random Cantor set*

$$\mathfrak{C} := \bigcap_{n \geq 0} \mathfrak{C}_n,$$

where $\mathfrak{C}_0 := J$ and

$$\mathfrak{C}_n := \bigcup_{|\mathsf{v}|=n-1} \bigcup_{i \geq 1} J_{\mathsf{v}i} = \bigcup_{|\mathsf{v}|=n} J_{\mathsf{v}}$$

for $n \geq 1$, is then obtained by the following recursive construction: In the first step, the set J, quasi-partitioned by $(J_{\mathsf{i}})_{i \geq 0}$, is reduced by the random set J_0. In the next step, each of the remaining J_{i}, quasi-partitioned by $(J_{\mathsf{ij}})_{j \geq 0}$, is reduced by the set $J_{\mathsf{i}0}$, and so on. The reader is asked in Problem 5.13(a) to verify that $\mathfrak{C}$ is nonempty with positive probability if, in addition to (RC-1)-(RC-3),

$$\mathbb{E}N(\mathsf{v}) > 1 \tag{RC-4}$$

holds true.

As an example, consider the unit square $[0,1]^2$ divided into four congruent subsquares of which one is removed at random. Then the same procedure is applied to the remaining three subsquares, and so on. The first three steps are illustrated in Fig. 5.1 below. As a second example consider the set $\mathfrak{C}_{a,b}$ introduced above, but with randomly chosen a,b.

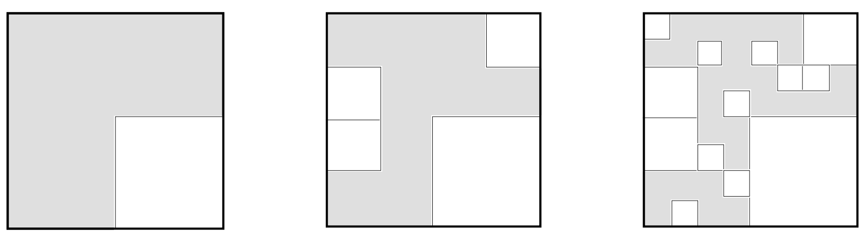


Fig. 5.1 Construction of a 2-dimensional random Cantor set: The removed squares are shown in white.

A rather simple argument [☞ Problem 5.13(c)] shows that

$$T_i(\mathsf{v})^d = \frac{\lambda^d(J_{\mathsf{v}i})}{\lambda^d(J_{\mathsf{v}})} = \frac{\lambda^d(int(J_{\mathsf{v}i}))}{\lambda^d(int(J_{\mathsf{v}}))} \quad \text{if } J_{\mathsf{v}} \neq \emptyset, \tag{5.21}$$

holds true and therefore

$$\sum_{i \geq 1} T_i(\mathsf{v})^d = \sum_{i \geq 1} \frac{\lambda^d(int(J_{\mathsf{v}i}))}{\lambda^d(int(J_{\mathsf{v}}))}$$

$$= \frac{\lambda\!\!\lambda^d(\sum_{i\ge 1} int(J_{\mathsf{v}i}))}{\lambda\!\!\lambda^d(int(J_{\mathsf{v}}))}$$
$$\le 1 - \frac{\lambda\!\!\lambda^d(int(J_{\mathsf{v}0}))}{\lambda\!\!\lambda^d(int(J_{\mathsf{v}}))} < 1$$

for all $\mathsf{v} \in \mathbb{T}$, where $int(J_{\mathsf{v}i}) \subset int(J_{\mathsf{v}})$ and $int(J_{\mathsf{v}i}) \cap int(J_{\mathsf{v}j}) = \emptyset$ for any $i, j \ge 0$, $i \ne j$ should be observed. In particular,

$$\mathfrak{m}(d) < 1, \tag{5.22}$$

where

$$\mathfrak{m}(\theta) := \mathbb{E}\left(\sum_{i\ge 1} T_i(\mathsf{v})^\theta\right) = \sum_{i\ge 1} \mathbb{E}T_i(\mathsf{v})^\theta$$

for $\theta \ge 0$ [recall that $T_i(\mathsf{v})^0 := \mathbf{1}_{\{T_i(\mathsf{v})>0\}}$ and thus $\mathfrak{m}(0) = \mathbb{E}N(\mathsf{v})$ holds true].

For $\mathsf{v} \in \mathbb{T}$ and $\theta \ge 0$, we now further define $X(\mathsf{v}) := 1$ and let $(Y_{\theta,n})_{n\ge 0}$ denote the WBP associated with $\boldsymbol{T}^\theta = (T(\mathsf{v})^\theta)_{\mathsf{v}\in\mathbb{T}}$. In particular,

$$Y_{\theta,1} := \sum_{i\ge 1} T_i(\varnothing)^\theta.$$

It follows with the help of (5.8) and (5.20) that

$$L(\mathsf{v}) = T_{\mathsf{v}_1}(\varnothing)\prod_{i=1}^{n-1} T_{\mathsf{v}_{i+1}}(\mathsf{v}|i) = \frac{diam(J_{\mathsf{v}_1})}{diam(J)}\prod_{i=1}^{n(\mathsf{v})-1}\frac{diam(J_{\mathsf{v}|i+1})}{diam(J_{\mathsf{v}|i})} = \frac{diam(J_{\mathsf{v}})}{diam(J)},$$

if $|\mathsf{v}| = n \ge 1$ and $n(\mathsf{v}) := \inf\{k \ge 1 : diam(J_{\mathsf{v}|k}) = 0 \text{ or } k = n\}$. Therefore,

$$Y_{\theta,n} = \sum_{|\mathsf{v}|=n} L(\mathsf{v})^\theta = \sum_{|\mathsf{v}|=n}\frac{diam(J_{\mathsf{v}})^\theta}{diam(J)^\theta} \tag{5.23}$$

for all $n \ge 1$ and $\theta \ge 0$. For the case $\theta = d$, we find as in (5.21) that

$$L(\mathsf{v})^d = \frac{diam(J_{\mathsf{v}})^d}{diam(J)^d} = \frac{\lambda\!\!\lambda^d(int(J_{\mathsf{v}}))}{\lambda\!\!\lambda^d(int(J))} = \frac{\lambda\!\!\lambda^d(J_{\mathsf{v}})}{\lambda\!\!\lambda^d(J)},$$

and thereby that

$$Y_{d,n} = \sum_{|\mathsf{v}|=n} L(\mathsf{v})^d = \sum_{|\mathsf{v}|=n}\frac{\lambda\!\!\lambda^d(J_{\mathsf{v}})}{\lambda\!\!\lambda^d(J)} \ge \frac{\lambda\!\!\lambda^d(\mathfrak{C}_n)}{\lambda\!\!\lambda^d(J)} \tag{5.24}$$

for all $n \ge 1$. By Lemma 5.5, $(\mathfrak{m}(d)^{-n} Y_{d,n})_{n\ge 0}$ forms a nonnegative and hence a.s. convergent martingale which in combination with (5.22) implies

$$\lim_{n\to\infty} \mathbb{E}Y_{d,n} = \lim_{n\to\infty} \mathfrak{m}(d)^n\, \mathbb{E}Y_{d,0} = 0$$

and then also $Y_{d,n} \to 0$ a.s. But (5.24) further gives $\lim_{n\to\infty} Y_{d,n} \ge \lambda\!\!\lambda^d(\mathfrak{C})/\lambda\!\!\lambda^d(J)$ and finally leads to the conclusion that

$$\lambda\!\!\lambda^d(\mathfrak{C}) = 0. \tag{5.25}$$

Every random Cantor set is therefore a Lebesgue null set under the stated conditions (RC-1)-(RC-4) and particularly contains no inner points. A far more difficult and surprising result is the following theorem by MAULDIN & WILLIAMS [85, Theorem 1.1] about the *Hausdorff dimension* of $\mathfrak{C}$ [☞ (5.28) and (5.29) below for its definition]. For related work see also the articles by HUTCHINSON [68], FALCONER [47], and GRAF [59].

Theorem 5.12. *Let $\boldsymbol{J} = \{J_v : v \in \mathbb{T}\}$ be a family of random compact subsets of $\mathbb{R}^d$ in a random recursive construction satisfying (RC-1)-(RC-4) and let $\mathfrak{C}$ be the associated random Cantor set having Hausdorff dimension* $\dim_H \mathfrak{C}$. *Then*

$$\alpha \ := \ \inf\{\theta \ge 0 : \mathfrak{m}(\theta) \le 1\} \ \in \ (0,d] \tag{5.26}$$

and

$$\mathbb{P}(\dim_H \mathfrak{C} = \alpha | \mathfrak{C} \neq \emptyset) = 1. \tag{5.27}$$

The θ*-dimensional Hausdorff (outer) measure on* $\mathbb{R}^d$ is defined by

$$\mathscr{H}_\theta^d(A) \ := \ \lim_{\varepsilon\downarrow 0} \ \inf\left\{\sum_{n\ge 1} \mathit{diam}(A_n)^\theta : (A_n)_{n\ge 1} \ \varepsilon\text{-covering of } A\right\} \tag{5.28}$$

for any $A \subset \mathbb{R}^d$ and constitutes a measure on $\mathscr{B}(\mathbb{R}^d)$, where $(A_n)_{n\ge 1}$ is called an ε*-covering of A* if

$$\mathit{diam}(A_n) \le \varepsilon \text{ for all } n \ge 1 \quad \text{and} \quad A \subset \bigcup_{n\ge 1} A_n.$$

The Hausdorff dimension of A is the unique number $\dim_H A \in [0,\infty]$ characterized by

$$\mathscr{H}_\theta^d(A) \ = \ \begin{cases} \infty, & \text{if } \theta < \dim_H A, \\ 0, & \text{if } \theta > \dim_H A. \end{cases} \tag{5.29}$$

More detailed information may be found in Appendix B.

Proof (the easy half). The difficult part is to show that $\mathbb{P}(\dim_H \mathfrak{C} \ge \alpha | \mathfrak{C} \neq \emptyset) = 1$ and must wait until Section 5.7. Here we confine ourselves to the simpler converse that may be stated as $\dim_H \mathfrak{C} \le \alpha$ a.s., and for which $\alpha < d$ can be assumed, for otherwise there is nothing to verify.

First note that (5.22) and (RC-4) imply $\alpha \in (0,d]$. We have $\mathfrak{m}(\theta) < 1$ for any $\theta \in (\alpha,d]$ because $\mathfrak{m}(\theta)$ is obviously decreasing on $\mathbb{D}_m := \{\theta \ge 0 : \mathfrak{m}(\theta) < \infty\}$.

As a consequence, $Y_{\theta,n} \to 0$ a.s. for any $\theta \in (\alpha, d]$ by a similar argument as given before (5.25) for $\theta = d$.

W.l.o.g. let $diam(J) = 1$. Use $Y_{d,n} \to 0$ a.s. and $0 \le \sup_{\mathsf{v}\in\mathbb{T}} L(\mathsf{v}) \le 1$ to infer

$$\lim_{n\to\infty} L_n^* \;=\; 0 \quad \text{a.s.}$$

for $L_n^* := \sup_{|\mathsf{v}|=n} L(\mathsf{v})$. Since $diam(J_\mathsf{v}) = L(\mathsf{v})$ for all $\mathsf{v} \in \mathbb{T}$ with $J_\mathsf{v} \ne \emptyset$, we have that $(J_\mathsf{v})_{|\mathsf{v}|=n}$ forms a (random) L_n^*-covering of $\mathfrak{C}$ for each $n \ge 1$, which together with $L_n^* \to 0$ a.s. and $Y_{\theta,n} \to 0$ a.s. for $\theta \in (\alpha, d]$ finally implies

$$\mathscr{H}_\theta^d(\mathfrak{C}) \;\le\; \lim_{n\to\infty} \sum_{|\mathsf{v}|=n} diam(J_\mathsf{v})^\theta \;=\; \lim_{n\to\infty} Y_{\theta,n} \;=\; 0 \quad \text{a.s.}$$

and therefore $\dim_H \mathfrak{C} \le \alpha$ a.s. as claimed. □

Problems

Problem 5.13. Let $\boldsymbol{J} = \{J_\mathsf{v} : \mathsf{v} \in \mathbb{T}\}$ be a family of random compact subsets of $\mathbb{R}^d$ in a random recursive construction with associated random Cantor set $\mathfrak{C}$. Prove the following assertions:

(a) If $\mathbb{E}N(\mathsf{v}) > 1$ then $\mathbb{P}(\mathfrak{C}_n \ne \emptyset \text{ f.a. } n \ge 0) > 0$ and thus (why?) $\mathbb{P}(\mathfrak{C} \ne \emptyset) > 0$.

(b) If $N(\mathsf{v}) = 1$ a.s., then $\mathfrak{C}$ is a.s. a singleton set.

(c) For any two measurable and geometrically similar $A, B \subset \mathbb{R}^d$ with $\lambda\!\!\lambda^d(B) > 0$

$$\frac{\lambda\!\!\lambda^d(A)}{\lambda\!\!\lambda^d(B)} \;=\; \frac{diam(A)^d}{diam(B)^d}$$

holds true and thus particularly (5.21).

Problem 5.14. Divide the unit square $[0,1]^2$ into four congruent subsquares and remove one of these at random with probability $1/4$ each. Apply the same procedure to the remaining three subsquares, and so on [☞ Fig. 5.1].

(a) Give the distribution of the $T(\mathsf{v})$.

(b) Use Theorem 5.12 to determine the Hausdorff dimension of the resulting random Cantor set.

(c) How does the result in (b) change if the subsquare to be removed is picked in accordance with an arbitrary distribution (p_1, p_2, p_3, p_4)?

5.3 The minimal L^p-metric

In the following, let $\mathscr{P}^p(\mathbb{R})$ for $p > 0$ be the set of probability distributions on $\mathbb{R}$ with finite p^{th} absolute moment as defined in (5.4). Given a probability space

$(\Omega,\mathfrak{A},\mathbb{P})$, let further $L^p(\mathbb{P}) = L^p(\Omega,\mathfrak{A},\mathbb{P})$ denote the vector space of p times integrable random variables on $(\Omega,\mathfrak{A},\mathbb{P})$. Then it is well-known that $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$ defines a complete (pseudo-)norm on $L^p(\mathbb{P})$ if $p \ge 1$, but fails to do so if $0 < p < 1$ [☞ Problem 5.27]. On the other hand, when setting $\|X\|_p := \mathbb{E}|X|^p$ in the latter case, so that generally

$$\|X\|_p \;:=\; (\mathbb{E}|X|^p)^{1\wedge(1/p)},$$

it is not difficult to verify that

$$\ell_p(X,Y) \;:=\; \|X-Y\|_p \tag{5.30}$$

provides us with a complete (pseudo-)metric[1] on $L^p(\mathbb{P})$ for each $p > 0$ [☞ again Problem 5.27].

Recall from Remark 4.4, that a pair (X,Y) of real-valued random variables defined on $(\Omega,\mathfrak{A},\mathbb{P})$ is called (F,G)-*coupling* if $\mathscr{L}(X) = F$ and $\mathscr{L}(Y) = G$. In this case we will use the shorthand notation $(X,Y) \sim (F,G)$ hereafter. Recall further that the *pseudo-inverse* of a cdf F is defined as $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \ge u\}$ for $u \in (0,1)$ and that $F^{-1}(U)$ has distribution F if $\mathscr{L}(U) = \mathit{Unif}(0,1)$.

Proposition 5.15. *For each $p > 0$, the mapping $\ell_p : \mathscr{P}^p(\mathbb{R}) \times \mathscr{P}^p(\mathbb{R}) \to \mathbb{R}_\ge$, defined by*

$$\ell_p(F,G) \;:=\; \inf_{(X,Y)\sim(F,G)} \|X-Y\|_p, \tag{5.31}$$

*is a metric on $\mathscr{P}^p(\mathbb{R})$, called **minimal L^p-metric** (also **Mallows metric** in [97]), and therefore possesses the following properties:*

(1) $\ell_p(F,G) = 0$ *iff* $F = G$.
(2) $\ell_p(F,G) = \ell_p(G,F)$ *for all* $F,G \in \mathscr{P}^p(\mathbb{R})$.
(3) $\ell_p(F,H) \le \ell_p(F,G) + \ell_p(G,H)$ *for all* $F,G,H \in \mathscr{P}^p(\mathbb{R})$.

The infimum in (5.31) *is attained, namely*

$$\begin{aligned}\ell_p(F,G) \;&=\; \|F^{-1}(U) - G^{-1}(U)\|_p \\ &=\; \left(\int_0^1 |F^{-1}(u) - G^{-1}(u)|^p \, du\right)^{1\wedge(1/p)}\end{aligned} \tag{5.32}$$

for any Unif$(0,1)$ random variable U.

Proof. It suffices to show that the infimum is indeed attained in the asserted manner which will be accomplished by four subsequent lemmata. That ℓ_p defines a metric follows easily from the corresponding properties of ℓ_p [☞ Problem 5.27]. □

---

[1] A pseudo-metric d has the same properties as a metric with one exception: $d(x,y) = 0$ does not necessarily imply $x = y$.

Lemma 5.16. *Let (X,Y) be a (F,G)-coupling with cdf H on $\mathbb{R}^2$. Then*

$$H(x,y) \le F(x) \wedge G(y)$$

for all $x,y \in \mathbb{R}$ and equality holds if $(X,Y) = (F^{-1}(U), G^{-1}(U))$ for a Unif$(0,1)$ variable U.

Proof. The first assertion is trivial because

$$H(x,y) \;=\; \mathbb{P}(X \le x, Y \le y) \;\le\; P(X \le x) \wedge \mathbb{P}(Y \le y) \;=\; F(x) \wedge G(y)$$

for all $x,y \in \mathbb{R}$. As for the second assertion, we first note that

$$\{u : F^{-1}(u) \le x\} \;=\; \{u : F(x) \ge u\}$$

holds true for each distribution F and all $x \in \mathbb{R}$. This implies

$$\mathbb{P}(F^{-1}(U) \le x, G^{-1}(U) \le y) \;=\; \mathbb{P}(U \le F(x), U \le G(y)) \;=\; F(x) \wedge G(y)$$

for all $x,y \in \mathbb{R}$. ☐

Lemma 5.17. *For each $p > 0$, the function $k_p : \mathbb{R}^2 \to \mathbb{R}_{\ge}$, $(x,y) \mapsto |x-y|^p$ is continuous, symmetric and Δ-antitone, viz.*

$$\Delta_{x,y}^{x',y'} k_p \;:=\; k_p(x',y') + k_p(x,y) - k_p(x',y) - k_p(x,y') \;\le\; 0$$

for all $(x,y),(x',y') \in \mathbb{R}^2$ with $x \le x'$ and $y \le y'$.

Proof. Problem 5.28. ☐

Every continuous and Δ-antitone function k defines a unique continuous Borel measure on $\mathbb{R}^2$ via

$$\mathscr{K}([x,x'] \times [y,y']) \;:=\; -\Delta_{x,y}^{x',y'} k \tag{5.33}$$

for all $(x,y),(x',y') \in \mathbb{R}^2$ with $x \le x'$ and $y \le y'$. This is now used to show the following result.

Lemma 5.18. *Given a (F,G)-coupling (X,Y) with cdf H, let*

$$A_H(x,y) \;:=\; F(x \wedge y) + G(x \wedge y) - H(x \wedge y, x \vee y) - H(x \vee y, x \wedge y)$$

for $x,y \in \mathbb{R}$. Then every continuous, symmetric and Δ-antitone function k satisfies

$$2\,\mathbb{E}k(X,Y) \;=\; \mathbb{E}k(X,X) \;+\; \mathbb{E}k(Y,Y) \;+\; \int A_H(x,y)\;\mathscr{K}(dx,dy), \qquad (5.34)$$

that is, $\mathbb{E}k(X,Y) = \int k(x,y)\,H(dx,dy)$ *depends on* H *only through its marginals* F,G *and* A_H.

Proof. On $\Omega \times \mathbb{R}^2$, let Z be defined by

$$Z(\omega,x,y) := \mathbf{1}_{[X(\omega)\wedge Y(\omega),X(\omega)\vee Y(\omega))^2}(x,y).$$

Using (5.33) and the symmetry of k, it follows that on $\{X \le Y\}$

$$\begin{aligned}\int_{\mathbb{R}^2} Z(\cdot,x,y)\;\mathscr{K}(dx,dy) \;&=\; \mathscr{K}([X,Y)\times[X,Y)) \\ &=\; -\big(k(X,X)+k(Y,Y)-k(X,Y)-k(Y,X)\big) \\ &=\; 2k(X,Y)-k(X,X)-k(Y,Y), \end{aligned} \qquad (5.35)$$

and the same result is obviously found on $\{X \ge Y\}$. Writing Z as

$$Z(\omega,x,y) \;=\; \mathbf{1}_{(-\infty,x\wedge y]}(X(\omega))\mathbf{1}_{(x\vee y,\infty)}(Y(\omega))+\mathbf{1}_{(x\vee y,\infty)}(X(\omega))\mathbf{1}_{(-\infty,x\wedge y]}(Y(\omega)),$$

we further obtain

$$\begin{aligned}\mathbb{E}Z(\cdot,x,y) \;&=\; \mathbb{P}(X\le x\wedge y, Y>x\vee y)+\mathbb{P}(X>x\vee y, Y\le x\wedge y) \\ &=\; F(x\wedge y)-H(x\wedge y,x\vee y)+G(x\wedge y)-H(x\vee y,x\wedge y) \\ &=\; A_H(x,y). \end{aligned} \qquad (5.36)$$

Finally, we conclude by a combination of (5.35) and (5.36) that

$$\begin{aligned}2\,\mathbb{E}k(X,Y)-\mathbb{E}k(X,X)-\mathbb{E}k(Y,Y) \;&=\; \int_\Omega\int_{\mathbb{R}^2} Z(\omega,x,y)\;\mathscr{K}(dx,dy)\;\mathbb{P}(d\omega) \\ &=\; \int_{\mathbb{R}^2}\mathbb{E}Z(\cdot,x,y)\;\mathscr{K}(dx,dy) \\ &=\; \int_{\mathbb{R}^2} A_H(x,y)\;\mathscr{K}(dx,dy), \end{aligned}$$

which proves the asserted identity (5.34). □

For $k_p(x,y)=|x-y|^p$, which vanishes if $x=y$, formula (5.34) simplifies to

$$\mathbb{E}k_p(X,Y) \;=\; \frac{1}{2}\int A_H(x,y)\;\mathscr{K}_p(dx,dy), \qquad (5.37)$$

where $\mathscr{K}_p$ has the obvious meaning. The last ingredient to the proof of Proposition 5.15 is now provided by

Lemma 5.19. *Given $F,G \in \mathscr{P}^p(\mathbb{R})$ for some $p > 0$ and $U \stackrel{d}{=} Unif(0,1)$, any (F,G)-coupling (X,Y) satisfies*

$$\mathbb{E}k_p(X,Y) \;\geq\; \mathbb{E}k_p(F^{-1}(U),G^{-1}(U)).$$

Proof. As before, let H denote the cdf of (X,Y) and $H_*(x,y) := F(x) \wedge G(y)$ the cdf of the (F,G)-coupling $(F^{-1}(U),G^{-1}(U))$. By Lemma 5.16, $H(x,y) \leq H_*(x,y)$ for all $x,y \in \mathbb{R}$ which in turn implies $A_H(x,y) \geq A_{H_*}(x,y)$ for all $x,y \in \mathbb{R}$ and hence the assertion by an appeal to (5.37). □

If $p = 1$, then (5.32) implies that $\ell_1(F,G)$ equals the area between F^{-1} and G^{-1}. Since F^{-1},G^{-1} may be obtained as the reflections of F,G at the bisecting line $y = x$, this area must be the same as between F and G, thus giving the following result.

Corollary 5.20. *For $F,G \in \mathscr{P}^1(\mathbb{R})$,*

$$\ell_1(F,G) \;=\; \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx. \tag{5.38}$$

For any distribution $F \in \mathscr{P}^1(\mathbb{R})$ with mean value $\mu_F = \int x\, F(dx)$, let F^0 denote its centering, that is $F^0(t) := F(t + \mu_F)$ for $t \in \mathbb{R}$. The next lemma provides information about the relation between $\ell_p(F,G)$ and $\ell_p(F^0,G^0)$ for $p \geq 1$ and will be needed in Section 5.5 for the case $p = 2$.

Lemma 5.21. *Given $p \geq 1$, distributions $F,G \in \mathscr{P}^p(\mathbb{R})$ with mean values μ_F, μ_G and a $Unif(0,1)$ random variable U, it holds true that*

$$\ell_p(F^0,G^0) \;=\; \|(F^{-1}(U) - \mu_F) - (G^{-1}(U) - \mu_G)\|_p, \tag{5.39}$$

$$\ell_p(F,G) \;=\; \|((F^0)^{-1}(U) + \mu_F) - ((G^0)^{-1}(U) + \mu_G)\|_p, \tag{5.40}$$

and therefore

$$|\ell_p(F,G) - \ell_p(F^0,G^0)| \;\leq\; |\mu_F - \mu_G|. \tag{5.41}$$

If $p = 2$, then furthermore

$$\ell_2^2(F,G) \;=\; \ell_2^2(F^0,G^0) + (\mu_F - \mu_G)^2. \tag{5.42}$$

Proof. For (5.39) and (5.40), it suffices to note that $F^0(t) = F(t + \mu_F)$ obviously implies $(F^0)^{-1}(t) = F^{-1}(t) - \mu_F$ for all $t \in \mathbb{R}$. If $p = 2$, then (5.40) with $X := (F^0)^{-1}(U)$ and $Y := (G^0)^{-1}(U)$ yields

$$\begin{aligned}\ell_2^2(F,G) &= \mathbb{E}\big((X-Y)+(\mu_F-\mu_G)\big)^2\\ &= \mathbb{E}(X-Y)^2+2(\mu_F-\mu_G)\,\mathbb{E}(X-Y)+(\mu_F-\mu_G)^2\\ &= \ell_2^2(F^0,G^0)+(\mu_F-\mu_G)^2,\end{aligned}$$

where $\mathbb{E}X=\mathbb{E}Y=0$ has been utilized. □

We proceed to a result that characterizes convergence with respect to ℓ_p ($\stackrel{\ell_p}{\longrightarrow}$).

Proposition 5.22. *Let $p>0$ and $(F_n)_{n\ge 0}$ be a sequence of distributions in $\mathscr{P}^p(\mathbb{R})$. Then the following assertions are equivalent:*

(a) $F_n \stackrel{\ell_p}{\longrightarrow} F$, *i.e.* $\lim_{n\to\infty}\ell_p(F_n,F)=0$.
(b) $F_n \stackrel{w}{\to} F$ *and* $\lim_{n\to\infty}\int |x|^p\,F_n(dx)=\int |x|^p\,F(dx)<\infty$.
(c) $F_n \stackrel{w}{\to} F$ *and* $x\mapsto |x|^p$ *is ui with respect to the* F_n, *that is*

$$\lim_{a\to\infty}\sup_{n\ge 1}\int_{(-a,a)^c}|x|^p\,F_n(dx) \;=\; 0.$$

Proof. As before, let U denote a Unif$\,(0,1)$ random variable. By Skorohod's representation theorem [☞ e.g. [16, Thm. 25.6]] $F_n \stackrel{w}{\to} F$ holds iff $F_n^{-1}(U)\to F^{-1}(U)$ a.s.. Therefore we may restate (b) and (c) as follows:

(b) $F_n^{-1}(U)\to F^{-1}(U)$ a.s. and $\mathbb{E}|F_n^{-1}(U)|^p\to\mathbb{E}|F^{-1}(U)|^p$.
(c) $F_n^{-1}(U)\to F^{-1}(U)$ a.s. and $(|F_n^{-1}(U)|^p)_{n\ge 0}$ is ui.

The asserted equivalences are now directly inferred with the help of [28, Thm. 4.2.3(i) and Cor. 4.2.5]. □

The proposition particularly states that ℓ_p metrizes weak convergence in combination with convergence of absolute moments of order p. Two straightforward consequences are given in two subsequent corollaries.

Corollary 5.23. *If $F_n \stackrel{\ell_p}{\longrightarrow} F$ for some $p>0$, then*

$$\lim_{n\to\infty}\int |x|^q\,F_n(dx) \;=\; \int |x|^q\,F(dx)$$

for each $q\in(0,p]$. Moreover, if $p\ge 1$, then

$$\lim_{n\to\infty}\int x^k\,F_n(dx) \;=\; \int x^k\,F(dx)$$

for each integral $k\in(0,p]$.

Proof. Problem 5.31. □

For the second corollary, note that $\|F_n^{-1}(U)-F^{-1}(U)\|_q \le \|F_n^{-1}(U)-F^{-1}(U)\|_p$ if $1 \le q \le p$, and $\|F_n^{-1}(U)-F^{-1}(U)\|_q \le \|F_n^{-1}(U)-F^{-1}(U)\|_p^{q/p}$ if $0 < q < p \le 1$ [☞ Problem 5.27(b)]. These immediately show:

Corollary 5.24. *If $p > 0$ and $F_n \xrightarrow{\ell_p} F$, then $F_n \xrightarrow{\ell_q} F$ for any $q \in (0,p)$.*

Concerning completeness of the metric space $(\mathscr{P}^p(\mathbb{R}), \ell_p)$, we finally note:

Proposition 5.25. *For each $p > 0$, the space $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ is complete.*

Proof. Given a Cauchy sequence $(F_n)_{n\ge 1}$ in $(\mathscr{P}^p(\mathbb{R}), \ell_p)$, we have that

$$\lim_{m,n\to\infty} \ell_p(F_m, F_n) \;=\; \lim_{m,n\to\infty} \|F_m^{-1}(U) - F_n^{-1}(U)\|_p \;=\; 0$$

for any Unif$(0,1)$ variable U on a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$. Since $L^p(\mathbb{P})$ is complete (modulo a.s. equality), we infer the convergence of $F_n^{-1}(U)$ in $L^p(\mathbb{P})$ as well as in probability to a random variable $X \in L^p(\mathbb{P})$ with distribution F, say, and thereupon $F_n \xrightarrow{w} F$. But the latter further implies $F_n^{-1}(U) \to F^{-1}(U)$ a.s. and thus $X = F^{-1}(U)$ a.s. Finally,

$$\lim_{n\to\infty} \ell_p(F_n, F) \;=\; \lim_{n\to\infty} \|F_n^{-1}(U) - F^{-1}(U)\|_p \;=\; \lim_{n\to\infty} \|F_n^{-1}(U) - X\|_p \;=\; 0,$$

which completes the proof. □

In the literature, the minimal L^p-metric often appears in connection with the so-called *Wasserstein-metric* or *Vasershtein-metric*[2], named after the Russian mathematician L.N. VASERSHTEIN who introduced the concept in 1969 in [109]. The name was coined one year later by DOBRUSHIN in [35]. We close this section with a brief discussion of the connection with ℓ_p.

Let $(\mathbb{X}, \rho)$ be a metric space with Borel σ-field $\mathscr{B}(\mathbb{X})$ and let $\mathscr{P}(\mathbb{X})$ be the set of probability measures on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$. Then

$$d_\rho(F,G) \;:=\; \inf_{(X,Y)\sim(F,G)} \mathbb{E}\rho(X,Y), \quad F,G \in \mathscr{P}(\mathbb{X}) \tag{5.43}$$

defines a distance function on $\mathscr{P}(\mathbb{X})$ which, however, may be infinite. We note the following result:

[2] Publications in English language mostly use the German spelling "Wasserstein" (attributed to the name "Vasershtein" being of Germanic origin).

Proposition 5.26. *Let F be a probability measure on $(\mathbb{X},\mathscr{B}(\mathbb{X}))$ and $\mathscr{U}(F):=\{G\in\mathscr{P}(\mathbb{X}):d_\rho(F,G)<\infty\}$. Then $(\mathscr{U}(F),d_\rho)$ forms a metric space which is complete if $(\mathbb{X},\rho)$ is separable. The metric d_ρ is called* ***Wasserstein metric***.

Proof. The reader is referred to [35, Thm. 2]. □

If $(\mathbb{X},\rho)=(\mathbb{R},|\cdot|^p)$ for $0<p\le 1$, then the Wasserstein metric d_ρ coincides exactly with the minimal L^p-metric ℓ_p, whereas for $p>1$ we have $d_\rho=\ell_p^p$ when keeping the definition of ρ, i.e. $\rho(x,y)=|x-y|^p$, and the definition of d_ρ in (5.43). However, it should be observed that ρ is no longer a metric on $\mathbb{R}$ for $p>1$. Some further historical background information may found in [100].

Problems

Problem 5.27. Let $p>0$, $(\Omega,\mathfrak{A},\mathbb{P})$ be a probability space and U a *Unif*$(0,1)$ random variable.

(a) For $0<p<1$, show that ℓ_p in (5.30) is a complete (pseudo-)metric on $L^p(\mathbb{P})$, but that $\ell_p(X,0)=\|X\|_p$ fails to be a norm.

(b) Show that $\|X-Y\|_q\le\|X-Y\|_p^{q/p}$ holds true for $0<q<p<1$ and $X,Y\in L^p(\mathbb{P})$.

(c) Use $\ell_p(F,G)=\|F^{-1}(U)-G^{-1}(U)\|_p$ to verify that ℓ_p is indeed a metric on $\mathscr{P}^p(\mathbb{R})$.

Problem 5.28. Prove Lemma 5.17.

Problem 5.29. For $F\in\mathscr{P}^1(\mathbb{R})$ and $c\in\mathbb{R}$, let F^c denote the translation of F with mean value c, viz. $F^c(t):=F^0(t-c)$ for $t\in\mathbb{R}$. Use Lemma 5.21 to show that

$$\ell_p(F^c,G^c)=\ell_p(F^0,G^0)$$

for all $p\ge 1$, $F,G\in\mathscr{P}^p(\mathbb{R})$ and $c\in\mathbb{R}$.

Problem 5.30. (Generalization) Let $p>0$ and $F,G\in\mathscr{P}^p(\mathbb{R})$. Prove that

$$\ell_p(F*H,G*H)\ \le\ \ell_p(F,G),$$

for every $H\in\mathscr{P}^p(\mathbb{R})$ and that equality holds if $H=\delta_a$ for some $a\in\mathbb{R}$, i.e.

$$\ell_p(F*\delta_a,G*\delta_a)\ =\ \ell_p(F,G)$$

for all $a\in\mathbb{R}$. This obviously generalizes the result in the previous problem.

Problem 5.31. Prove Corollary 5.23.

Problem 5.32. Let $\mathscr{P}^\infty(\mathbb{R})$ denote the set of distributions on $\mathbb{R}$ with compact support. Further, let U be a $Unif(0,1)$ random variable on a probability space $(\Omega,\mathfrak{A},\mathbb{P})$ and $\|\cdot\|_\infty$ the usual L^∞-norm (essential supremum) on the vector space of $\mathbb{P}$-a.s. bounded random variables. Prove the following assertions:

(a) The mapping $\ell_\infty : \mathscr{P}^\infty(\mathbb{R})\times\mathscr{P}^\infty(\mathbb{R})\to\mathbb{R}_\geq$, $(F,G)\mapsto\|F^{-1}(U)-G^{-1}(U)\|_\infty$, defines a complete metric on $\mathscr{P}^\infty(\mathbb{R})$. [Hint: Use that $\|X\|_p\uparrow\|X\|_\infty$ as $p\uparrow\infty$ for any $X\in L^\infty(\mathbb{P})$.]

(b) Equivalence of

(i) $\lim_{n\to\infty}\ell_\infty(F_n,F)=0$;

(ii) $F_n\stackrel{w}{\to}F$ and $F\in\mathscr{P}^\infty(\mathbb{R})$;

(iii) $F_n\stackrel{w}{\to}F$ and $F_n(K)=1$ for all $n\geq 1$ and a compact $K\subset\mathbb{R}$.

holds true.

5.4 Conditions for $\mathscr{S}$ to be a self-map of $\mathscr{P}^p(\mathbb{R})$

In order to study the contractive behavior of $\mathscr{S}$ on $\mathscr{P}^p(\mathbb{R})$ for $p>0$, we must first provide conditions that ensure that $\mathscr{S}$ is a self-map on this subset of distributions on $\mathbb{R}$. In other words, we need conditions on $(C,T)=(C,(T_i)_{i\geq1})$ such that

$$\sum_{i\geq1}T_iX_i+C\ \in\ L^p$$

whenever the iid $X_1,X_2,\ldots$ are in L^p. Choosing $X_1=X_2=\ldots=0$, we see that $C\in L^p$ is necessary, so that we are left with the problem of finding conditions on T such that $\sum_{i\geq1}T_iX_i\in L^p$ if this is true for the X_i. The main result is stated as Proposition 5.33 below and does not need that $N=\sum_{i\geq1}\mathbf{1}_{\{T_i\neq0\}}$ is a.s. finite. We therefore remark that $\sum_{i\geq1}T_iX_i\in L^p$ is generally to be understood in the sense of L^p-convergence of the finite partial sums $\sum_{i=1}^n T_iX_i$, which particularly implies convergence in probability. Before stating the result let us define

$$\mathscr{P}^p_c(\mathbb{R})\ :=\ \left\{F\in\mathscr{P}^p(\mathbb{R}):\int x\,F(dx)=c\right\}$$

and also $L^p_c:=\{X\in L^p:\mathbb{E}X=c\}$ for $p\geq1$ and $c\in\mathbb{R}$.

Proposition 5.33. *Let $T=(T_i)_{i\geq1}$ and $(X_i)_{i\geq1}$ be independent sequences on a given probability space $(\Omega,\mathfrak{A},\mathbb{P})$ such that $X_1,X_2,\ldots$ are iid and in L^p. Then each of the following set of conditions implies $\sum_{i\geq1}T_iX_i\in L^p$:*

(i) $0<p\leq1$ *and* $\sum_{i\geq1}|T_i|^p\in L^1$.

(ii) $1<p\leq2$, $\sum_{i\geq1}T_i\in L^p$ *and* $\sum_{i\geq1}|T_i|^p\in L^1$.

(iii) $2 \le p < \infty$, $\sum_{i\ge1} T_i \in L^p$ *and* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.
(iv) $1 < p \le 2$, $\sum_{i\ge1} |T_i|^p \in L^1$ *and* $\mathbb{E}X_1 = 0$.
(v) $2 \le p < \infty$, $\sum_{i\ge1} T_i^2 \in L^{p/2}$ *and* $\mathbb{E}X_1 = 0$.

Conversely, if $1 < p < \infty$, *then*

(a) $\sum_{i\ge1} T_iX_i \in L^p$ *for any choice of* T*-independent and iid* $X_1, X_2, \ldots$ *in* L^p *implies* $\sum_{i\ge1} T_i \in L^p$ *and* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.
(b) $\sum_{i\ge1} T_iX_i \in L^p$ *for any choice of* T*-independent and iid* $X_1, X_2, \ldots \in L_0^p$ *implies* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.

Remark 5.34. The reader should observe that, in view of (iii) and (v), the implications in the converse parts (a) and (b) are in fact equivalences. It is tacitly understood there that the underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ is rich enough to carry T-independent iid $X_1, X_2, \ldots$ with arbitrary distribution in $\mathscr{P}^p(\mathbb{R})$, which is obviously the case if it carries a sequence of iid $\mathit{Unif}(0,1)$ variables. Our proof will show that it is even enough if there exist T-independent iid $X_1, X_2, \ldots$ taking values ±1 with probability $1/2$ each.

Proof. (i) If $0 < p \le 1$, the subadditivity of $x \mapsto x^p$ for $x \ge 0$ immediately implies under the given assumptions that

$$\mathbb{E}\left(\sum_{i\ge1} |T_iX_i|\right)^p \;\le\; \sum_{i\ge1} \mathbb{E}|T_iX_i|^p \;=\; \mathbb{E}|X_1|^p \sum_{i\ge1} \mathbb{E}|T_i|^p \;<\; \infty$$

and thus the almost sure absolute convergence of $\sum_{i\ge1} T_iX_i$ as well as its integrability of order p.

(ii) Here we argue that $(\sum_{i=1}^n T_iX_i)_{n\ge1}$ forms a Cauchy sequence in $(L^p(\mathbb{P}), \|\cdot\|_p)$ and is therefore L^p-convergent. First note that $\mathbb{E}(\sum_{i\ge1} |T_i|^p) = \sum_{i\ge1} \mathbb{E}|T_i|^p$ implies $T_i \in L^p$ for each $i \ge 1$, which in combination with $X_i \in L^p$ for each $i \ge 1$ ensures that $\sum_{i=m}^n T_iX_i \in L^p$ for all $n \ge m \ge 1$. Denoting by μ the expectation of the X_i, we have that $(\sum_{i=m}^k T_i(X_i - \mu))_{m\le k\le n}$ conditioned upon T forms an L^p-martingale, for T and $(X_i)_{i\ge1}$ are independent. Since $1 < p \le 2$, the even function $x \mapsto |x|^p$ is convex with concave derivative on $\mathbb{R}_\ge$ which allows us to make use of the Topchiĭ-Vatutin inequality B.1 in the Appendix. This yields

$$\mathbb{E}\left(\left|\sum_{i=m}^n T_i(X_i - \mu)\right|^p \middle| T\right) \;\le\; 2\,\mathbb{E}|X_1 - \mu|^p \sum_{i=m}^n |T_i|^p \quad \text{a.s.}$$

and then by taking unconditional expectations

$$\left\|\sum_{i=m}^n T_i(X_i - \mu)\right\|_p \;\le\; 2\,\|X_1 - \mu\|_p \left\|\sum_{i=m}^n |T_i|^p\right\|_1^{1/p}.$$

Since $\sum_{i\ge 1}|T_i|^p \in L^1$, the right-hand side converges to zero as $m,n\to\infty$. By using the second assumption $\sum_{i\ge 1} T_i \in L^p$, we infer that $\lim_{m,n\to\infty}\|\sum_{i=m}^n T_i\|_p = 0$ as well, whence finally

$$\left\|\sum_{i=m}^n T_i X_i\right\|_p \;\le\; \left\|\sum_{i=m}^n T_i(X_i-\mu)\right\|_p + |\mu|\left\|\sum_{i=m}^n T_i\right\|_p \;\to\; 0 \tag{5.44}$$

as $m,n\to\infty$.

(iii) Here we use the same Cauchy sequence argument as in (ii), but make use of the famous Burkholder inequality B.4 in the Appendix.. This yields

$$\mathbb{E}\left(\left|\sum_{i=m}^n T_i(X_i-\mu)\right|^p \,\middle|\, T\right) \;\le\; b_p^p\,\mathbb{E}\left(\left(\sum_{i=m}^n T_i^2(X_i-\mu)^2\right)^{p/2} \,\middle|\, T\right) \quad \text{a.s.}$$

for a constant $b_p \in \mathbb{R}_>$ which only depends on p. Next, put $\Sigma_{m:n}^2 := \sum_{i=m}^n T_i^2$ for $n\ge m\ge 1$. Given T and $\Sigma_{m:n}\ne 0$, the vector

$$\left(\frac{T_m^2}{\Sigma_{m:n}^2},\ldots,\frac{T_n^2}{\Sigma_{m:n}^2}\right)$$

defines a discrete probability distribution on $\{m,\ldots,n\}$, which in combination with the independence of T and $(X_i)_{i\ge 1}$, the convexity of $x\mapsto x^{p/2}$ for $x\ge 0$ and $p\ge 2$ and an appeal to Jensen's inequality yields

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{i=m}^n T_i^2(X_i-\mu)^2\right)^{p/2} \,\middle|\, T\right) \;&=\; \mathbb{E}\left(\left(\sum_{i=m}^n \frac{T_i^2}{\Sigma_{m:n}^2}\Sigma_{m:n}^2(X_i-\mu)^2\right)^{p/2} \,\middle|\, T\right)\\
&\le\; \mathbb{E}\left(\sum_{i=m}^n \frac{T_i^2}{\Sigma_{m:n}^2}\Sigma_{m:n}^p|X_i-\mu|^p \,\middle|\, T\right)\\
&=\; \left(\Sigma_{m:n}^p\sum_{i=m}^n\frac{T_i^2}{\Sigma_{m:n}^2}\right)\mathbb{E}|X_1-\mu|^p\\
&=\; \Sigma_{m:n}^p\,\mathbb{E}|X_1-\mu|^p \quad \text{a.s. on } \{\Sigma_{m:n}>0\}.
\end{aligned}$$

But if $\Sigma_{m:n}=0$, the inequality is trivially satisfied. Since, by assumption, $\mathbb{E}\Sigma_{m,n}^p\to 0$ as $m,n\to\infty$, we now obtain by taking unconditional expectations and letting m,n tend to infinity that

$$\lim_{m,n\to\infty}\mathbb{E}\left|\sum_{i=m}^n T_i(X_i-\mu)\right|^p \;\le\; b_p^p\,\mathbb{E}|X_1-\mu|^p\lim_{m,n\to\infty}\mathbb{E}\Sigma_{m,n}^p \;=\; 0.$$

The remaining argument via (5.44) is identical to the one in the previous case and thus not repeated here.

(iv), (v) If $\mu = \mathbb{E}X_1 = 0$, the assumption in $\sum_{i\ge1} T_i \in L^p$ can be dropped because then the second term on the right-hand side in (5.44) vanishes.

The converse part:

(a) By choosing $X_i = 1$ for $i \ge 1$, we find that $\sum_{i\ge1} T_i \in L^p$ and are thus left with a proof of $\sum_{i\ge1} T_i^2 \in L^{p/2}$. Let now $X_1, X_2, \ldots$ be iid random variables taking values ±1 with probability $1/2$ each. Then $\mathbb{E}X_1 = 0$, $X_1 \in L^p$ for any $p > 1$, and $(\sum_{i=1}^n T_i X_i)_{n\ge0}$ conditioned on T forms a L^p-bounded martingale. By another appeal to Burkholder's inequality B.4 (lower bound) and observing $X_1^2 = 1$, it follows that

$$\mathbb{E}\left(\left|\sum_{i=1}^n T_i X_i\right|^p \middle| T\right) \ \ge\ a_p^p \left(\sum_{i=1}^n T_i^2\right)^{p/2} \quad \text{a.s.}$$

for a constant $a_p \in \mathbb{R}_>$ which only depends on p. Consequently,

$$\mathbb{E}\left(\sum_{i\ge1} T_i^2\right)^{p/2} \ \le\ \frac{1}{a_p^p}\,\mathbb{E}\left|\sum_{i\ge1} T_i X_i\right|^p \ <\ \infty$$

which proves the remaining assertion.

(b) Here it suffices to refer to the last argument. □

In the following, we say that *the smoothing transform $\mathscr{S}$ exists in L^p-sense* if $\mathscr{S}$ is a self-map on $\mathscr{P}^p(\mathbb{R})$. Then we obtain as a direct consequence of Proposition 5.33:

Corollary 5.35. *The smoothing transform $\mathscr{S}$ exists*

- *in L^p-sense for $0 < p \le 1$ if $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *in L^p-sense for $1 < p < 2$ if C, $\sum_{i\ge1} T_i \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ for $1 < p < 2$ if $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ for $1 < p \le 2$ if $C \in L_0^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *in L^p-sense for $2 \le p < \infty$ iff $C, \sum_{i\ge1} T_i \in L^p$ and $\sum_{i\ge1} T_i^2 \in L^{p/2}$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ for $2 \le p < \infty$ iff $C \in L_0^p$ and $\sum_{i\ge1} T_i^2 \in L^{p/2}$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ for $2 \le p < \infty$ iff $C \in L^p$ and $\sum_{i\ge1} T_i^2 \in L^{p/2}$.*

Conversely, if $\mathscr{S}$ exists

- *in L^p-sense for $1 < p < 2$, then $C, \sum_{i\ge1} T_i \in L^p$ and $\sum_{i\ge1} T_i^2 \in L^{p/2}$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ for $1 < p < 2$, then $C \in L_0^p$ and $\sum_{i\ge1} T_i^2 \in L^{p/2}$.*

Proof. Problem 5.39. □

In the particularly important case when $T_1, T_2, \ldots$ are nonnegative, a necessary and sufficient condition for $\mathscr{S}$ to exist in L^p-sense can be given for *all* $p \geq 1$ and follows directly from the previous result.

Corollary 5.36. *Let $T_1, T_2, \ldots$ be nonnegative and $1 \leq p < \infty$. Then the smoothing transform $\mathscr{S}$ exists in L^p-sense iff $C, \sum_{i\geq 1} T_i \in L^p$.*

Proof. Problem 5.39. □

Problems

Problem 5.37. Consider the situation of Proposition 5.33 with $p = 2$ and show by direct computation that

$$\mathbb{E}\left(\sum_{i=m}^{n} T_i X_i\right)^2 = \mathbb{V}\mathrm{ar} X_1 \, \mathbb{E}\left(\sum_{i=m}^{n} T_i^2\right) + \left(\mathbb{E}X_1\right)^2 \mathbb{E}\left(\sum_{i=m}^{n} T_i\right)^2$$

for all $n \geq m \geq 1$. Use this to infer that $(\sum_{i=1}^{n} T_i X_i)_{n\geq 1}$ forms a Cauchy sequence in $(L^2(\mathbb{P}), \|\cdot\|_2)$.

Problem 5.38. Again in the situation of Proposition 5.33, prove the following assertions for $1 < p < \infty$:

(a) If $\sum_{i\geq 1} |T_i| \in L^p$, then $\sum_{i\geq 1} T_i \in L^p$ and $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.
(b) If $0 < p \leq 2$, then $\sum_{i\geq 1} |T_i|^p \in L^1$ implies $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.
(c) If $2 \leq p < \infty$, then $\sum_{i\geq 1} T_i^2 \in L^{p/2}$ implies $\sum_{i\geq 1} |T_i|^p \in L^1$.

Problem 5.39. Prove Corollaries 5.35 and 5.36.

Problem 5.40. Give conditions on (C, T) which entail that $\mathscr{S}$ is a self-map of $\mathscr{P}^2_{0,1}(\mathbb{R})$, the set of distributions on $\mathbb{R}$ with mean zero and variance one.

Problem 5.41. Conditions on (C, T) for $\mathscr{S}$ to be a self-map may also be studied for other subsets of $\mathscr{P}(\mathbb{R})$ which are not characterized by moments. Do so for the following subsets:

(a) $\mathscr{P}_{cont}(\mathbb{R})$, the set of continuous distributions.
(b) $\mathscr{P}_{ac}(\mathbb{R})$, the set of absolutely continuous distributions.
(c) $\mathscr{P}(\mathbb{Z})$, the set of discrete distributions concentrated on $\mathbb{Z}$.
(d) $\mathscr{P}_{symm}(\mathbb{R})$, the set of symmetric distributions F satisfying $F(B) = F(-B)$ for all $B \in \mathscr{B}(\mathbb{R})$.
(e) the set of normal distributions.
(f) the set of Poisson distributions.

(g) the set of Cauchy distributions.
(h) the set of *linear fractional distributions*

$$F_{b,p} \;=\; \frac{1-b-p}{1-p}\,\delta_0+\sum_{n\geq 1} bp^n\,\delta_n$$

for $b,p\in(0,1)$ with $b+p\leq 1$.

Problem 5.42. Prove that $\mathscr{S}$ is a self-map of $\mathscr{P}(\mathbb{R})$ iff (A0), i.e. $N<\infty$ a.s., holds true. [Hint: Use *Kolmogorov's three series theorem* which states that, given independent random variables $X_1,X_2,\ldots$, the series $\sum_{i\geq 1} X_i$ is a.s. convergent iff the following three conditions hold true:

(TST-1) $\sum_{i\geq 1}\mathbb{P}(|X_i|>1)<\infty$.
(TST-2) the series $\sum_{i\geq 1}\mathbb{E}X_i\mathbf{1}_{\{|X_i|\leq 1\}}$ is a.s. convergent.
(TST-3) $\sum_{i\geq 1}\mathbb{V}\mathrm{ar}\,(X_i\mathbf{1}_{\{|X_i|\leq 1\}})<\infty$.

A proof of this result can be found e.g. in [28, Thm. 5.1.2]].

Problem 5.43. Prove that, if $\mathscr{S}$ is a self-map on $\mathscr{P}(\mathbb{R})$, then it is weakly continuous, that is, $F_n\stackrel{w}{\to}F$ implies $\mathscr{S}(F_n)\stackrel{w}{\to}\mathscr{S}(F)$.

5.5 Contraction conditions for $\mathscr{S}$

We now turn to the central question under which conditions on (C,T) the smoothing transform $\mathscr{S}$ is a contraction or quasi-contraction on $\mathscr{P}^p(\mathbb{R})$, or a subset thereof like $\mathscr{P}_0^p(\mathbb{R})$, and therefore possesses a unique geometrically attracting fixed point on this set by Banach's fixed point theorem [☞ Appendix A]. More precisely, we will study the problem of finding conditions which ensure (geometric) convergence of the sequence $(\mathscr{S}^n(F))_{n\geq 0}$ for $F\in\mathscr{P}^p(\mathbb{R})$. Since the behavior of $\mathscr{S}$ is different for $0<p\leq 1$, $1<p\leq 2$ and $p>2$, these cases will be treated separately.

5.5.1 Convergence of iterated mean values

By Theorem C.5 in the Appendix, the convergence of $\mathscr{S}^n(F)$ to a fixed point in $(\mathscr{P}^p(\mathbb{R}),\ell_p)$ follows if $\mathscr{S}$ is a continuous self-map of this space and

$$\ell_p(\mathscr{S}^{n+1}(F),\mathscr{S}^n(F))\;\leq\;c\,\alpha^n \tag{5.45}$$

holds true for suitable $c\geq 0$, $\alpha\in[0,1)$ and all $n\geq 0$. A contraction lemma will ensure the continuity of $\mathscr{S}$ in each of the cases mentioned above, but in order to infer uniqueness of the fixed point we must consider expected values if $p\geq 1$, which provides the motivation behind the subsequent lemma [☞ [97, Lemma 1]]. In slight abuse of language, we will use the convenient notation $\mathbb{E}F := \int x\,F(dx)$ for the mean value of a distribution $F \in \mathscr{P}^1(\mathbb{R})$.

Lemma 5.44. *Suppose that $\mathscr{S}$ exists in L^p-sense for some $p \ge 1$ and let $F \in \mathscr{P}^p(\mathbb{R})$. Then*

(a) $\mathbb{E}(\sum_{i\ge1} T_i) \in (-1,1)$ *implies*

$$\lim_{n\to\infty} \mathbb{E}\mathscr{S}^n(F) \;=\; \frac{\mathbb{E}C}{1-\mathbb{E}(\sum_{i\ge1} T_i)},$$

and the convergence rate is geometric.

(b) $|\mathbb{E}(\sum_{i\ge1} T_i)| > 1$ *and* $\mathbb{E}F + (\mathbb{E}(\sum_{i\ge1} T_i) - 1)^{-1}\,\mathbb{E}C \ne 0$ *imply*

$$\lim_{n\to\infty} |\mathbb{E}\mathscr{S}^n(F)| \;=\; \infty.$$

(c) $|\mathbb{E}(\sum_{i\ge1} T_i)| > 1$ *and* $\mathbb{E}F + (\mathbb{E}(\sum_{i\ge1} T_i) - 1)^{-1}\,\mathbb{E}C = 0$ *imply*

$$\lim_{n\to\infty} \mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}F \;=\; \frac{\mathbb{E}C}{1-\mathbb{E}(\sum_{i\ge1} T_i)}.$$

(d) $\mathbb{E}(\sum_{i\ge1} T_i) = 1$ *and* $\mathbb{E}C \ne 0$ *imply*

$$\lim_{n\to\infty} |\mathbb{E}\mathscr{S}^n(F)| \;=\; \infty.$$

(e) $\mathbb{E}(\sum_{i\ge1} T_i) = 1$ *and* $\mathbb{E}C = 0$ *imply* $\mathbb{E}\mathscr{S}^n(F) = \mathbb{E}F$ *for all* $n \ge 0$.

(f) $\mathbb{E}(\sum_{i\ge1} T_i) = -1$ *implies* $\mathbb{E}S^{2n}(F) = \mathbb{E}F$ *and* $\mathbb{E}\mathscr{S}^{2n+1} = \mathbb{E}C - \mathbb{E}F$ *for all* $n \ge 0$.

Proof. Fix any $n \ge 1$ and let (C,T), $X_1, X_2, \ldots$ be independent such that $\mathscr{L}(X_i) = \mathscr{S}^{n-1}(F)$ for each $i \ge 1$. Since $\sum_{i\ge1} T_i \in L^1$ by Corollary 5.35, we infer upon setting $\beta := \mathbb{E}(\sum_{i\ge1} T_i)$ that

$$\mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}C + \mathbb{E}\left(\sum_{i\ge1} T_i X_i\right) \;=\; \mathbb{E}C + \beta\,\mathbb{E}X_1 \;=\; \mathbb{E}C + \beta\,\mathbb{E}\mathscr{S}^{n-1}(F) \qquad (5.46)$$

and then inductively

$$\mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}C \sum_{k=0}^{n-1} \beta^k + \beta^n \mathbb{E}F.$$

All assertions are easily derived from this equation, which is left as an exercise to the reader [☞ Problem 5.57]. □

5.5.2 Contraction conditions if $0 < p \le 1$

We are now ready to derive contraction results for $\mathscr{S}$ on the space $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and begin with the simplest case when $0 < p \le 1$.

Theorem 5.45. *Let $0 < p \le 1$. If*

$$C \in L^p \quad \textit{and} \quad \left\| \sum_{i\ge1} |T_i|^p \right\|_1 < 1,$$

then $\mathscr{S}$ defines a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

The proof of this result is furnished by the following contraction lemma.

Lemma 5.46. *Let $0 < p \le 1$, $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$. Then*

$$\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \ \le \ \left\| \sum_{i\ge1} |T_i|^p \right\|_1 \ell_p(F,G) \tag{5.47}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$.

Proof. Pick any $F, G \in \mathscr{P}^p(\mathbb{R})$ and let $(X_1,Y_1),(X_2,Y_2),\ldots$ be iid and (C,T)-independent random variables with $\mathscr{L}(X_1) = F$, $\mathscr{L}(Y_1) = G$ and $\|X_1 - Y_1\|_p = \ell_p(F,G)$. We note that $\mathscr{S}$ exists in L^p-sense by Corollary 5.35. Since $x \mapsto x^p$ is subadditive for $x \ge 0$ and $(\sum_{i\ge1} T_i X_i + C, \sum_{i\ge1} T_i Y_i + C) \sim (\mathscr{S}(F), \mathscr{S}(G))$, we infer

$$\begin{aligned}
\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \ &\le \ \left\| \sum_{i\ge1} T_i X_i - \sum_{i\ge1} T_i Y_i \right\|_p \\
&= \ \mathbb{E} \left| \sum_{i\ge1} T_i (X_i - Y_i) \right|^p \\
&\le \ \|X_1 - Y_1\|_p \, \mathbb{E} \left(\sum_{i\ge1} |T_i|^p \right) \\
&= \ \left\| \sum_{i\ge1} |T_i|^p \right\|_1 \ell_p(F,G),
\end{aligned}$$

which is the assertion. ◻

Proof (of Theorem 5.45). By virtue of the previous lemma, $\mathscr{S}$ forms an α-contraction with $\alpha := \|\sum_{i\ge1}|T_i|^p\|_1$. Hence, the assertions follow from Banach's fixed point theorem C.2 in combination with (5.47) (or Corollary C.3). □

5.5.3 Contraction conditions if $p > 1$

Having settled the case $0 < p \le 1$ with just one condition, viz. $\left\|\sum_{i\ge1}|T_i|^p\right\|_1 < 1$, giving contraction of $\mathscr{S}$ and a unique fixed point on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$, the case $1 < p < \infty$ exhibits a more complex picture as shown by three subsequent theorems, which for $p = 2$ are all due to RÖSLER [97]. The afore-mentioned contraction condition, which figured in the previous subsection, is now replaced with

$$\mathscr{C}_p(T) \;:=\; \mathbb{E}\left(\sum_{i\ge1}|T_i|^p\right) \vee \mathbb{E}\left(\sum_{i\ge1}T_i^2\right)^{p/2} \tag{5.48}$$

which is still $\left\|\sum_{i\ge1}|T_i|^p\right\|_1$ if $1 < p \le 2$, but equals $\left\|\sum_{i\ge1}T_i^2\right\|_{p/2}^{p/2}$ if $p \ge 2$ [☞ Problem 5.38(c)]. Plainly, the conditions are exactly the same if $p = 2$.

Theorem 5.47. *Let $p > 1$. If*

$$C \in L_0^p \quad \textit{and} \quad \mathscr{C}_p(T) < 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Theorem 5.48. *Let $p > 1$. If*

$$C, \sum_{i\ge1}T_i \in L^p, \quad \mathscr{C}_p(T) < 1 \quad \textit{and} \quad \left|\mathbb{E}\left(\sum_{i\ge1}T_i\right)\right| < 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Theorem 5.49. *Let $p > 1$ and $c \in \mathbb{R}$. If*

$$C \in L_0^p, \quad \sum_{i\ge1}T_i \in L^p, \quad \mathscr{C}_p(T) < 1, \quad \textit{and} \quad \mathbb{E}\left(\sum_{i\ge1}T_i\right) = 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}_c^p(\mathbb{R}),\ell_p)$ and has a unique geometrically attracting fixed point G_c in this space. Moreover, if even $\sum_{i\ge1}T_i=1$ a.s. holds true, then the G_c form a translation family, i.e. $G_c=\delta_c*G_0$ for all $c\in\mathbb{R}$.

We proceed to the derivation of two contraction lemmata, treating the cases

- $p=2$ and $\mathscr{C}_p(T)=\|\sum_{i\ge1}|T_i|^p\|_1=\|\sum_{i\ge1}T_i^2\|_{p/2}^{p/2}<1$.
- $p>1$ and $\mathscr{C}_p(T)<1$.

The proofs of the previous theorems require only the last of these lemmata, but we have included the other one because the provided contraction constant is better for $p=2$. Recall that F^0 denotes the centering of F if $F\in\mathscr{P}^1(\mathbb{R})$.

Lemma 5.50. *Assuming $C\in L^2$ and $\sum_{i\ge1}T_i^2\in L^1$, the following assertions hold true:*

(a) $\mathscr{S}$ exists from $\mathscr{P}_0^2(\mathbb{R})\to\mathscr{P}^2(\mathbb{R})$ and

$$\ell_2^2(\mathscr{S}(F^0),\mathscr{S}(G^0)) \le \left\|\sum_{i\ge1}T_i^2\right\|_1 \ell_2^2(F^0,G^0) \tag{5.49}$$

for all $F,G\in\mathscr{P}^2(\mathbb{R})$.

(b) If also $\sum_{i\ge1}T_i\in L^2$, then $\mathscr{S}$ exists in the L^2-sense and

$$\ell_2^2(\mathscr{S}(F),\mathscr{S}(G)) \le \left\|\sum_{i\ge1}T_i^2\right\|_1 \ell_2^2(F^0,G^0)+\left\|\sum_{i\ge1}T_i\right\|_2^2\big(\mathbb{E}F-\mathbb{E}G\big)^2 \tag{5.50}$$

for all $F,G\in\mathscr{P}^2(\mathbb{R})$.

Proof. In both parts, the existence of $\mathscr{S}$ in the claimed sense follows again from Corollary 5.35.

(a) Given any $F,G\in\mathscr{P}^2(\mathbb{R})$, let $(X_1,Y_1),(X_2,Y_2),\dots$ be iid random vectors independent of (C,T) and satisfying $(X_1,Y_1)\sim(F^0,G^0)$ and $\|X_1-Y_1\|_2=\ell_2(F^0,G^0)$. By a similar estimation as in the proof of Lemma 5.46, it then follows that

$$\begin{aligned}\ell_2^2(\mathscr{S}(F^0),\mathscr{S}(G^0)) &\le \left\|\sum_{i\ge1}T_iX_i-\sum_{i\ge1}T_iY_i\right\|_2^2\\ &= \mathbb{E}\left(\sum_{i\ge1}T_i(X_i-Y_i)\right)^2\end{aligned}$$

$$= \mathbb{E}(X_1 - Y_1)^2 \, \mathbb{E}\left(\sum_{i\geq 1} T_i^2\right)$$
$$= \left\|\sum_{i\geq 1} T_i^2\right\|_1 \ell_2^2(F^0, G^0),$$

where $\mathbb{E}(X_1 - Y_1) = 0$ has been utilized for the penultimate line.

(b) Keeping the notation of part (a), note that $(X_1 + \mathbb{E}F, Y_1 + \mathbb{E}G) \sim (F, G)$. Therefore, we obtain

$$\begin{aligned}
\ell_2^2(\mathscr{S}(F), \mathscr{S}(G)) &\leq \left\|\sum_{i\geq 1} T_i(X_i + \mathbb{E}F) - \sum_{i\geq 1} T_i(Y_i + \mathbb{E}G)\right\|_2^2 \\
&= \mathbb{E}\left(\sum_{i\geq 1} T_i\Big((X_i - Y_i) + (\mathbb{E}F - \mathbb{E}G)\Big)\right)^2 \\
&= \mathbb{E}(X_1 - Y_1)^2 \, \mathbb{E}\left(\sum_{i\geq 1} T_i^2\right) + (\mathbb{E}F - \mathbb{E}G)^2 \, \mathbb{E}\left(\sum_{i\geq 1} T_i\right)^2 \\
&= \left\|\sum_{i\geq 1} T_i^2\right\|_1 \ell_2^2(F^0, G^0) + \left\|\sum_{i\geq 1} T_i\right\|_2^2 (\mathbb{E}F - \mathbb{E}G)^2,
\end{aligned}$$

where again $\mathbb{E}(X_1 - Y_1) = 0$ has entered to get the penultimate line. □

The corresponding lemma for $p > 1$ is technically more difficult to prove because p^{th} powers of sums can be written out term-wise only for integral p.

Lemma 5.51. *Let $1 < p < \infty$, $C \in L^p$ and $\sum_{i\geq 1} |T_i|^p \in L^1$. Then the following assertions hold true:*

(a) $\mathscr{S}$ *exists from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ and*

$$\ell_p(\mathscr{S}^n(F^0), \mathscr{S}^n(G^0)) \leq b_p \mathscr{C}_p(T)^{n/p} \, \ell_p(F^0, G^0) \tag{5.51}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$ and $n \geq 1$.

(b) *If also $\sum_{i\geq 1} T_i \in L^p$, then $\mathscr{S}$ exists in L^p-sense and*

$$\begin{aligned}
&\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \\
&\quad \leq b_p \left[\mathscr{C}_p(T)^{n/p} \, \ell_p(F^0, G^0) + n\lambda_p \kappa_p^{n-1} \left|\mathbb{E}F - \mathbb{E}G\right|\right] &(5.52) \\
&\quad \leq b_p \left(\frac{n\lambda_p}{\kappa_p} + 2\right) \kappa_p^n \, \ell_p(F, G) &(5.53)
\end{aligned}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$ and $n \geq 1$, where

$$\kappa_p := \left|\mathbb{E}\left(\sum_{i\ge1} T_i\right)\right| \vee \mathscr{C}_p(T)^{1/p}$$
$$\textit{and} \quad \lambda_p := \left\|\sum_{i\ge1}(T_i - \mathbb{E}T_i)\right\|_p + b_p^{-1}\left\|\sum_{i\ge1} T_i\right\|_p .$$

If $1 < p \le 2$*, we can choose* $b_p = 2^{1/p}$ *in both parts.*

Proof. The existence of $\mathscr{S}$ in the claimed sense is again guaranteed by Corollary 5.35.

(a) Given any $F,G \in \mathscr{P}^p(\mathbb{R})$, let $(X(\mathsf{v}),Y(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ be a family of iid random vectors which is independent of $\boldsymbol{C}\otimes\boldsymbol{T} = (C(\mathsf{v}),T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ (having the usual meaning) and satisfies $(X(\mathsf{v}),Y(\mathsf{v})) \sim (F^0,G^0)$ and $\|X(\mathsf{v})-Y(\mathsf{v})\|_p = \ell_p(F^0,G^0)$. Consider two WBP $(Z_n')_{n\ge0}$ and $(Z_n'')_{n\ge0}$ associated with $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X} = (C(\mathsf{v}),T(\mathsf{v}),X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ and $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{Y}$, respectively, so that $\mathscr{L}(Z_n') = \mathscr{S}^n(F^0)$ and $\mathscr{L}(Z_n'') = \mathscr{S}^n(G^0)$ for each $n \ge 0$ [☞ Lemma 5.2]. Furthermore,

$$Z_n := Z_n' - Z_n'' = \sum_{|\mathsf{v}|=n} L(\mathsf{v})(X(\mathsf{v}) - Y(\mathsf{v})), \quad n \ge 0$$

defines a WBP associated with $\boldsymbol{T}\otimes\boldsymbol{X}-\boldsymbol{Y} = (T(\mathsf{v}),X(\mathsf{v})-Y(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ such that

$$\ell_p(\mathscr{S}^n(F^0),\mathscr{S}^n(G^0)) \le \|Z_n' - Z_n''\|_p = \|Z_n\|_p$$

for all $n \ge 0$, because $(Z_n',Z_n'') \sim (\mathscr{S}^n(F^0),\mathscr{S}^n(G^0))$. Write Z_n as

$$Z_n = L^p\text{-}\lim_{k\to\infty} \sum_{j=1}^{k} L(\mathsf{v}^j)(X(\mathsf{v}^j) - Y(\mathsf{v}^j))$$

for a suitable enumeration $\mathsf{v}^1,\mathsf{v}^2,\ldots$ of $\mathbb{N}^n$ and observe that, conditioned on $\boldsymbol{T}$, the right-hand sum forms an L^p-martingale in $k \ge 1$. As in the proof of Proposition 5.33, we must distinguish the cases $1 < p \le 2$ and $p \ge 2$ to complete our argument.

CASE 1: $1 < p \le 2$. Then we infer with the help of the Topchiĭ-Vatutin inequality B.1 in the Appendix that

$$\begin{aligned}
\mathbb{E}(|Z_n|^p|\boldsymbol{T}) &\le 2\lim_{k\to\infty}\sum_{j=1}^{k} |L(\mathsf{v}^j)|^p\,\mathbb{E}|X(\mathsf{v}^j) - Y(\mathsf{v}^j)|^p \\
&= 2\sum_{j\ge1} |L(\mathsf{v}^j)|^p\,\mathbb{E}|X(\mathsf{v}^j) - Y(\mathsf{v}^j)|^p \\
&= 2\,\ell_p(F^0,G^0)^p \sum_{|\mathsf{v}|=n} |L(\mathsf{v})|^p \quad \text{a.s.}
\end{aligned}$$

Since, furthermore, $\mathbb{E}(\sum_{|\mathsf{v}|=n}|L(\mathsf{v})|^p) = \|\sum_{i\geq 1}|T_i|^p\|_1^n$ [use Lemma 5.5], we obtain (5.51) by taking unconditional expectation in the previous estimation.

CASE 2: $p \geq 2$. Put $\Sigma_1^2 := \sum_{i\geq 1} T_i(\varnothing)^2$. By proceeding as in the proof of Proposition 5.33(iii), but with $X(i) - Y(i)$ instead of $X_i - \mu$ and $m = 1$, $n = \infty$, it then follows by use of Burkholder's inequality B.4 and Jensen's inequality that

$$\begin{aligned}
\mathbb{E}\left(\left|\sum_{i\geq 1} T_i(\varnothing)(X(i)-Y(i))\right|^p \middle| \boldsymbol{T}\right)
&\leq b_p^p\,\mathbb{E}\left(\left(\sum_{i\geq 1} T_i(\varnothing)^2(X(i)-Y(i))^2\right)^{p/2} \middle| \boldsymbol{T}\right)\\
&\leq b_p^p\,\Sigma_1^p\,\mathbb{E}\left(\left(\sum_{i\geq 1} \frac{T_i(\varnothing)^2}{\Sigma_1^2}(X(i)-Y(i))^2\right)^{p/2} \middle| \boldsymbol{T}\right)\\
&\leq b_p^p\,\Sigma_1^p\,\mathbb{E}|X(1)-Y(1)|^p\\
&\leq b_p^p\,\Sigma_1^p\,\ell_p^p(F^0,G^0) \quad \text{a.s.}
\end{aligned}$$

and thereby

$$\ell_p(\mathscr{S}(F^0),\mathscr{S}(G^0)) \;\leq\; \left\|\sum_{i\geq 1} T_i(X(i)-Y(i))\right\|_p \;\leq\; b_p\,\|\Sigma\|_p\,\ell_p(F^0,G^0),$$

where b_p only depends on p. This proves (5.51) for $n = 1$. But in the same manner, we obtain for general n

$$\begin{aligned}
\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G)) &\leq \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(X(\mathsf{v})-Y(\mathsf{v}))\right\|_p\\
&\leq b_p\,\|\Sigma_n\|_p\,\ell_p(F^0,G^0),
\end{aligned}$$

where $\Sigma_n^2 := \sum_{|\mathsf{v}|=n} L(\mathsf{v})^2$. Hence, the proof of (5.51) will be complete once we have shown that

$$\|\Sigma_n\|_p \leq \|\Sigma\|_p^n \tag{5.54}$$

for all $n \geq 1$. To this end put $\Sigma(\mathsf{v}) := \sum_{i\geq 1} T_i(\mathsf{v})^2$ for $\mathsf{v} \in \mathbb{T}$ and recall from (5.17) that $\mathscr{F}_k = \sigma(T(\mathsf{v}) : |\mathsf{v}| \leq k-1)$ for $k \geq 1$. Then

$$\begin{aligned}
\mathbb{E}(\Sigma_n^p|\mathscr{F}_{n-1}) &= \mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n-1} L(\mathsf{v})^2\Sigma(\mathsf{v})^2\right)^{p/2} \middle| \mathscr{F}_{n-1}\right)\\
&= \Sigma_{n-1}^p\,\mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n-1} \frac{L(\mathsf{v})^2}{\Sigma_{n-1}^2}\Sigma(\mathsf{v})^2\right)^{p/2} \middle| \mathscr{F}_{n-1}\right)
\end{aligned}$$

$$
\begin{aligned}
&\le\ \Sigma_{n-1}^{p}\,\mathbb{E}\left(\sum_{|\mathsf{v}|=n-1}\frac{L(\mathsf{v})^2}{\Sigma_{n-1}^2}\Sigma(\mathsf{v})^p\,\middle|\,\mathscr{F}_{n-1}\right)\\
&=\ \Sigma_{n-1}^{p}\,\|\Sigma\|_p^p\quad\text{a.s.}
\end{aligned}
$$

for each $n\ge 2$, which clearly gives (5.54) upon taking expectations and iteration.

(b) Let us first note that it suffices to show (5.52) because then (5.53) can be easily deduced with the help of (5.41) and the obvious inequality $|\mathbb{E}F-\mathbb{E}G|\le\ell_p(F,G)$, namely

$$
\begin{aligned}
&\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0)+n\lambda_p\kappa_p^{n-1}\left|\mathbb{E}F-\mathbb{E}G\right|\\
&\qquad\le\ \mathscr{C}_p(T)^{n/p}\,\ell_p(F,G)+\left(\frac{n\lambda_p}{\kappa_p}+1\right)\kappa_p^n\left|\mathbb{E}F-\mathbb{E}G\right|\\
&\qquad\le\ \left(\frac{n\lambda_p}{\kappa_p}+2\right)\kappa_p^n\,\ell_p(F,G)
\end{aligned}
$$

for all $F,G\in\mathscr{P}^p(\mathbb{R})$.

Similar to the proof of part (b) of the previous lemma, we obtain with the help of part (a) and Minkowski's inequality that

$$
\begin{aligned}
\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G))\ &\le\ \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\Big(\big(X(\mathsf{v})-Y(\mathsf{v})\big)+(\mathbb{E}F-\mathbb{E}G)\Big)\right\|_p\\
&\le\ \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\big(X(\mathsf{v})-Y(\mathsf{v})\big)\right\|_p+|\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\right\|_p\\
&=\ \|Z_n\|_p+|\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\right\|_p\\
&\le\ b_p\,\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0)+|\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\right\|_p
\end{aligned}\tag{5.55}
$$

for all $n\ge 1$, where b_p can be chosen as $2^{1/p}$ if $1<p\le 2$. This leaves us with the task to give an estimate for $a_n:=\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\|_p$, which will be accomplished by another martingale argument involving the Topchiĭ-Vatutin inequality if $1<p\le 2$, and the Burkholder inequality if $p\ge 2$.

CASE 1: $1<p\le 2$. We put $U(\mathsf{v}):=\sum_{i\ge 1}T_i(\mathsf{v})$, $\alpha:=\mathscr{C}_p(T)^{1/p}$, $\beta:=\mathbb{E}U(\mathsf{v})$ and $\gamma:=\|U(\mathsf{v})-\beta\|_p=\|\sum_{i\ge 1}T_i-\beta\|_p$. Since $\sum_{i\ge 1}T_i\in L^p$ and $p>1$, we have $|\beta|\le a_1<\infty$. By a similar argument as in (a), we see that $\sum_{|v|=n}L(\mathsf{v})(U(\mathsf{v})-\beta)$ conditioned on $\mathscr{F}_n$ is the limit of an L^p-martingale (use that $U(\mathsf{v})$ is independent of $\mathscr{F}_n$), whence the Topchiĭ-Vatutin inequality yields

$$\mathbb{E}\left(\left|\sum_{|v|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right|^p \middle| \mathscr{F}_n\right) \;\le\; 2\,\gamma^p \sum_{|\mathsf{v}|=n} |L(\mathsf{v})|^p \quad \text{a.s.}$$

As a consequence,

$$\begin{aligned} a_{n+1} \;&=\; \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})U(\mathsf{v})\right\|_p \\ &\le\; \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right\|_p + |\beta| a_n \\ &\le\; 2^{1/p}\gamma \left\|\sum_{|\mathsf{v}|=n} |L(\mathsf{v})|^p\right\|_1^{1/p} + |\beta| a_n \\ &=\; 2^{1/p}\gamma\alpha^n + |\beta| a_n \end{aligned} \tag{5.56}$$

for all $n \ge 1$, which leads to

$$\begin{aligned} a_{n+1} \;&\le\; 2^{1/p}\gamma \sum_{k=0}^{n-1} |\beta|^k \alpha^{n-k} + |\beta|^n a_1 \\ &\le\; (n+1)(2^{1/p}\gamma + a_1)(|\beta| \vee \alpha)^n \;=\; (n+1)2^{1/p}\lambda_p \kappa_p^n \end{aligned} \tag{5.57}$$

for all $n \ge 1$. Since this inequality trivially holds for $n = 0$, we finally obtain the asserted inequality (5.52) from (5.55) and (5.57).

CASE 2: $p \ge 2$. In this case, we obtain with the Burkholder inequality that

$$\begin{aligned} \mathbb{E}\left(\left|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right|^p \middle| \mathscr{F}_n\right) \;&\le\; b_p^p\, \mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n} L(\mathsf{v})^2 (U(\mathsf{v})-\beta)^2\right)^{p/2} \middle| \mathscr{F}_n\right) \\ &\le\; b_p^p \gamma^p \Sigma_n^p \quad \text{a.s.} \end{aligned}$$

which upon taking expectations on both sides and using (5.56) provides us with

$$\left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right\|_p \;\le\; b_p\,\gamma \|\Sigma_1\|_p^n \;=\; b_p\,\gamma\alpha^n$$

and thus [☞ also (5.56)]

$$a_{n+1} \;\le\; \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right\|_p + |\beta| a_n \;\le\; b_p\,\gamma\alpha^n + |\beta| a_n \tag{5.58}$$

for all $n \ge 1$. For the remaining arguments we can refer to the previous case. □

Now we can turn to the proofs of the theorems stated above.

Proof (of Theorem 5.47). As $\mathbb{E}C = 0$ is assumed, $\mathscr{S}$ defines a self-map of $\mathscr{P}_0^p(\mathbb{R})$ by Corollary 5.35. It is also an α-contraction on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ with $\alpha := \|\sum_{i\ge 1}(T_i)^2\|_{p/2}^{1/p}$ if $p = 2$ [by Lemma 5.50(a)], and an α_m-quasi-contraction with $\alpha_m := b_p\,\alpha^m$ for suitable $m \ge 1$ if $p > 1$ [by Lemma 5.51(a)]. Therefore, the assertion follows from Banach's fixed point theorem C.2 or its generalization C.4 in combination with the contraction inequality (5.49) or (5.51), respectively. □

Proof (of Theorem 5.48). The existence of $\mathscr{S}$ in L^p-sense follows again from Corollary 5.35, while contraction inequality (5.53) shows that $\mathscr{S}$ is a quasi-contraction on $\mathscr{P}^p(\mathbb{R})$, viz.

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; c\,\kappa^n\,\ell_p(F,G)$$

for any $\kappa \in (0, \kappa_p)$, $F, G \in \mathscr{P}^p(\mathbb{R})$, $n \ge 1$ and a suitable $c = c(\kappa) > 0$. All assertions now follow from Banach's fixed point theorem C.4 for quasi-contractions. □

Proof (of Theorem 5.49). First note that $\mathbb{E}C = 0$ and $\mathbb{E}(\sum_{i\ge 1} T_i) = 1$ entail $\mathbb{E}\mathscr{S}(F) = \mathbb{E}F = c$ for all $F \in \mathscr{P}_c^p(\mathbb{R})$. Hence, $\mathscr{S}$ is a self-map of $\mathscr{P}_c^p(\mathbb{R})$ for any $c \in \mathbb{R}$. Moreover, (5.52) simplifies to

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; b_p \left\| \sum_{i\ge 1} T_i^2 \right\|_{p/2}^{n/2} \ell_p(F,G)$$

for all $n \ge 1$ and $F, G \in \mathscr{P}_c^p(\mathbb{R})$ because $\ell_p(F,G) = \ell_p(F^0, G^0)$. Hence $\mathscr{S}$ is also a quasi-contraction on $\mathscr{P}_c^p(\mathbb{R})$ and therefore has a unique fixed point G_c by Theorem C.4. It remains to verify that $G_c = \delta_c * G_0$ in the case when $\sum_{i\ge 1} T_i = 1$ a.s. By the uniqueness property of G_c, it suffices to verify that $\mathscr{S}(\delta_c * G_0) = \delta_c * G_0$. Choose iid (C,T)-independent random variables $X_1, X_2, \ldots$ with law G_0. Then

$$\begin{aligned}
\mathscr{S}(\delta_c * G_0) \;&=\; \mathscr{L}\left(\sum_{i\ge 1} T_i(X_i + c) + C\right) \\
&=\; \mathscr{L}\left(\sum_{i\ge 1} T_i X_i + c + C\right) \\
&=\; \delta_c * \mathscr{L}\left(\sum_{i\ge 1} T_i X_i + C\right) \\
&=\; \delta_c * \mathscr{S}(G_0) \\
&=\; \delta_c * G_0
\end{aligned}$$

yields the desired conclusion. □

5.5.4 Contraction conditions if $p > 2$ and $\sum_{i\ge1}|T_i| \in L^p$

If $p > 2$ and $\sum_{i\ge1} T_i \in L^p$ is replaced by the generally stronger condition $\sum_{i\ge1}|T_i| \in L^p$, then we can trade in the contraction condition $\|\sum_{i\ge1} T_i^2\|_{p/2} < 1$ for a weaker one as the following results due to RÖSLER [97] show. However, proofs are becoming more involved. Let us define

$$\mathfrak{m}(\theta) := \mathbb{E}\left(\sum_{i\ge1}|T_i|^\theta\right)$$

and note that $\mathfrak{m}(q)\vee\mathfrak{m}(p) < 1$ for $0 < q < p < \infty$ implies $\mathfrak{m}(r) < 1$ for any $r \in [q,p]$ because $\mathfrak{m}$ is convex on $[2,p]$ [☞ Problem 5.9].

Theorem 5.52. *Let $p > 2$. If*

$$C \in L_0^p, \quad \sum_{i\ge1}|T_i| \in L^p \quad \text{and} \quad \mathfrak{m}(2)\vee\mathfrak{m}(p) < 1,$$

then $\mathscr{S}$ is a self-map of $\mathscr{P}_0^p(\mathbb{R})$ with a unique geometrically ℓ_p-attracting fixed point G_0 in this set.

Theorem 5.53. *Let $p > 2$. If*

$$C, \sum_{i\ge1}|T_i| \in L^p, \quad \mathfrak{m}(2)\vee\mathfrak{m}(p) < 1 \quad \text{and} \quad \left|\mathbb{E}\left(\sum_{i\ge1} T_i\right)\right| < 1,$$

then $\mathscr{S}$ exists in L^p-sense and has a unique geometrically ℓ_p-attracting fixed point G_0 in $(\mathscr{P}^p(\mathbb{R}),\ell_p)$.

Proof (of Theorem 5.52). Here we will proceed in a different way than before and prove that $\mathscr{S}$ is locally contractive on $(\mathscr{P}_0^p(\mathbb{R}),\ell_p)$ in the sense of Theorem C.5 [☞ (5.59) below]. In particular, we will not make use of the contraction lemma 5.51. The first step is to show the result for integral $p > 2$ (the only case actually considered by RÖSLER in [97]).

So let $2 < p \in \mathbb{N}$. We prove by induction that, for each $q \in \{1,...,p\}$, there exists $\rho_q \in (0,1)$ such that

$$\ell_q^q(\mathscr{S}^n(F),\mathscr{S}^n(G)) \le c_q\,\rho_q^n \tag{5.59}$$

for all $F,G \in \mathscr{P}_0^p(\mathbb{R})$, $n \ge 1$ and a suitable $c_q \in \mathbb{R}_>$ which may depend on F,G. Observe that this corresponds to (C.5) when choosing $F = \mathscr{S}(G)$.

Hereafter, $K \in \mathbb{R}_>$ shall denote a generic constant which may differ from line to line but does not depend on n. Recall from above that $\mathfrak{m}(2) \vee \mathfrak{m}(p) < 1$ entails $\mathfrak{m}(q) < 1$ for all $q \in [2, p]$.

If $q = 1$ or $= 2$, we may invoke Lemma 5.50 to find

$$\ell_1^2(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; \ell_2^2(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; \mathfrak{m}(2)^n\, \ell_2^2(F, G)$$

for all $n \ge 1$ and $F, G \in \mathscr{P}_0^2(\mathbb{R})$, which clearly shows (5.59) in this case. We further see that $\mathscr{S}$ forms a contraction on $(\mathscr{P}_0^2(\mathbb{R}), \ell_2)$ and hence possesses a unique fixed point G_0 in this space. Since $\mathscr{P}_0^2(\mathbb{R}) \supset \mathscr{P}_0^p(\mathbb{R})$, it follows that G_0 is also the unique fixed point in $\mathscr{P}^p(\mathbb{R})$ once (5.59) has been verified for $q = p$.

For the inductive step suppose that (5.59) holds for any $r \in \{1, ..., q-1\}$ and let $(U_i)_{i \ge 1}$ be a sequence of iid $\mathit{Unif}(0,1)$ random variables which are further independent of (C, T). Fixing any $F, G \in \mathscr{P}_0^q(\mathbb{R})$ throughout the rest of the proof, put

$$Y_{n,i} \;:=\; \mathscr{S}^n(F)^{-1}(U_i) - \mathscr{S}^n(G)^{-1}(U_i), \quad n \ge 1$$

and note that $\|Y_{n,i}\|_r = \ell_r(\mathscr{S}^n(F), \mathscr{S}^n(G))$ for all $i \ge 1, n \ge 0$ and $r \in [1, q]$. Since

$$\ell_q^q(\mathscr{S}^{n+1}(F), \mathscr{S}^{n+1}(G)) \;\le\; \mathbb{E}\left|\sum_{i \ge 1} T_i Y_{n,i}\right|^q \;\le\; \lim_{m \to \infty} \mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|\right)^q$$

we will further estimate the last expectation for arbitrary $m \in \mathbb{N}$ by making use of the multinomial formula which provides us with

$$\mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|\right)^q \;=\; \mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|^q\right) + \mathbb{E}\left(\sum_{\substack{0 \le r_1, ..., r_m < q,\\ r_1 + ... + r_m = q}} \frac{q!}{r_1! \cdot ... \cdot r_m!} \prod_{j=1}^m |T_j Y_{n,j}|^{r_j}\right).$$

The first term on the right-hand side obviously equals $\mathfrak{m}(q)\, \ell_q^q(\mathscr{S}^n(F), \mathscr{S}^n(G))$, while the second may be further computed as follows by conditioning upon T and using the fact that the $Y_{n,i}$ for any fixed n are iid:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{\substack{0 \le r_1, ..., r_m < q,\\ r_1 + ... + r_m = q}} \frac{q!}{r_1! \cdot ... \cdot r_m!} \prod_{j=1}^m |T_j Y_{n,j}|^{r_j}\right)\\
&\quad = \mathbb{E}\left(\sum_{\substack{0 \le r_1, ..., r_m < q,\\ r_1 + ... + r_m = q}} \frac{q!\, \mathbb{E}|Y_{n,1}|^{r_1} \cdot ... \cdot \mathbb{E}|Y_{n,1}|^{r_m}}{r_1! \cdot ... \cdot r_m!} \prod_{j=1}^m |T_j|^{r_j}\right)\\
&\quad = \left(\prod_{j=1}^m \ell_{r_j}^{r_j}(\mathscr{S}^n(F), \mathscr{S}^n(G))\right) \mathbb{E}\left(\sum_{\substack{0 \le r_1, ..., r_m < q,\\ r_1 + ... + r_m = q}} \frac{q!}{r_1! \cdot ... \cdot r_m!} \prod_{j=1}^m |T_j|^{r_j}\right)
\end{aligned}$$

$$\leq\ K\rho^n\,\mathbb{E}\left(\sum_{i=1}^m |T_i|\right)^q$$

where the inductive hypothesis has been utilized to give the last estimate with $\rho := \max_{1\le s\le q-1}\rho_s$. The reader should notice that the constant K is not only independent of n but of m as well. Hence, by taking the limit $m\to\infty$, we find that

$$\ell_q^q(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G))\ \leq\ \mathfrak{m}(q)\,\ell_q^q(\mathscr{S}^n(F),\mathscr{S}^n(G))+K\rho^n$$

for all $n\ge 0$ and thereupon

$$\begin{aligned}\ell_q^q(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G))\ &\leq\ \mathfrak{m}(q)^{n+1}\,\ell_q^q(F,G)+K\sum_{k=1}^n \rho^k\mathfrak{m}(q)^{n-k}\\ &\leq\ \Big(\ell_q^q(F,G)+Kn\Big)\,(\mathfrak{m}(q)\vee\rho)^{n+1}\end{aligned}$$

for all $n\ge 0$ which implies (5.59) for any $\rho_q\in(\mathfrak{m}(q)\vee\rho,1)$. By an appeal to Theorem C.5, we conclude that, for any $F\in\mathscr{P}_0^p(\mathbb{R})$, $\mathscr{S}^n(F)$ converges to a fixed point in this this set which must be unique by what has been stated above.

We turn to the second step which aims at an extension of the assertion to general $p>2$ with integer part $\widehat{p}$, say. Let $r\in\mathbb{N}$ be such that $2^r<p\le 2^{r+1}$ and $s := p/2^{r+1}\in(0,1]$. From the first part of the proof, we know that (5.59) holds true for every $q\in\{1,...,\widehat{p}\}$, and since $\ell_\alpha(\cdot,\cdot)$ is nondecreasing in α, this readily extends to all $q\in[1,\widehat{p}]$. We will show hereafter that (5.59) is also true for $q=p$ (and thus for all $q\in[1,p]$) which finally proves the theorem in full generality.

Let us introduce the following operator Δ and its k-fold iterations Δ^k: For any nonnegative random variable W define

$$\Delta W := (W-\mathbb{E}W)^2,\quad \Delta^2 W = \Big((W-\mathbb{E}W)^2-\mathbb{V}\mathrm{ar}\,W\Big)^2,\quad \text{etc.}$$

and $\Delta^0 W := W$. Naturally, $\Delta W=\infty$ is stipulated if $\mathbb{E}W=\infty$. We note that

$$\mathbb{E}\Delta^k W\ \leq\ \mathbb{E}(\Delta^{k-1}W)^2\ \leq\ 2\,\mathbb{E}(\Delta^{k-2}W)^4\ \leq ...\leq\ 2^{k-1}\mathbb{E}W^{2^k} \tag{5.60}$$

holds true for any $k\ge 1$. The reader is asked for a proof in Problem 5.60.

By repeated use of the Burkholder inequality B.4 (in the by now familiar manner after conditioning on T) and the subadditivity of $x\mapsto x^\alpha$ for $x\ge 0$ and $0<\alpha\le 1$, we now obtain

$$\left\|\sum_{i\ge 1} T_iY_{n,i}\right\|_p\ \leq\ K\,\left\|\sum_{i\ge 1} T_i^2Y_{n,i}^2\right\|_{p/2}^{1/2}$$

$$
\begin{aligned}
&\le\; K\left(\left\|\sum_{i\ge1}T_i^2(Y_{n,i}^2-\mathbb{E}Y_{n,i}^2)\right\|_{p/2}^{1/2}+\left(\mathbb{E}Y_{n,1}^2\right)^{1/2}\left\|\sum_{i\ge1}T_i^2\right\|_{p/2}^{1/2}\right)\\
&\le\; K\left(\left\|\sum_{i\ge1}T_i^4\Delta Y_{n,i}^2\right\|_{p/4}^{1/4}+\left(\mathbb{E}Y_{n,1}^2\right)^{1/2}\left\|\sum_{i\ge1}|T_i|\right\|_{p}^{1/2}\right)\\
&\;\;\vdots\\
&\le\; K\left(\left\|\sum_{i\ge1}T_i^{2^{r+1}}\Delta^r Y_{n,i}^2\right\|_{s}^{1/2^{r+1}}+\sum_{j=0}^{r-1}\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}}\left\|\sum_{i\ge1}|T_i|\right\|_{p}^{1/2}\right)\\
&\le\; K\left(\left\|\sum_{i\ge1}|T_i|^p\left(\Delta^r Y_{n,i}^2\right)^s\right\|_{1}^{1/2^{r+1}}+\sum_{j=0}^{r-1}\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}}\left\|\sum_{i\ge1}|T_i|\right\|_{p}^{1/2}\right)\\
&\le\; K\left(\left\|\Delta^r Y_{n,1}^2\right\|_s^{1/2^{r+1}}\left\|\sum_{i\ge1}|T_i|^p\right\|_{1}^{1/2^{r+1}}+\sum_{j=0}^{r-1}\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}}\left\|\sum_{i\ge1}|T_i|\right\|_{p}^{1/2}\right)
\end{aligned}
$$

for all $n\ge1$. Use (5.60), the definition of $Y_{n,1}$, and (5.59) for $\widehat{p}$ to infer

$$
\begin{aligned}
\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}} \;&\le\; \left(2^{j-1}\,\mathbb{E}Y_{n,1}^{2^{j+1}}\right)^{1/2^{j+1}} \;\le\; 2\,\|Y_{n,1}\|_{2^{j+1}}\\
&\le\; 2\,\|Y_{n,1}\|_{\widehat{p}} \;=\; 2\,\ell_{\widehat{p}}(\mathscr{S}^n(F),\mathscr{S}^n(G)) \;\le\; 2\,c_{\widehat{p}}\,\rho_{\widehat{p}}^n
\end{aligned}
$$

for any $j\in\{0,\dots,r-1\}$ and $n\ge0$. By combining this with $\|\sum_{i\ge1}|T_i|^p\|_1=\mathfrak{m}(p)<1$, the above estimation finally provides us with

$$
\ell_p(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G)) \;\le\; \left\|\sum_{i\ge1}T_iY_{n,i}\right\|_p \;\le\; K\rho^{n+1}
$$

for all $n\ge0$ and a suitable $\rho\in(0,1)$. □

Proof (of Theorem 5.53). We are now in a more comfortable situation because the bulk of work has already been carried out in the previous proof. First note that all assumptions of Theorem 5.48 with $p=2$ are fulfilled which allows us to infer the existence of a unique fixed point $G_0\in\mathscr{P}^2(\mathbb{R})$. By Lemma 5.44(a), its mean value equals $c:=\mathbb{E}G_0=(1-\beta)^{-1}\mathbb{E}C$ with $\beta:=\mathbb{E}(\sum_{i\ge1}T_i)$. The reader can easily check that, if $F\in\mathscr{P}_c^p(\mathbb{R})$, then $\mathbb{E}S^n(F)=c$ for all $n\ge0$ and that this further implies $\mathscr{S}^n(F)^c=\mathscr{S}^n(F^c)$ (recall that $F^c=F^0(\cdot-c)$) and thereupon [☞ Problem 5.29]

$$
\begin{aligned}
\ell_p(\mathscr{S}^{n+1}(F^c),\mathscr{S}^n(F^c)) \;&=\; \ell_p(\mathscr{S}^{n+1}(F)^c,\mathscr{S}^n(F)^c)\\
&=\; \ell_p(\mathscr{S}^{n+1}(F)^0,\mathscr{S}^n(F)^0)
\end{aligned}
\tag{5.61}
$$

for all $F\in\mathscr{P}^p(\mathbb{R})$ and $n\ge0$.

Now fix any $F \in \mathscr{P}^p(\mathbb{R})$, define $Y_{n,i}$ as in the previous proof, but for the pair $(\mathscr{S}(F^c), F^c)$. Then (5.59) for $q = p$ can be shown as in the previous proof, giving

$$\ell_p^p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) \;\le\; \left\|\sum_{i\ge 1} T_i Y_{n,i}\right\|_p^p \;\le\; c_p \rho_p^n$$

for all $n \ge 0$ and suitable constants $c_p \in \mathbb{R}_>$ and $\rho_p \in (0,1)$. Note further that

$$\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F) \;=\; \beta^n\,(\mathbb{E}\mathscr{S}(F) - \mathbb{E}F)$$

for all $n \ge 0$, as has been shown in the proof of Lemma 5.44 [☞ (5.46)]. By combining these fact with (5.41) and (5.61), we finally obtain

$$\begin{aligned}
\ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) &\le \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\
&= \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\
&= \ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\
&\le c_p^{1/p}\rho_p^{n/p} + \beta^n\,|\mathbb{E}\mathscr{S}(F) - \mathbb{E}F|
\end{aligned}$$

for all $n \ge 0$, that is geometric contraction of every iteration sequence in $\mathscr{P}^p(\mathbb{R})$. By invoking Theorem C.5, we conclude that G_0 is the unique geometrically ℓ_p-attracting fixed point in this set. ☐

5.5.5 A global contraction condition if $p \ge 1$

None of the previous results has provided conditions ensuring global contraction of $\mathscr{S}$ on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ if $p > 1$. We make up for this by the following theorem, again taken from [97]. It should be compared with its counterpart Theorem 5.45 for the case $0 < p \le 1$.

Theorem 5.54. *Let* $p \ge 1$. *If*

$$C \in L^p \quad \textit{and} \quad \left\|\sum_{i\ge 1} |T_i|\right\|_p < 1,$$

then $\mathscr{S}$ *is a contraction on* $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ *and has a unique geometrically attracting fixed point in this space.*

Proof. In view of Banach's fixed point theorem C.2 it suffices to prove that $\mathscr{S}$ forms a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$. Pick any $F, G \in \mathscr{P}^p(\mathbb{R})$ and then as usual iid and (C,T)-independent random variables $(X_1,Y_1),(X_2,Y_2),\ldots$ such that $(X_1,Y_1) \sim (F,G)$

and $\|X_1 - Y_1\|_p = \ell_p(F,G)$. Setting $\Sigma_n := \sum_{i=1}^n |T_i|$, it follows as in the proof of Proposition 5.33(iv) that

$$\mathbb{E}\left(\left(\sum_{i=1}^n |T_i(X_i - Y_i)|\right)^p \middle| T\right) \le \Sigma_n^p\, \mathbb{E}|X_1 - Y_1|^p = \Sigma_n^p\, \ell_p^p(F,G) \quad \text{a.s.}$$

for all $n \ge 1$ and therefore upon taking expectations, letting $n \to \infty$ and using the monotone convergence theorem

$$\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \le \left\| \sum_{i\ge 1} |T_i(X_i - Y_i)| \right\|_p \le \left\| \sum_{i\ge 1} |T_i| \right\|_p \ell_p(F,G).$$

which proves that $\mathscr{S}$ is indeed a contraction on on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$. □

The following list summarizes the results of the last two subsections concerning contraction properties of $\mathscr{S}$:

(a) Conditions of Theorem 5.47 $\Rightarrow$ $\mathscr{S}$ is a quasi-contraction on $\mathscr{P}_0^p(\mathbb{R})$ and has a unique fixed point in this set $(p > 1)$.
(b) Conditions of Theorem 5.48 $\Rightarrow$ $\mathscr{S}$ is a quasi-contraction on $\mathscr{P}^p(\mathbb{R})$ and has a unique fixed point in this set $(p > 1)$.
(c) Conditions of Theorem 5.49 $\Rightarrow$ $\mathscr{S}$ is a quasi-contraction on $\mathscr{P}_c^p(\mathbb{R})$ for each $c \in \mathbb{R}$ and has a unique fixed point in each of these sets $(p > 1)$.
(d) Conditions of Theorem 5.52 $\Rightarrow$ $\mathscr{S}$ is a self-map of $\mathscr{P}_0^p(\mathbb{R})$ and has a unique fixed point in this set $(p > 2)$.
(e) Conditions of Theorem 5.53 $\Rightarrow$ $\mathscr{S}$ exists in L^p-sense and has a unique fixed point in $\mathscr{P}^p(\mathbb{R})$ $(p > 2)$.
(f) Conditions of Theorem 5.54 $\Rightarrow$ $\mathscr{S}$ is a contraction on $\mathscr{P}^p(\mathbb{R})$ and has a unique fixed point in this set $(p \ge 1)$.

5.5.6 Existence of exponential moments

We close this chapter with a discussion of the question which conditions on (C,T) entail that a fixed point G_0 of $\mathscr{S}$ has exponential moments, that is $\int e^{tx}\, G_0(dx) < \infty$ for all $t \in [-t_0, t_0]$ and some $t_0 > 0$. In other words, we ask for conditions implying that 0 is an inner point of the natural domain $\{t \in \mathbb{R} : \psi_{G_0}(t) < \infty\}$ of the mgf of G_0. Our analysis will include exponential moments of an iteration sequence $(\mathscr{S}^n(F))_{n\ge 0}$. The two results to be shown thereafter, once again due to RÖSLER [97], are stated first.

Theorem 5.55. *Suppose that*

$$\sup_{i\geq 1}|T_i| \leq 1 \text{ a.s.} \quad \text{and} \quad \left\|\sum_{i\geq 1} T_i^2\right\|_1 < 1.$$

Suppose further that

$$\mathbb{E}\exp\left(t\sum_{i\geq 1}|T_i|\right) < \infty \quad \text{and} \quad \mathbb{E}e^{tC} < \infty$$

for all t in an open neighborhood of zero and that $\mathbb{E}C = 0$ if $\mathbb{E}(\sum_{i\geq 1} T_i) = 1$. Let $a := (1-\mathbb{E}(\sum_{i\geq 1} T_i))^{-1}\mathbb{E}C$ if $\mathbb{E}(\sum_{i\geq 1} T_i) \neq 1$, and arbitrary otherwise. Finally, let F be any distribution with mgf ψ_F satisfying

$$\psi_F(t) \leq e^{at+b_0t^2} \tag{5.62}$$

for some $b_0 > 0$ and all t in an open neighborhood of zero, and let G denote the ℓ_2-limit of $\mathscr{S}^n(F)$, thus a fixed point of $\mathscr{S}$ in $(\mathscr{P}^2(\mathbb{R}), \ell_2)$. Then there exists an open neighborhood $(-t_0, t_0)$ and some $b_1 \geq b_0$ such that

$$\psi_H(t) \leq e^{at+b_1t^2} \tag{5.63}$$

whenever $t \in (-t_0, t_0)$ and $H \in \{\mathscr{S}^n(F) : n \geq 0\} \cup \{G\}$.

Additional conditions on (C,T) lead to the following improvement of the previous theorem.

Theorem 5.56. *Given the assumptions of Theorem 5.55, suppose further that*

$$\sum_{i\geq 1} T_i^2 \leq 1 \text{ a.s.} \quad \text{and} \quad \mathbb{E}e^{3s|C|} < \infty \text{ for some } s > 0$$

and, in the case $\mathbb{E}(\sum_{i\geq 1} T_i) \neq 1$, that

$$\mathbb{E}\exp\left(3\left|a\sum_{i\geq 1} T_i\right|\right) < \infty.$$

Then assertion (5.63) holds true for all $t \in [-s,s]$, some $b_1 > 0$, and with $a = 0$ if $\mathbb{E}(\sum_{i\geq 1} T_i) = 1$.

The reader should notice that property (5.62) entails that F has absolute moments of arbitrary order, i.e. $F \in \mathscr{P}^p(\mathbb{R})$ for all $p > 0$. This follows because

$$\mathbb{E}e^{t|X|} \;\le\; \psi_F(t) + \psi_F(-t)$$

for some $t > 0$ and any random variable X with law F.

Proof (of Theorem 5.55). Since $|T_i| \le 1$ for all $i \ge 1$, we have $\sum_{i\ge1} T_i^2 \le \sum_{i\ge1} |T_i|$ and thus

$$\mathbb{E}\exp\left(t\sum_{i\ge1} T_i^2\right) \;\le\; \mathbb{E}\exp\left(t\sum_{i\ge1} |T_i|\right) \;<\; \infty$$

for all t in an open neighborhood of zero. Defining $R_1 := \sum_{i\ge1} T_i - 1$ and $R_2 := \sum_{i\ge1} T_i^2 - 1$, consider the function

$$f_b(t) := \mathbb{E}e^{tC+atR_1+bt^2R_2}$$

for any $b > 0$, which is finite for all t in an open neighborhood of zero (depending on b). Then f_b is infinitely often differentiable on this set and

$$\begin{aligned} f_b'(t) &= \mathbb{E}(C + aR_1 + 2btR_2)e^{tC+atR_1+bt^2R_2},\\ f_b''(t) &= \mathbb{E}\Big(2bR_2 + (C + aR_1 + 2btR_2)^2\Big)e^{tC+atR_1+bt^2R_2}. \end{aligned}$$

By choice of a, this implies

$$f_b'(0) = \mathbb{E}(C + aR_1) = 0 \quad \text{and} \quad f_b''(0) = 2b\,\mathbb{E}R_2 + \mathbb{E}(C + aR_1)^2,$$

and since $\mathbb{E}R_2 < 0$ by assumption, we can fix $b_1 \ge b_0$ such that $f_{b_1}''(0) < 0$. By combining these facts with $f_{b_1}(0) = 1$, we find that $f_{b_1}(t) \le 1$ for all t in a sufficiently small neighborhood $(-t_0, t_0)$ of zero.

Now let F be a distribution satisfying (5.62) on $(-t_0, t_0)$ which may require to further reduce the given t_0. We prove by induction over $n \ge 0$ that (5.63) holds for $H = \mathscr{S}^n(F)$. If $n = 0$, this follows from (5.62), for $b_1 \ge b_0$. So assume validity of (5.63) for some $\mathscr{S}^n(F)$ and pick a (C,T)-independent sequence $(X_i)_{i\ge1}$ of iid random variables with common law $\mathscr{S}^n(F)$. Using $\sup_{i\ge1} |T_i| \le 1$ a.s. and the inductive hypothesis, we infer

$$\begin{aligned} \mathbb{E}\exp\left(t\left(\sum_{i\ge1} T_iX_i + C\right)\middle| C,T\right) \;&=\; e^{tC}\prod_{i\ge1}\psi_{\mathscr{S}^n(F)}(tT_i)\\ &\le\; e^{at+b_1t^2}\, e^{tC+atR_1+bt^2R_2} \quad \text{a.s.} \end{aligned}$$

and then

$$\psi_{\mathscr{S}^{n+1}(F)}(t) \;=\; \mathbb{E}\exp\left(t\left(\sum_{i\ge1} T_iX_i + C\right)\right) \;\le\; e^{at+b_1t^2} f_{b_1}(t) \;\le\; e^{at+b_1t^2}$$

for all $t \in (-t_0, t_0)$ which is the desired inequality. Finally, use $\mathscr{S}^n(F) \xrightarrow{w} G$ and thus $\mathscr{S}^n(F)^{-1}(U) \to G^{-1}(U)$ a.s. for any $Unif(0,1)$ variable U to conclude with the help of Fatou's lemma that

$$\psi_G(t) \le \liminf_{n\to\infty} \psi_{\mathscr{S}^n(F)}(t) \le e^{at+b_1t^2}$$

for all $t \in (-t_0, t_0)$ holds true as well. ◻

We could have arrived at the very last conclusion also by observing that (5.63) for the $\mathscr{S}^n(F)$ implies uniform integrability of $(\exp(t\mathscr{S}^n(F)^{-1}(U)))_{n\ge 0}$ for all $t \in (-t_0, t_0)$, which in turn gives $\psi_{\mathscr{S}^n(F)}(t) \to \psi_G(t)$ and thus again validity of (5.63) for $\psi_G(t)$ and all t from this interval.

Proof (of Theorem 5.56). Keeping the notation from before, it is obviously enough to show $f_b(t) \le t$ for all $t \in [-s, s]$ and a sufficiently large b. By Theorem 5.55, this is true for all $|t| \le \varepsilon < s$ and a suitable ε, which does not depend on b because $R_2 \le 0$ a.s. here. If $|t| \in (\varepsilon, s]$, then use the generalized Hölder inequality [☞ e.g. [12, p. 78]] and the afore-mentioned property of R_2 to infer

$$\begin{aligned} f_b(t) &\le \left\|e^{tC}\right\|_3 \left\|e^{atR_1}\right\|_3 \left\|e^{bt^2R_2}\right\|_3 \\ &\le \left\|e^{t|C|}\right\|_3 \left\|\exp\left(as\left|\sum_{i\ge 1} T_i\right| + as\right)\right\|_3 \left\|e^{b\varepsilon^2R_2}\right\|_3 \end{aligned}$$

where the last product is finite by the conditions of the theorem. Since, furthermore, $\mathbb{P}(R_2 < 0) > 0$, the last factor converges to 0 as $b \to \infty$. Obviously, this yields the desired conclusion. ◻

Problems

Problem 5.57. Complete the proof of Lemma 5.44.

Problem 5.58. Consider the situation stated in Lemma 5.50(b) and prove the following assertions:

(a) $\mathbb{E}\mathscr{S}(F) - \mathbb{E}\mathscr{S}(G) = \mathbb{E}(\sum_{i\ge 1} T_i)(\mathbb{E}F - \mathbb{E}G)$ for all $F, G \in \mathscr{P}^1(\mathbb{R})$.

(b) If $(X_1, Y_1), (X_2, Y_2), \ldots$ are iid and (C, T)-independent random variables such that $(X_1, Y_1) \sim (F^0, G^0)$ and $\|X_1 - Y_1\|_2 = \ell_2(F^0, G^0)$ for any $F, G \in \mathscr{P}^2(\mathbb{R})$, then

$$\ell_2^2(\mathscr{S}(F)^0, \mathscr{S}(G)^0) \le \left\|\sum_{i\ge 1} T_i(X_i - Y_i)\right\|_2^2 + \mathbb{V}\mathrm{ar}\left(\sum_{i\ge 1} T_i\right)(\mathbb{E}F - \mathbb{E}G)^2.$$

(c) Use part (b), (5.49) and

$$\ell_2^2(\mathscr{S}(F),\mathscr{S}(G)) \;=\; \ell_2^2(\mathscr{S}(F)^0,\mathscr{S}(G)^0) + (\mathbb{E}\mathscr{S}(F) - \mathbb{E}\mathscr{S}(G))^2,$$

valid by (5.42) of Lemma 5.21, to give an alternative proof of inequality (5.50), the difference being that $\mathscr{S}$ and the centering operation are applied in reverse order.

Finally, consider the situation in Lemma 5.51(b) for $p = 2$ and adopt the notation of its proof, viz. $a_n = \|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\|_2$ for $n \ge 1$, $\alpha = \|\sum_{i\ge1} T_i^2\|_1^{1/2}$, $\beta = \mathbb{E}(\sum_{i\ge1} T_i)$ and $\gamma^2 = \mathbb{V}\mathrm{ar}\,(\sum_{i\ge1} T_i)$. By improving the argument from there, prove the following:

(d) For all $n \ge 1$ and $F,G \in \mathscr{P}^2(\mathbb{R})$,

$$\ell_2^2(\mathscr{S}^n(F),\mathscr{S}^n(G)) \;=\; a_n^2\Big(\ell_2^2(F^0,G^0) + (\mathbb{E}F - \mathbb{E}G)^2\Big)$$

as well as $a_n^2 = \beta^{2(n-1)} a_1 + \gamma^2 \sum_{k=0}^{n-2} \beta^{2k}\alpha^{2(n-k-1)}$ hold true. What is the resulting estimate for $\ell_2^2(\mathscr{S}^n(F),\mathscr{S}^n(G))$ alternative to (5.50)?

Problem 5.59. Replacing the condition $\mathbb{E}(\sum_{i\ge1} T_i) = 1$ with $\mathbb{E}(\sum_{i\ge1} T_i) = -1$ in Theorem 5.49, prove the following assertions:

(a) $\mathscr{S}^2$ fulfills the assumptions of Theorem 5.49 and hence possesses a unique geometrically ℓ_p-attracting fixed point G_c in $\mathscr{P}_c^p(\mathbb{R})$ for each $c \in \mathbb{R}$.

(b) $\mathscr{S}$ itself has at most one fixed point, namely G_0. [Hint: Lemma 5.44]

Problem 5.60. Given any nonnegative random variable W, define $\Delta^0 W := W$ and

$$\Delta W := (W - \mathbb{E}W)^2, \quad \Delta^2 W = \Big((W - \mathbb{E}W)^2 - \mathbb{V}\mathrm{ar}\,W\Big)^2, \quad \text{etc.}$$

with the usual convention that $\Delta W := \infty$ if $\mathbb{E}W = \infty$. Show that

$$\mathbb{E}\Delta^k W \;\le\; \mathbb{E}(\Delta^{k-1}W)^2 \;\le\; 2\,\mathbb{E}(\Delta^{k-2}W)^4 \;\le \ldots \le\; 2^{k-1}\mathbb{E}W^{2^k}$$

for all $k \ge 1$ [☞ (5.60)].

5.6 An application: `Quicksort` asymptotics

In Section 1.4, we gave a brief introduction of the sorting algorithm `Quicksort` and provided an outline of how to determine the asymptotic behavior, as $n \to \infty$, of the number of key comparisons X_n necessary to sort a random list of n distinct numbers, w.l.o.g. a random permutation of $1,\ldots,n$. Here is a short summary of the major findings from there. First of all, the crucial random recursive equation (1.20) must be recalled, viz.

$$X_n \;\stackrel{d}{=}\; X'_{Z_n-1} \;+\; X''_{n-Z_n} \;+\; n-1$$

for all $n \ge 1$, where $X_0' = X_0'' = 0$ and $(X_n')_{n\ge 1}, (X_n'')_{n\ge 1}, (Z_n)_{n\ge 1}$ are independent with $\mathscr{L}(X_n') = \mathscr{L}(X_n'') = \mathscr{L}(X_n)$ and $\mathscr{L}(Z_n) = Unif\{1,...,n\}$ for each $n \ge 1$. We further recall that $\mathbb{E}X_n = 2n\log n + O(n)$ [☞ Lemma 1.13] and $\mathbb{V}\mathrm{ar}X_n = (7-\frac{2}{3}\pi^2)n^2 + o(n^2)$ [☞ Problem 1.16]. Defining the normalization

$$\widehat{X}_n := \frac{X_n - \mathbb{E}X_n}{n},$$

the above distributional equation may be rewritten in terms of $\widehat{X}_n$, namely [☞ (1.24) and (1.25)]

$$\widehat{X}_n \stackrel{d}{=} \frac{Z_n-1}{n}\widehat{X}'_{Z_n-1} + \frac{n-Z_n}{n}\widehat{X}''_{n-Z_n} + g_n(Z_n) \tag{5.64}$$

for $n \ge 2$, where $\widehat{X}_0 = \widehat{X}_1 := 0$ and $g_n : \{1,...,n\} \to \mathbb{R}$ is defined by

$$g_n(k) := \frac{n-1}{n} + \frac{1}{n}(\mathbb{E}X_{k-1} + \mathbb{E}X_{n-k} - \mathbb{E}X_n).$$

It is easily verified that $Z_n/n \stackrel{d}{\to} Unif(0,1)$, and we will show in Lemma 5.62 that

$$\lim_{n\to\infty} g_n(\lceil nt \rceil) = g(t) := 1 + 2t\log t + 2(1-t)\log(1-t) \tag{5.65}$$

for all $t \in (0,1)$ uniformly on compact subsets, where $\lceil x \rceil := \inf\{n \in \mathbb{Z} : x \le n\}$. Therefore it seems plausible and will actually be the main result to be shown in this section that $\widehat{X}_n \stackrel{d}{\to} \widehat{X}_\infty$ for some random variable $\widehat{X}_\infty$ with a law which in $\mathscr{P}_0^2(\mathbb{R})$ uniquely solves the so-called Quicksort equation

$$\widehat{X}_\infty \stackrel{d}{=} U\widehat{X}'_\infty + (1-U)\widehat{X}''_\infty + g(U) \tag{5.66}$$

where $\widehat{X}'_\infty, \widehat{X}''_\infty$ and U are independent with $\widehat{X}'_\infty \stackrel{d}{=} \widehat{X}''_\infty \stackrel{d}{=} \widehat{X}_\infty$ and $U \stackrel{d}{=} Unif(0,1)$. Plainly, this is the SFPE pertaining to the smoothing transform $\mathscr{S}$ with

$$T_1 = U, \quad T_2 = 1-U, \quad T_3 = T_4 = ... = 0 \quad \text{and} \quad C = g(U),$$

and one can immediately assess that

(1) $\sum_{i\ge 1} T_i = 1$.
(2) $\sum_{i\ge 1} T_i^p = U^p + (1-U)^p < 1$ and thus $\mathscr{C}_p(T) < 1$ for all $p > 1$.
(3) $\mathbb{E}C = 0$ because $\mathbb{E}U\log U = \mathbb{E}(1-U)\log(1-U) = -\frac{1}{4}$ [☞ Problem 5.63].
(4) $\|C\|_p < \infty$ for all $p > 0$ because g is bounded on $(0,1)$.

Hence the conditions of Theorem 5.49 are fulfilled for each $p > 1$ and we conclude from this result that $\mathscr{S}$ has indeed a fixed point in $\bigcap_{p>1}\mathscr{P}_0^p(\mathbb{R})$ which is unique up to translations and has moments of arbitrary order. Denote by F the unique fixed point with mean 0, and let F_n be the law of $\widehat{X}_n$. Here is our main result of this section:

Theorem 5.61. *In the given notation,* $F_n \xrightarrow{\ell_p} F$ *for each* $p > 0$.

For its proof, we first give a technical lemma that shows (5.65), that is the convergence of $g_n(\lceil nt \rceil)$ to $g(t) = 1 + 2t\log t + 2(1-t)\log(1-t)$ for t in the open unit interval.

Lemma 5.62. *The function* $g_n(\lceil n\cdot \rceil)$ *converges pointwise to* g *on* $(0,1)$*, and the convergence is uniform on compact subsets. Moreover, the* g_n *are uniformly bounded on* $[0,1]$*, that is*

$$\sup_{n\ge 1}\max_{1\le i\le n} g_n(i) < \infty.$$

Proof. Putting $\ell(t) := t\log t$ and $R(k) := \mathbb{E}X_k - 2\ell(k)$, write

$$\begin{aligned} g_n(\lceil nt \rceil) &= \frac{n-1}{n} + \frac{1}{n}\Big(2\,\ell(\lceil nt \rceil - 1) + R(\lceil nt \rceil - 1) \\ &\quad + 2\,\ell(n - \lceil nt \rceil) + R(n - \lceil nt \rceil) - 2\,\ell(n) - R(n)\Big) \\ &= \frac{n-1}{n} + 2\,\ell\left(\frac{\lceil nt \rceil - 1}{n}\right) + 2\,\ell\left(\frac{n - \lceil nt \rceil}{n}\right) - \frac{2}{n}\log n \\ &\quad + \frac{\lceil nt \rceil - 1}{n}\cdot\frac{R(\lceil nt \rceil - 1)}{\lceil nt \rceil - 1} + \frac{n - \lceil nt \rceil}{n}\cdot\frac{R(n - \lceil nt \rceil)}{n - \lceil nt \rceil} - \frac{R(n)}{n} \end{aligned} \tag{5.67}$$

for all $t \in (0,1)$ and $n \in \mathbb{N}$. This implies the pointwise convergence of $g_n(\lceil nt \rceil)$ as $n \to \infty$, because $\lim_{n\to\infty} R(n)/n$ exists and is finite by Lemma 1.13. We leave it to the reader to verify that the convergence is uniform on compact subsets of $(0,1)$ [☞ Problem 5.64]. Replacing $\lceil nt \rceil$ with i in (5.67), the last assertion of the lemma is immediate when using the convergence of $R(n)/n$ in combination with the uniform boundedness of ℓ on $[0,1]$. ☐

Proof (of Theorem 5.61). The asserted convergence will be shown by an induction over even integers p, which suffices by the monotonicity of $\ell_p(\cdot,\cdot)$ in p.

STEP 1: $p = 2$.
Let U, V, W be independent *Unif*$(0,1)$ variables, $G_n = \text{Unif}\{1/n, 2/n, ..., 1\}$ for $n \ge 1$ and

$$\begin{aligned} &\widehat{X}_n' := F_n^{-1}(V),\ n \in \mathbb{N}, \quad \widehat{X}_\infty' := F^{-1}(V), \\ &\widehat{X}_n'' := F_n^{-1}(W),\ n \in \mathbb{N}, \quad \widehat{X}_\infty'' := F^{-1}(W), \\ &Z_n := n\,G_n^{-1}(U),\ n \in \mathbb{N}. \end{aligned}$$

Obviously, $\mathscr{L}(Z_n) = Unif\{1,...,n\}$, and we have for $n \geq 2$ that

$$\begin{aligned}\ell_2(F_n,F) &\leq \left\| \frac{Z_n-1}{n}\,\widehat{X}'_{Z_n-1} + \frac{n-Z_n}{n}\,\widehat{X}''_{Z_n-1} + g_n(Z_n) - U\widehat{X}'_\infty - (1-U)\widehat{X}''_\infty - g(U)\right\|_2 \\ &\leq \Upsilon_{1,n} + \Upsilon_{2,n} + \Upsilon_{3,n},\end{aligned}$$

where

$$\begin{aligned}\Upsilon_{1,n} &:= \left\| \frac{Z_n-1}{n}\,(\widehat{X}'_{Z_n-1} - \widehat{X}'_\infty) + \frac{n-Z_n}{n}\,(\widehat{X}''_{n-Z_n} - \widehat{X}''_\infty)\right\|_2, \\ \Upsilon_{2,n} &:= \left\| \left(\frac{Z_n-1}{n} - U\right)\widehat{X}'_\infty + \left(\frac{n-Z_n}{n} - (1-U)\right)\widehat{X}''_\infty\right\|_2, \\ \text{and}\quad \Upsilon_{3,n} &:= \|g_n(Z_n) - g(U)\|_2,\end{aligned}$$

which are now further estimated individually.

Since the $\widehat{X}'_n, \widehat{X}''_n$ for $n \in \overline{\mathbb{N}}_0$ are independent with mean zero, we infer

$$\begin{aligned}\Upsilon_{1,n}^2 &= \sum_{i=1}^n \mathbb{P}(Z_n = i)\,\mathbb{E}\left(\frac{i-1}{n}\,(\widehat{X}'_{i-1} - \widehat{X}'_\infty) + \frac{n-i}{n}\,(\widehat{X}''_{n-i} - \widehat{X}''_\infty)\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^n \left(\left(\frac{i-1}{n}\right)^2 \mathbb{E}(\widehat{X}'_{i-1} - \widehat{X}'_\infty)^2 + \left(\frac{n-i}{n}\right)^2 \mathbb{E}(\widehat{X}''_{n-i} - \widehat{X}''_\infty)^2\right) \\ &= \frac{1}{n}\sum_{i=1}^n \left(\left(\frac{i-1}{n}\right)^2 \ell_2^2(F_{i-1},F) + \left(\frac{n-i}{n}\right)^2 \ell_2^2(F_{n-i},F)\right) \\ &= \frac{2}{n}\sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^2 \ell_2^2(F_i,F).\end{aligned}$$

As for $\Upsilon_{2,n}$, we have the estimate

$$\begin{aligned}\Upsilon_{2,n} &\leq \left(\left\|\frac{Z_n-1}{n} - U\right\|_2 + \left\|\frac{n-Z_n}{n} - (1-U)\right\|_2\right)\|\widehat{X}'_\infty\|_2 \\ &= \left(\|G_n^{-1}(U) - U\|_2 + \frac{1}{n} + \|(1-G_n^{-1}(U)) - (1-U)\|_2\right)\|\widehat{X}'_\infty\|_2 \\ &= \left(2\,(\|G_n^{-1}(U) - U\|_2 + \frac{1}{n}\right)\|\widehat{X}'_\infty\|_2.\end{aligned}$$

Since $G_n \stackrel{w}{\to} U$ implies $G_n^{-1}(U) \to U$ a.s., this further implies

$$\lim_{n\to\infty} \Upsilon_{2,n} = 0.$$

For the third term $\Upsilon_{3,n}$, we first point out that $G_n^{-1}(w) \to w$ actually holds true *for all* $w \in (0,1)$. Fixing w, we thus find $\varepsilon_w > 0$ and $n_0 \in \mathbb{N}$ such that $G_n^{-1}(w) \in [w-\varepsilon_w, w+\varepsilon_w] \subset (0,1)$ for all $n \geq n_0$. By combining this with Lemma 5.62, we infer with the help of the dominated convergence theorem that

$$\lim_{n\to\infty}\Upsilon_{3,n} = \lim_{n\to\infty}\|g_n(Z_n)-g(U)\|_2 = 0.$$

Summarizing, we have shown that

$$\ell_2(F_n,F) \;\le\; \left(\frac{2}{n}\sum_{i=1}^{n-1}\left(\frac{i}{n}\right)^2 \ell_2^2(F_i,F)\right)^{1/2} + b_n \tag{5.68}$$

for all $n\ge 2$ and a suitable sequence $(b_n)_{n\ge 2}$ with limit zero. Setting $a_n := \ell_2(F_n,F)$ and $a_n^* := \max_{1\le i\le n} a_i$, it follows upon using the well-known formula $\sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1)$ in (5.68) that

$$\begin{aligned} a_n \;&\le\; a_{n-1}^*\left(\sum_{i=1}^{n-1}\left(\frac{i}{n}\right)^2\right)^{1/2} + b_n \\ &=\; \frac{a_{n-1}^*}{n}\left(\frac{1}{3}(n-1)(2n-1)\right)^{1/2} + b_n \\ &\le\; \left(\frac{2}{3}\right)^{1/2} a_{n-1}^* + b_n \end{aligned}$$

and thus the boundedness of the a_n. We leave it to the reader to further conclude from this and (5.68) that $a_n\to 0$ as $n\to\infty$ [☞ Problem 5.65].

STEP 2: $p-2\to p$.
Suppose that $\ell_q(F_n,F)\to 0$ for any $q\le p-2$ (inductive hypothesis). It is clear that we still have

$$\ell_p(F_n,F) \;\le\; \Upsilon_{1,n}+\Upsilon_{2,n}+\Upsilon_{3,n}$$

if we replace $\|\cdot\|_2$ by $\|\cdot\|_p$ in the definitions of the $\Upsilon_{i,n}$. It is also immediate that $\Upsilon_{2,n}$ and $\Upsilon_{3,n}$ are still convergent to zero which leaves us with a study of

$$\Upsilon_{1,n} := \left\|\frac{Z_n-1}{n}\,(\widehat{X}'_{Z_n-1}-\widehat{X}'_\infty)+\frac{n-Z_n}{n}\,(\widehat{X}''_{n-Z_n}-\widehat{X}''_\infty)\right\|_p.$$

Keeping in mind that $\|\widehat{X}'_i-\widehat{X}'_\infty\|_q = \|\widehat{X}''_i-\widehat{X}''_\infty\|_q = \ell_q(F_i,F)$ for any $q\le p$ and $\mathbb{E}(\widehat{X}'_i-\widehat{X}'_\infty) = \mathbb{E}(\widehat{X}''_i-\widehat{X}''_\infty)=0$, we obtain

$$\begin{aligned} \Upsilon_{1,n} \;&=\; \left(\sum_{i=1}^n \mathbb{P}(Z_n=i)\,\mathbb{E}\left(\frac{i-1}{n}\,(\widehat{X}'_{i-1}-\widehat{X}'_\infty)+\frac{n-i}{n}\,(\widehat{X}''_{n-i}-\widehat{X}''_\infty)\right)^p\right)^{1/p} \\ &=\; \frac{1}{n^{1/p}}\left(\sum_{i=1}^n\left[\left(\frac{i-1}{n}\right)^p \mathbb{E}(\widehat{X}'_{i-1}-\widehat{X}'_\infty)^p + \left(\frac{n-i}{n}\right)^p \mathbb{E}(\widehat{X}''_{n-i}-\widehat{X}''_\infty)^p\right]\right. \end{aligned}$$

$$+\sum_{k=1}^{p-1}\binom{p}{k}\sum_{i=1}^{n}\left[\left(\frac{i-1}{n}\right)^{k}\mathbb{E}(\widehat{X}'_{i-1}-\widehat{X}'_{\infty})^{k}\left(\frac{n-i}{n}\right)^{p-k}\mathbb{E}(\widehat{X}''_{n-i}-\widehat{X}''_{\infty})^{p-k}\right]\Bigg)^{1/p}$$
$$\le \left(\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{i-1}{n}\right)^{p}\ell_p^p(F_{i-1},F)+\left(\frac{n-i}{n}\right)^{p}\ell_p^p(F_{n-i},F)\right]\right)^{1/p}$$
$$+\left(\frac{1}{n}\sum_{k=2}^{p-2}\binom{p}{k}\sum_{i=1}^{n}\left[\left(\frac{i-1}{n}\right)^{k}d_k^k(F_{i-1},F)\left(\frac{n-i}{n}\right)^{p-k}\ell_{p-k}^{p-k}(F_{n-i},F)\right]\right)^{1/p}$$
$$=\left(\frac{2}{n}\sum_{i=1}^{n-1}\left(\frac{i}{n}\right)^{p}\ell_p^p(F_{i-1},F)\right)^{1/p}+c_n,$$

where

$$c_n := \left(\frac{1}{n}\sum_{k=2}^{p-2}\binom{p}{k}\sum_{i=1}^{n}\left[\left(\frac{i-1}{n}\right)^{k}d_k^k(F_{i-1},F)\left(\frac{n-i}{n}\right)^{p-k}\ell_{p-k}^{p-k}(F_{n-i},F)\right]\right)^{1/p}$$

is easily seen to converge to zero as $n\to\infty$ by making use of the inductive hypothesis. Therefore we have shown that

$$\ell_p(F_n,F) \le \left(\frac{2}{n}\sum_{i=1}^{n-1}\left(\frac{i}{n}\right)^{p}\ell_p^p(F_{i-1},F)\right)^{1/p}+b_n$$

for all $n\ge 2$ and suitable b_n convergent to zero as $n\to\infty$. The remaining argument to conclude $\ell_p(F_n,F)\to 0$ is the same as in the case $p=2$. □

Problems

Problem 5.63. Check that $\mathbb{E}U\log U=-\frac{1}{4}$ if $U\stackrel{d}{=}Unif(0,1)$.

Problem 5.64. Given the situation of Lemma 5.62, prove that the asserted convergence $g_n(\lceil nt\rceil)\to g(t)$ is uniform on compact subsets of $(0,1)$.

Problem 5.65. Use (5.68) and the boundedness of $a_n=\ell_2(F_n,F)$ to show that $a_n\to 0$ as $n\to\infty$.

Problem 5.66. If F as before denotes the Quicksort distribution and ψ_F its mgf, prove that

$$\psi_F(t)\le e^{bt^2}$$

for all $t\ge 0$ and a suitable $b>0$. [Remark: For further information the reader may consult the article by Fill & Janson [53].]

Problem 5.67. Consider the homogeneous version of the `Quicksort` equation, viz.

$$X \stackrel{d}{=} UX_1 + (1-U)X_2, \quad U \stackrel{d}{=} \mathit{Unif}(0,1), \tag{5.69}$$

and show that

(a) every Cauchy distribution $\mathit{Cauchy}(a,b)$, $(a,b) \in \mathbb{R} \times \mathbb{R}_>$, with $\lambda\!\!\lambda$-density $g_{a,b}(x) = \frac{1}{\pi}\frac{b}{(x-a)^2+b^2}$ forms a solution to this equation.

(b) every convolution of the `Quicksort` distribution F with a Cauchy distribution solves the `Quicksort` equation. [Remark: This is part of a stronger result due to FILL & JANSON [52].]

(c) Stipulating $\mathit{Cauchy}(a,0) = \delta_a$, parts (a) and (b) remain true for these degenerate distributions. How does this relate to Theorem 5.49.

5.7 The Hausdorff dimension of random Cantor sets: completing the proof of Theorem 5.12

We return to the situation of Section 5.2 and will now complete the proof of Theorem 5.12 by MAULDIN & WILLIAMS [85] about the Hausdorff dimension of random Cantor sets obtained from a *random recursive construction* [☞ (RC-1)-(RC-4)]. Naturally adopting the notation from there, we are given a family $\boldsymbol{J} = \{J_{\mathsf{v}} : \mathsf{v} \in \mathbb{T}\}$ of random compact subsets of $\mathbb{R}^d$ which defines a random Cantor set $\mathfrak{C}$ via

$$\mathfrak{C} \;=\; \bigcap_{n\ge 0} \bigcup_{|\mathsf{v}|=n} J_{\mathsf{v}}.$$

Assuming (RC-1)-(RC-4), Theorem 5.12 claims that the Hausdorff dimension of $\mathfrak{C}$ satisfies

$$\mathbb{P}(\dim_H \mathfrak{C} = \alpha | \mathfrak{C} \ne \emptyset) = 1,$$

where

$$\alpha \;:=\; \inf\{\theta \ge 0 : \mathfrak{m}(\theta) \le 1\} \;\in\; (0,d].$$

The simple half, viz. $\mathbb{P}(\dim_H \mathfrak{C} \le \alpha | \mathfrak{C} \ne \emptyset) = 1$, has already been settled in 5.2. This section is therefore devoted to the proof of the more difficult assertion that $\mathbb{P}(\dim_H \mathfrak{C} \ge \alpha | \mathfrak{C} \ne \emptyset) = 1$.

As usual, let $T = (T_1, T_2, \ldots)$ and $N = \sum_{i\ge 1} \mathbf{1}_{\{T_i>0\}}$ denote generic copies of $T(\mathsf{v})$ and $N(\mathsf{v})$, respectively. Recall from (5.23) that

$$Y_{\theta,n} \;=\; \sum_{|\mathsf{v}|=n} L(\mathsf{v})^\theta \;=\; \sum_{|\mathsf{v}|=n} \frac{\mathit{diam}(J_{\mathsf{v}})^\theta}{\mathit{diam}(J)^\theta}$$

for $n \ge 1$ and $\theta \ge 0$, and put $Y_n := Y_{1,n}$. If $\mathfrak{m}(\alpha) = 1$, then $(Y_{\alpha,n})_{n\ge 0}$ constitutes a nonnegative martingale [☞ Lemma 5.5], and we begin with two lemmata about this martingale and its almost sure limit.

Lemma 5.68. *In addition to the assumptions of Theorem 5.12, suppose that* $\mathfrak{m}(\alpha) = 1$ *and* $Y_{\alpha,1} \in L^p$ *for some* $p > 1$*. Then the martingale* $(Y_{\alpha,n})_{n\geq 0}$ *is* L^p*-bounded, i.e.*

$$\sup_{n\geq 0} \|Y_{\alpha,n}\|_p < \infty.$$

Proof. W.l.o.g. we may assume $\alpha = 1$, thus $Y_{\alpha,n} = Y_n$. Recall that $\mathbb{E}Y_{\theta,n} = \mathfrak{m}(\theta)^n$ for all $\theta > 0$ and $n \in \mathbb{N}_0$, and that $\mathfrak{m}(\theta) < 1$ if $\theta > 1$. Put $\nu(r) := \|Y_1 - 1\|_r$ for $r \leq p$ and $Y(\mathsf{v}) := \sum_{i\geq 1} T_i(\mathsf{v})$ for $\mathsf{v} \in \mathbb{T}$. Then the $Y(\mathsf{v})$ are iid copies of Y_1 and $(Y(\mathsf{v}))_{|\mathsf{v}|=n}$ independent of $\mathscr{F}_n$ (defined in (5.17)) for each n.

We first show that each Y_n is in L^p which must be done only for $n \geq 2$, for $Y_0 = 1$ and $Y_1 \in L^p$ by assumption. Using Jensen's inequality, we infer on $\{Y_{n-1} > 0\}$

$$\begin{aligned} Y_n^p &= \left(\sum_{|\mathsf{v}|=n-1} \frac{L(\mathsf{v})}{Y_{n-1}} Y_{n-1} Y(\mathsf{v})\right)^p \\ &\leq \sum_{|\mathsf{v}|=n-1} \frac{L(\mathsf{v})}{Y_{n-1}} Y_{n-1}^p Y(\mathsf{v})^p \\ &= Y_{n-1}^{p-1} \sum_{|\mathsf{v}|=n-1} L(\mathsf{v}) Y(\mathsf{v})^p. \end{aligned}$$

Since $\{Y_n > 0\} \subset \{Y_{n-1} > 0\}$, this implies

$$\mathbb{E}Y_n^p \;\leq\; \mathbb{E}Y_1^p\, \mathbb{E}\left(Y_{n-1}^{p-1} \sum_{|\mathsf{v}|=n-1} L(\mathsf{v})\right) \;=\; \mathbb{E}Y_1^p\, \mathbb{E}Y_{n-1}^p$$

and thus $a_n = \|Y_n\|_p < \infty$ for all $n \geq 0$ by an inductive argument.

Now recall (5.58), which in the present notation reads

$$a_{n+1} \;\leq\; a_n + \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(Y(\mathsf{v}) - 1)\right\|_p \tag{5.70}$$

for each $n \geq 0$. Use inequality (B.5) in the Appendix (the Burkholder-Davis-Gundy inequality specialized to weighted sums of iid random variables) after conditioning with respect to $\mathscr{F}_n$ to infer

$$\left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(Y(\mathsf{v}) - 1)\right\|_p^p \;\leq\; c_p \left(\nu(2)^p\, \mathbb{E}Y_{2,n}^{p/2} + \nu(p)^p\, \mathfrak{m}(p)^n\right)$$

for all $n \geq 0$ and some constant c_p only depending on p, and then with the help of the Cauchy-Schwarz inequality

$$\begin{aligned}\mathbb{E}Y_{2,n}^{p/2} &\le \mathbb{E}\left(\max_{|\mathsf{v}|=n} L(\mathsf{v}) \sum_{|\mathsf{v}|=n} L(\mathsf{v})\right)^{p/2}\\ &\le \left(\mathbb{E}\left(\max_{|\mathsf{v}|=n} L(\mathsf{v})^p\right)\right)^{1/2} \left(\mathbb{E}Y_n^p\right)^{1/2}\\ &\le \mathfrak{m}(p)^{n/2} a_n^{p/2}\end{aligned}$$

for all $n \ge 0$. By combining both estimates, we find in (5.70) that

$$a_{n+1} \le a_n + c_p^{1/p}\left(\nu(2)\,\mathfrak{m}(p)^{n/2p} a_n^{1/2} + \nu(p)\,\mathfrak{m}(p)^{n/p}\right)$$

and thereupon

$$\frac{a_{n+1}}{a_n} \le 1 + K\,\mathfrak{m}(p)^{n/2p}$$

for all $n \ge 0$ and a suitable constant $K \in \mathbb{R}_>$, because $\mathfrak{m}(p) < 1$ and $a_1 \le a_2 \le \ldots$ (as $(Y_n^p)_{n\ge 0}$ forms a submartingale). Finally,

$$\sup_{n\ge 1} \|Y_n\|_p = \prod_{n\ge 0} \frac{a_{n+1}}{a_n} \le \prod_{n\ge 0} (1 + K\,\mathfrak{m}(p)^{n/2p}) < \infty$$

completes our proof. ◻

Defining $(Y_{\alpha,n}(\mathsf{v}))_{n\ge 0}$ for any $\mathsf{v} \in \mathbb{T}$ in the usual manner, namely

$$Y_{\alpha,n}(\mathsf{v}) := \sum_{|\mathsf{w}|=n} [L(\mathsf{w})]_{\mathsf{v}}^{\alpha}, \quad n \ge 0,$$

the previous lemma implies that, for all $\mathsf{v} \in \mathbb{T}$,

$$Y_{\alpha,n}(\mathsf{v}) \to W(\mathsf{v}) \quad \text{a.s. and in } L^p,$$

where $\{W(\mathsf{v}) : \mathsf{v} \in \mathbb{T}\}$ forms a family of identically distributed random variables in L_1^p which are further independent if $\mathsf{v} \in \{\mathsf{u} : |\mathsf{u}| = n\}$ for any $n \ge 0$. Denote by $G \in \mathscr{P}^p(\mathbb{R}_\ge)$ their common distribution and by W a generic copy.

Lemma 5.69. *Given the assumptions of Lemma 5.68, the following assertions hold true:*

(a) *For all $\mathsf{v} \in \mathbb{T}$ and $n \ge 0$, $W(\mathsf{v}) = \sum_{|\mathsf{w}|=n} [L(\mathsf{w})]_{\mathsf{v}} W(\mathsf{vw})$, whence G constitutes a fixed point of the homogeneous smoothing transform $\mathscr{S}$ associated with $T^\alpha = (T_i^\alpha)_{i\ge 1}$.*

(b) *$q := \mathbb{P}(W = 0)$ is the unique fixed point in $[0,1)$ of the gf $f(s) = \mathbb{E}s^N$ of the counting variable N, where $s^\infty := 0$.*

Proof. (a) Naturally, it suffices to show the asserted equation for $\mathsf{v} = \varnothing$. As the WBP associated with $(T(\mathsf{v})^{\alpha})_{\mathsf{v}\in\mathbb{T}}$, $(Y_{\alpha,n})_{n\ge 0} = (Y_{\alpha,n}(\varnothing))_{n\ge 0}$ satisfies the backward equation

$$Y_{\alpha,n} \;=\; \sum_{|\mathsf{w}|=k} L(\mathsf{w})^{\alpha} Y_{\alpha,n-k}(\mathsf{w})$$

for all $n \ge k \ge 1$ [☞ (5.16) and (5.18)]. Hence, by an appeal to Fatou's lemma,

$$W(\varnothing) \;=\; \lim_{n\to\infty} Y_{\alpha,n} \;\ge\; \sum_{|\mathsf{w}|=k} L(\mathsf{w})^{\alpha} \lim_{n\to\infty} Y_{\alpha,n-k} \;=\; \sum_{|\mathsf{w}|=k} L(\mathsf{w})^{\alpha} W(\mathsf{w}) \quad \text{a.s.}$$

for all $k \ge 1$. On the other hand, the previous inequality turns into an identity when taking expectations, for $\mathfrak{m}(\alpha) = 1$ and $\mathbb{E}W = 1$, and then also without doing this, i.e.

$$W(\varnothing) \;=\; \sum_{|\mathsf{w}|=k} L(\mathsf{w})^{\alpha} W(\mathsf{w}) \quad \text{a.s.}$$

for any $k \ge 1$, in particular $\mathscr{S}(G) = G$.

(b) First note that $\mathbb{E}W = 1$ implies $q < 1$. Using (a) and recalling our convention that $T_1 \ge T_2 \ge \ldots$, we then obtain

$$\begin{aligned} q \;&=\; \mathbb{P}\left(\sum_{i=1}^{N} T_i(\varnothing) W(i) = 0\right) \\ &=\; \sum_{n\ge 0} \mathbb{P}(N = n)\, \mathbb{P}(W(i) = 0 \text{ for } i = 1,\ldots,n) \\ &=\; \sum_{n\ge 0} \mathbb{P}(N = n)\, q^n \\ &=\; f(q), \end{aligned}$$

that is the fixed-point property of q. The convexity of f on $[0,1]$ in combination with $f(1) = \mathbb{P}(N < \infty) \le 1$ then also shows the uniqueness of q in $[0,1)$. $\square$

Turning to the core of the proof of $\mathbb{P}(\dim_H \mathfrak{C} \ge \alpha | \mathfrak{C} \ne \emptyset) = 1$, we will do so first under the additional condition

$$\exists \xi > 0 : \quad \inf_{1\le i\le N} T_i \;\ge\; \xi \quad \text{a.s.} \tag{RC-5}$$

As $Y_{d,1} = \sum_{i\ge 1} T_i(\varnothing)^d < 1$ [☞ after (5.21)], we then have

$$Y_{\theta,n} \;=\; \sum_{i\ge 1} T_i(\varnothing)^{\theta} \;\le\; \xi^{\theta-d} \sum_{i\ge 1} T_i(\varnothing)^d \;\le\; \xi^{\theta-d} \tag{5.71}$$

for all $0 \le \theta \le d$ and therefore $\|Y_{\theta,1}\|_p \le \|Y_{\theta,1}\|_{\infty} < \infty$ for all $p \ge 1$. In particular, $N = Y_{0,1} \le \xi^{-d}$ a.s. This allows an application of Lemmata 5.68 and 5.69.

We now define

$$\Lambda_n(A) \;:=\; \sum_{|\mathsf{v}|=n,\, A\cap J_{\mathsf{v}}\ne\emptyset} L(\mathsf{v})^{\alpha} W(\mathsf{v})$$

for $n \geq 0$ and $A \subset [0,1]^d$. Then

$$\begin{aligned}\Lambda_{n+1}(A) &= \sum_{|\mathsf{v}|=n} L(\mathsf{v})^\alpha \sum_{i\geq 1, A\cap J_{\mathsf{v}i}\neq\emptyset} T_i(\mathsf{v})^\alpha W(\mathsf{v}i)\\ &= \sum_{|\mathsf{v}|=n, A\cap J_{\mathsf{v}}\neq\emptyset} L(\mathsf{v})^\alpha \sum_{i\geq 1, A\cap J_{\mathsf{v}i}\neq\emptyset} T_i(\mathsf{v})^\alpha W(\mathsf{v}i)\\ &\leq \sum_{|\mathsf{v}|=n, A\cap J_{\mathsf{v}}\neq\emptyset} L(\mathsf{v})^\alpha \sum_{i\geq 1} T_i(\mathsf{v})^\alpha W(\mathsf{v}i)\\ &= \sum_{|\mathsf{v}|=n, A\cap J_{\mathsf{v}}\neq\emptyset} L(\mathsf{v})^\alpha W(\mathsf{v})\\ &= \Lambda_n(A),\end{aligned}$$

which implies the existence of

$$\Lambda(A) := \lim_{n\to\infty} \Lambda_n(A) = \inf_{n\geq 0} \Lambda_n(A)$$

for all $A \subset [0,1]^d$, called *random construction measure*. The following lemma provides an explanation for this name.

Lemma 5.70. *Under the additional assumption* (RC-5)*, the random set function*

$$\Lambda : \Omega \times \mathfrak{P}([0,1]^d) \to \mathbb{R}_\geq$$

a.s. defines a metric outer measure [☞ Definition D.1 in the Appendix], and it is a.s. a measure on $\mathscr{B}([0,1]^d)$ with $\Lambda([0,1]^d) = \Lambda(\mathfrak{C}) = W(\varnothing)$ and thus $\Lambda(\mathfrak{C}^c) = 0$. Furthermore,

$$\{\mathfrak{C} \neq \emptyset\} = \{\Lambda(\mathfrak{C}) > 0\} = \{W(\varnothing) > 0\} \quad a.s. \tag{5.72}$$

holds true.

Proof. We leave it to the reader to verify that $\Lambda(\omega,\cdot)$ forms an outer measure for each $\omega \in \Omega$ [☞ Problem 5.73]. Given two subsets A, B of $[0,1]^d$ with positive distance ε, choose $\tau = \tau(\omega)$ large enough to have $diam(J_{\mathsf{v}}(\omega)) < \varepsilon/2$ for all v of length $|\mathsf{v}| \geq \tau$. Then τ is a.s. finite because

$$\lim_{n\to\infty} \sup_{|\mathsf{v}|=n} L(\mathsf{v}) = 0 \quad \text{a.s.}$$

as we have shown already in the first half of the proof of Theorem 5.12 in Section 5.2. Consequently, any J_{v} with $|\mathsf{v}| \geq \tau$ can intersect at most one of the two sets A, B, giving

$$\Lambda_n(A\cup B) = \Lambda_n(A) + \Lambda_n(B) \quad \text{a.s.}$$

for all $n \geq \tau$. By letting n tend to ∞, we find $\Lambda(A\cup B) = \Lambda(A) + \Lambda(B)$ a.s. and have thus shown that Λ a.s. defines a metric outer measure. By Theorem D.2 in the

Appendix, it is then also a measure on $\mathscr{B}([0,1]^d)$ a.s. Moreover,

$$\Lambda([0,1]^d) \;=\; \lim_{n\to\infty}\Lambda_n([0,1]^d) \;=\; \sum_{|\mathsf{v}|=n} L(\mathsf{v})^{\alpha}W(\mathsf{v}) \;=\; W(\varnothing) \quad \text{a.s.}$$

For the verification of $\Lambda(\mathfrak{C}^c)=0$ a.s. we refer once again to Problem 5.73.

As for (5.72), note that $\mathfrak{C}\neq\emptyset$ holds true iff the GWP $(Y_{0,n})_{n\ge 0}$ survives, the probability being $1-q$ where q is the unique fixed point of $f(s)=\mathbb{E}s^N$ in $[0,1)$. But by what has been shown before, we also have that

$$\{\mathfrak{C}\neq\emptyset\} \;\supseteq\; \{\Lambda(\mathfrak{C})>0\} \;=\; \{W(\varnothing)>0\} \quad \text{a.s.,}$$

and $\mathbb{P}(W(\varnothing)>0)=1-q$ by Lemma 5.69(b). Hence, all events in (5.72) have the same probability and must therefore be a.s. equal. ☐

Lemma 5.71. *Under the additional assumption* (RC-5)*,*

$$\mathbb{P}\left(\sup_{|\mathsf{v}|=n} L(\mathsf{v})^{\alpha-\theta}W(\mathsf{v})>b \textit{ infinitely often}\right)=0 \tag{5.73}$$

holds true for any $\theta\in[0,\alpha)$ and $b>0$, whence

$$\kappa \;:=\; \sup\left\{n\ge 1:\ \sup_{|\mathsf{v}|\ge n} L(\mathsf{v})^{\alpha-\theta}W(\mathsf{v})>b\right\} \;<\; \infty \quad \textit{a.s.} \tag{5.74}$$

Proof. We remind the reader that $\|Y_{\theta,1}\|_p<\infty$ for all $\theta\in[0,d]$ and $p\ge 1$. Hence, by Lemma 5.68 and what has been pointed out after its proof, $\|W\|_p<\infty$ for all $p\ge 1$. In the following, let θ,b be fixed and $p\ge 1$ further below be suitably chosen. Using the Markov inequality in combination with the independence of $L(\mathsf{v}),W(\mathsf{v})$, we find

$$\mathbb{P}(L(\mathsf{v})^{\alpha-\theta}W(\mathsf{v})>b) \;\le\; \frac{\mathbb{E}L(\mathsf{v})^{p(\alpha-\theta)}\,\mathbb{E}W^p}{b^p}$$

for any $\mathsf{v}\in\mathbb{T}$ and thereupon

$$\sum_{|\mathsf{v}|=n}\mathbb{P}(L(\mathsf{v})^{\alpha-\theta}W(\mathsf{v})>b) \;\le\; \frac{\mathbb{E}W^p}{b^p}\sum_{|\mathsf{v}|=n}\mathbb{E}L(\mathsf{v})^{p(\alpha-\theta)} \;=\; \frac{\mathbb{E}W^p}{b^p}\,\mathfrak{m}(p(\alpha-\theta))^n$$

for any $n\ge 1$. Now fixing p large enough such that $\mathfrak{m}(p(\alpha-\theta))<1$, we finally arrive at

$$\sum_{n\ge 1}\mathbb{P}\left(\sup_{|\mathsf{v}|=n} L(\mathsf{v})^{\alpha-\theta}W(\mathsf{v})>b\right) \;\le\; \sum_{n\ge 1}\sum_{|\mathsf{v}|=n}\mathbb{P}(L(\mathsf{v})^{\alpha-\theta}W(\mathsf{v})>b)$$

$$\leq \frac{\mathbb{E}W^p}{b^p} \sum_{n \geq 1} \mathfrak{m}(p(\alpha - \theta))^n < \infty$$

which in turn implies (5.73) by an appeal to the Borel-Cantelli lemma. □

The next proposition is the crucial step towards our ultimate goal of this section.

Proposition 5.72. *Let $\theta \in [0, \alpha)$. Under the additional assumption* (RC-5)*, there exists a $\mathbb{P}$-null set $\mathscr{N}$ such that the implication*

$$\mathscr{H}_\theta(A) < \infty, \text{ i.e. } \dim_H A \leq \theta \quad \Longrightarrow \quad \Lambda(\omega, A) = 0$$

holds true for all $A \in \mathscr{B}([0,1]^d)$ and $\omega \in \mathscr{N}^c$.

Proof. Fixing any $\theta \in [0, \alpha)$ and $A \in \mathscr{B}([0,1]^d)$ with $\mathscr{H}_\alpha(A) < \infty$, let κ be defined as in (5.74) for an arbitrarily small $b > 0$. Then pick any positive

$$\varepsilon \leq 1 \wedge \min\{L(\mathsf{v}) : L(\mathsf{v}) > 0 \text{ and } |\mathsf{v}| \leq \kappa\}$$

and observe that this number is random. The following considerations are based on a pathwise analysis and thus valid only outside a $\mathbb{P}$-null set which, however, does not depend on A. Having this said, the annex "a.s." will be omitted hereafter.

Pick an arbitrary $(\varepsilon/2)$-covering $(A_i)_{i \geq 1}$ of A, consisting of closed balls. For each $i \geq 1$, we then find a random $r_i \in \mathbb{N}_0$ such that

$$2^{-1-r_i} \leq \mathit{diam}(A_i) < 2^{-r_i} \leq 2\,\mathit{diam}(A_i) \leq \varepsilon.$$

For any $x \in \mathfrak{C}$ and $r \in \mathbb{N}$, let $V_{x,r}$ be the unique node v of $\mathbb{T}$ with the properties

$$x \in J_\mathsf{v}, \quad L(\mathsf{v}) = \mathit{diam}(J_\mathsf{v}) < 2^{-r} \quad \text{and} \quad L(\mathsf{v}|k-1) \geq 2^{-r} \text{ if } |\mathsf{v}| = k.$$

Define the random sets

$$\mathscr{V}_i := \{V_{x,r_i} : x \in \mathfrak{C} \cap J_{V_{x,r_i}} \cap A_i\} \subset \{\mathsf{v} \in \mathbb{T} : |\mathsf{v}| > \kappa\}$$

for $i \geq 1$. Note for the stated inclusion that $|V_{x,r_i}| \leq \kappa$ is indeed impossible, for otherwise (recall from above the choice of ε)

$$\varepsilon \leq \mathit{diam}(J_{V_{x,r_i}}) < 2^{-r_i} \leq \varepsilon.$$

For all distinct $\mathsf{v}, \mathsf{w} \in \mathscr{V}_i$, we further have either $J_\mathsf{v} = J_\mathsf{w}$ or $\mathit{int}(J_\mathsf{v}) \cap \mathit{int}(J_\mathsf{w}) = \emptyset$, because the remaining alternative $J_\mathsf{w} \subset J_\mathsf{v}$ (w.l.o.g. $|\mathsf{v}| < |\mathsf{w}| = k$) would yield $\mathsf{w} = \mathsf{vu}$ for some $\mathsf{u} \in \mathbb{T} \backslash \{\varnothing\}$ and then the contradiction

$$\mathit{diam}(J_{\mathsf{w}|k-1}) \leq \mathit{diam}(J_\mathsf{v}) < 2^{-r_i}.$$

Therefore, we conclude that $\{J_{\mathsf{v}} : \mathsf{v} \in \mathscr{V}_i\}$ forms a family of nonoverlapping sets satisfying

$$|x-y| \le diam(J_{\mathsf{v}}) + diam(A_i) < 2^{1-r_i}$$

for any choice $y \in A_i$ and $x \in J_{\mathsf{v}}$, $\mathsf{v} \in \mathscr{V}_i$, as $A_i \cap J_{\mathsf{v}} \neq \emptyset$. This yields

$$\bigcup_{\mathsf{v}\in\mathscr{V}_i} J_{\mathsf{v}} \subset \mathbb{B}(y, 2^{1-r_i}) \tag{5.75}$$

for all $y \in A_i$, where $\mathbb{B}(y,r)$ denotes the closed ball centered at y with radius r. Now condition (RC-5) comes once again into play and provides us with

$$diam(J_{\mathsf{v}}) = T_{\mathsf{v}_n}(\mathsf{v}|n-1)\, diam(J_{\mathsf{v}|n-1}) \ge \xi\, 2^{-r_i}$$

for any $\mathsf{v} = \mathsf{v}_1...\mathsf{v}_n \in \mathscr{V}_i$ and then [☞ also (5.21)]

$$\frac{\lambda\!\!\lambda^d(int(J_{\mathsf{v}}))}{\lambda\!\!\lambda^d(int(J))} = \frac{diam(J_{\mathsf{v}})^d}{diam(J)^d} \ge \left(\frac{\xi 2^{-r_i}}{diam(J)}\right)^d. \tag{5.76}$$

Setting $N_i := |\mathscr{V}_i|$, we obtain with the help of (5.75) and (5.76) that

$$\begin{aligned} N_i \left(\frac{\xi 2^{-r_i}}{diam(J)}\right)^d \lambda\!\!\lambda^d(int(J)) &\le \sum_{\mathsf{v}\in\mathscr{V}_i} \lambda\!\!\lambda^d(int(J_{\mathsf{v}})) \\ &\le \lambda\!\!\lambda^d(\mathbb{B}(y, 2^{1-r_i})) \\ &= 2^{-r_i d}\, \lambda\!\!\lambda^d(\mathbb{B}(0,2)) \end{aligned}$$

and thereby the estimate

$$N_i \le K := \left(\frac{diam(J)}{\xi}\right)^d \frac{\lambda\!\!\lambda^d(\mathbb{B}(0,2))}{\lambda\!\!\lambda^d(int(J))} < \infty \tag{5.77}$$

for all $i \ge 1$ and some constant K which, unlike the random variables N_i, is fixed and not depending on $b > 0$, nor on the particular choice of $(A_i)_{i\ge 1}$.

Using $\mathscr{V}_i \subset \{\mathsf{v} : |\mathsf{v}| > \kappa\}$ and the definition of κ, we infer

$$L(\mathsf{v})^\alpha W(\mathsf{v}) \le bL(\mathsf{v})^\theta \le b2^{-\theta r_i} \le b2^\theta diam(A_i)^\theta$$

for all $\mathsf{v} \in \mathscr{V}_i$ and then upon summation

$$\sum_{\mathsf{v}\in\mathscr{V}_i} L(\mathsf{v})^\alpha W(\mathsf{v}) \le bK 2^\theta diam(A_i)^\theta$$

for any $i \ge 1$. Put $n(i) := \max\{|\mathsf{v}| : \mathsf{v} \in \mathscr{V}_i\}$. Then

$$\begin{aligned} \Lambda(A \cap A_i) &= \Lambda(A \cap A_i \cap \mathfrak{C}) \\ &\le \Lambda_{n(i)}(A \cap A_i \cap \mathfrak{C}) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathsf{v}\in\mathbb{T}:|\mathsf{v}|=n(i),J_{\mathsf{v}}\cap A\cap A_i\cap\mathfrak{C}\neq\emptyset} L(\mathsf{v})^{\alpha}W(\mathsf{v})\\
&\le \sum_{\mathsf{v}\in\mathscr{V}_i} L(\mathsf{v})^{\alpha}W(\mathsf{v})\\
&\le bK2^{\theta}\mathit{diam}(A_i)^{\theta},
\end{aligned}
$$

where, for the penultimate line, we have utilized Lemma 5.69(a) and the fact that any vertex v of length $n(i)$ and with $J_{\mathsf{v}}\cap A_i\cap\mathfrak{C}\neq\emptyset$ must be in $\mathscr{V}_i$ or stemming from an element of $\mathscr{V}_i$ (by the defining property of $n(i)$). Summing over $i\ge 1$, we finally arrive at

$$\Lambda(A) \ \le\ \sum_{i\ge1}\Lambda(A\cap A_i) \ \le\ bK2^{\theta}\sum_{i\ge1}\mathit{diam}(A_i)^{\theta}$$

and thus

$$\Lambda(A) \ \le\ bK2^{\theta}\,\mathscr{H}_{\theta}(A) \ <\ \infty.$$

Since b was arbitrary and K independent of b, the proof is complete. □

Now we are in the position to complete the proof of Theorem 5.12.

Proof (of Theorem 5.12, 2^{nd} half). Pick any $\theta\in[0,\alpha)$. Under the additional assumption (RC-5), we infer with the help of the previous proposition that $\mathscr{H}_{\theta}(\mathfrak{C}(\omega))=\infty$ and thus $\dim_H(\mathfrak{C}(\omega))\ge\alpha$ holds true for $\mathbb{P}$-almost all $\omega\in\{\Lambda(\mathfrak{C})>0\}$, which is the desired conclusion as the last event a.s. equals $\{\mathfrak{C}\neq\emptyset\}$ by (5.72).

In order to get rid of assumption (RC-5), the obvious idea is to pick an arbitrarily small $\xi>0$ and consider the random recursive construction based on the family

$$\{J_{\mathsf{v}}^{\xi}:\mathsf{v}\in\mathbb{T}\},$$

where $J_{\varnothing}^{\xi}:=J_{\varnothing}=J$ and, for $\mathsf{v}=\mathsf{v}_1...\mathsf{v}_n\in\mathbb{T}\backslash\{\varnothing\}$,

$$J_{\mathsf{v}}^{\xi} \ :=\ \begin{cases} J_{\mathsf{v}}, & \text{if } T_{\mathsf{v}_1}(\varnothing)\wedge T_{\mathsf{v}_2}(\mathsf{v}_1)\wedge\ldots\wedge T_{\mathsf{v}_n}(\mathsf{v}_1...\mathsf{v}_{n-1})\ge\xi,\\ \emptyset, & \text{otherwise.}\end{cases}$$

Define $T^{\xi}(\mathsf{v}):=\big(T_i(\mathsf{v})\mathbf{1}_{\{T_i(\mathsf{v})\ge\xi\}}\big)_{i\ge1}$ with generic copy $T^{\xi}=\big(T_i^{\xi}\big)_{i\ge1}$ and then the corresponding $L^{\xi}(\mathsf{v}),N^{\xi}(\mathsf{v})$ in the obvious manner. Recall that $N^{\xi}\le\xi^{-d}$ a.s. and put further

$$\mathfrak{m}^{\xi}(\theta) \ :=\ \sum_{i\ge1}\mathbb{E}\big(T_i^{\xi}\big)^{\theta}$$

for $\theta\ge0$ with associated unique α^{ξ} satisfying $\mathfrak{m}^{\xi}(\alpha^{\xi})=1$ for any ξ small enough such that $\mathbb{E}N^{\xi}>1$. Plainly,

$$\mathbb{E}N^{\xi}\uparrow\mathbb{E}N,\quad \mathfrak{m}^{\xi}(\cdot)\uparrow\mathfrak{m}(\cdot)\text{ pointwise},\quad\text{and}\quad \alpha^{\xi}\uparrow\alpha$$

as $\xi\downarrow0$. Furthermore,

$$\mathfrak{C}^{\xi} := \bigcap_{n\geq 0}\bigcup_{|\mathsf{v}|=n} J_{\mathsf{v}}^{\xi} \subset \mathfrak{C}$$

and therefore, by what has been already shown,

$$\dim_H \mathfrak{C} \geq \dim_H \mathfrak{C}^{\xi} = \alpha^{\xi} \quad \text{a.s. on } \{\mathfrak{C}^{\xi} \neq \emptyset\}$$

for all $\xi > 0$. By letting ξ tend to 0, it follows $\dim_H \mathfrak{C} \geq \alpha$ a.s. on the event

$$\lim_{\xi\downarrow 0}\{\mathfrak{C}^{\xi} \neq \emptyset\} = \bigcup_{\xi>0}\{\mathfrak{C}^{\xi} \neq \emptyset\} \subset \{\mathfrak{C} \neq \emptyset\},$$

which leaves us with a proof of the last inclusion to be an identity. Obviously, this follows if

$$q^{\xi} := \mathbb{P}(\mathfrak{C}^{\xi} = \emptyset) \to \mathbb{P}(\mathfrak{C} = \emptyset) =: q \quad \text{as } \xi \to 0.$$

But we have that $q = \mathbb{P}(Y_{0,n} = 0 \text{ eventually})$. It is therefore the unique fixed point of the gf f of N in $[0,1)$, while q^{ξ}, for ξ sufficiently small, is the unique fixed point of the gf f^{ξ}, say, of N^{ξ} in $[0,1)$ [☞ (5.72) in Lemma 5.70 and 5.69(b)]. Since $N^{\xi} \uparrow N$ and thus $f^{\xi} \downarrow f$ on $[0,1]$, we finally arrive at the desired conclusion. □

Problems

Problem 5.73. For the random set function Λ considered in Lemma 5.70, prove the following assertions:

(a) $\Lambda(\omega,\cdot)$ is an outer measure for any $\omega \in \Omega$.
(b) $\Lambda(\omega,\mathfrak{C}^c) = 0$ for $\mathbb{P}$-almost all $\omega \in \Omega$.

Chapter 6
The contraction method for a class of distributional recursions

The probabilistic performance analysis of `Quicksort` described in Section 5.6 can be seen as a paradigm for a whole class of similar problems dealing with distributional recursions arising in the analysis of divide and conquer algorithms, random data structures, random trees or combinatorial probability. A review of the `Quicksort` analysis reveals that it actually involves *two* contraction arguments:

- the first one concerns the smoothing transform $\mathscr{S}$, say, related to the `Quicksort` equation (5.66) and ensures that $\mathscr{S}$ has a unique fixed point in the class $\mathscr{P}_0^2(\mathbb{R})$ which thus qualifies as the limit law of the quantity of interest,
- the second and more delicate one provides a proof of the `Quicksort` equation being indeed the limit of the finite-size model equation (5.64) which bears the intrinsic recursive structure of the algorithm.

This approach works in many other applications as well. As to the first argument, requiring information on the contraction properties of some smoothing transform, one can draw on the results of the previous chapter, based to a fair extent on the work in [97]. The present chapter aims at a discussion of the second argument in a quite general framework. It is nowadays known as the *contraction method*, a name coined by RACHEV & RÜSCHENDORF [95]. For a suitable class of model equations, which naturally bear the typical distributional recursion particularly present in the `Quicksort` example, we will derive two contraction theorems due to NEININGER & RÜSCHENDORF [92, 93] and then give a number of applications ranging from random recursive trees, m-ary search trees to `Quickselect` (also known as FIND) as another example of a dived-and-conquer algorithm. For related work see also the article [95] mentioned above, [91], and the surveys [99, 98]. After the introduction of the basic setup in the next section, the *Zolotarev metric* is defined and studied to some extent [☞ Sections 6.2, 6.3]. It forms an important tool, alternative to the minimal L^s-metric and better suited to situations where convergence is to be proved for random variables with specified moments up to some integral order m like in the frequently encountered case of normalized random variables with mean zero and variance one ($m = 2$). The proofs of the main contraction theorems, derived in Sections 6.4 and 6.5, provide illuminating examples.

6.1 The setup: a distributional recursion of general additive type

Consider a sequence $(X_n)_{n\ge 0}$ of real-valued random variables, which satisfies the random recursive equation

$$X_n \stackrel{d}{=} \sum_{i=1}^{N_n} A_{n,i} X_{i,\tau_{n,i}} + B_n \tag{6.1}$$

for $n \ge n_0 \in \mathbb{N}_0$, where

(A1) $B_n, A_{n,1}, A_{n,2}, \ldots$ are real-valued random variables such that $A_{n,i} > 0$ for $1 \le i \le N_n$ and $A_{i,n} = 0$ otherwise, thus

$$N_n := \sum_{i\ge 1} \mathbf{1}_{\{A_{n,i}\ne 0\}}.$$

(A2) $\tau_{n,1}, \tau_{n,2}, \ldots$ are random variables taking values in $\{0, \ldots, n\}$.
(A3) $\mathscr{L}((X_{i,n})_{n\ge 0}) = \mathscr{L}((X_n)_{n\ge 0})$ for each $i \ge 1$.
(A4) $(B_n, (A_{n,1}, \tau_{n,1}), (A_{n,2}, \tau_{n,2}), \ldots)$, $(X_{1,n})_{n\ge 0}$, $(X_{2,n})_{n\ge 0}$,... are independent.

Like in the `Quicksort` example, X_n typically represents a random quantity of interest for a random discrete structure parametrized by n as a measure of its size or complexity. The intrinsic nature of random structures satisfying (6.1) is that they may be divided into N_n substructures of similar kind but random size or cardinality $\tau_{n,i}$. In the `Quicksort` example, $N_n = 2$ and $\tau_{n,1} = Z_n - 1$, $\tau_{n,2} = n - Z_n$, are the lengths of the two sublists created by comparison with the pivot Z_n.

Defining the normalization

$$\widehat{X}_n := \frac{X_n - \mu_n}{\sigma_n}$$

for some $(\mu_n, \sigma_n^2) \in \mathbb{R} \times \mathbb{R}_>$, we may rewrite (6.1) as

$$\widehat{X}_n \stackrel{d}{=} \sum_{i=1}^{N_n} T_{n,i} \widehat{X}_{i,\tau_{n,i}} + C_n \tag{6.2}$$

for $n \ge n_0$, where $\widehat{X}_{i,n}$ has the obvious meaning,

$$T_{n,i} := \frac{\sigma_{\tau_{n,i}}}{\sigma_n} A_{n,i} \quad \text{and} \quad C_n := \frac{1}{\sigma_n}\left(B_n - \mu_n + \sum_{i=1}^{N_n} T_{n,i}\, \mu_{\tau_{n,i}}\right).$$

Of course, if $\mathbb{P}(N_n = \infty) > 0$, we tacitly assume that $\sum_{i=1}^{N_n} T_{n,i} \widehat{X}_{i,\tau_{n,i}}$ and $\sum_{i=1}^{N_n} T_{n,i}\, \mu_{\tau_{n,i}}$ are well-defined on $\{N_n = \infty\}$. In the common situation that X_n is square-integrable, we essentially choose μ_n as the mean of X_n and σ_n^2 as its variance, entailing $\mathscr{L}(\widehat{X}_n) \in \mathscr{P}^2_{0,1}(\mathbb{R})$. Our goal, to be accomplished in Section 6.4, is to prove a convergence result for $\mathscr{L}(\widehat{X}_n)$ by means of a contraction argument. Naturally, this requires appropriate assumptions about the input parameters $T_{n,i}, \tau_{n,i}$ and C_n in equation (6.2) as well as an appropriate class of metrics to be introduced next [Section 6.2] and then discussed in comparison with the already known minimal L^p-metrics ℓ_p [Section 6.3].

6.2 The Zolotarev metric

As usual, let $\mathscr{C}^0(\mathbb{R})$ denote the space of continuous functions $f:\mathbb{R}\to\mathbb{R}$ and $\mathscr{C}^m(\mathbb{R})$ for $m\in\mathbb{N}$ the subspace of m times continuously differentiable real- or complex-valued functions. For $s=m+\alpha$ with $m\in\mathbb{N}_0$ and $0<\alpha\le 1$, put

$$\mathfrak{F}_s := \left\{ f\in\mathscr{C}^m(\mathbb{R}) : |f^{(m)}(x)-f^{(m)}(y)| \le |x-y|^\alpha \text{ for all } x,y\in\mathbb{R}\right\}. \tag{6.3}$$

which obviously contains the monomials $x\mapsto x^k$ for $k=1,\dots,m$ as well as $x\mapsto sign(x)|x|^s/c_s$ and $x\mapsto |x|^s/c_s$ for some $c_s\in\mathbb{R}_>$ [☞ Problem 6.5]. Finally, if $s>1$ and thus $m\in\mathbb{N}$, then denote by $\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$, $\mathbf{z}=(z_1,\dots,z_m)\in\mathbb{R}^m$, the set of distributions on $\mathbb{R}$ having k^{th} moment z_k for $k=1,\dots,m$.

ZOLOTAREV [116] introduced the metric ζ_s on $\mathscr{P}^s(\mathbb{R})$, defined by

$$\zeta_s(F,G) := \sup_{f\in\mathfrak{F}_s,\,(X,Y)\sim(F,G)} \left|\mathbb{E}\big(f(X)-f(Y)\big)\right| \tag{6.4}$$

and nowadays named after him. Via a Taylor expansion of the functions $f\in\mathfrak{F}_s$ in (6.4), it can be shown that $\zeta_s(F,G)$ is finite for all $F,G\in\mathscr{P}^s(\mathbb{R})$ if $0<s\le 1$, and for all $F,G\in\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$ and $\mathbf{z}\in\mathbb{R}^m$ if $s>1$ [☞ Problem 6.6]. On the other hand, in the last case $\zeta_s(F,G)=\infty$ for distributions $F,G\in\mathscr{P}^s(\mathbb{R})$ that do not have the same integral moments up to order m [☞ Problem 6.7]. We thus see that ζ_s defines a proper probability metric on $\mathscr{P}^s(\mathbb{R})$ only for $0<s\le 1$ and on $\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$ for any $\mathbf{z}\in\mathbb{R}^m$, otherwise. Here we should add that $\zeta_s(F,G)=0$ implies $F=G$ because $\mathscr{C}^m_b(\mathbb{R}):=\{f\in\mathscr{C}^m(\mathbb{R}): f^{(m)} \text{ is bounded}\}$ is a measure determining class for each $m\in\mathbb{N}_0$ [☞ Problem 6.8].

Given a probability space $(\Omega,\mathfrak{A},\mathbb{P})$, ζ_s can also be defined on $L^s=L^s(\mathbb{P})$, viz.

$$\zeta_s(X,Y) := \sup_{f\in\mathfrak{F}_s} \left|\mathbb{E}\big(f(X)-f(Y)\big)\right|, \tag{6.5}$$

and constitutes a *pseudo-metric* there if $0<s\le 1$. If $p>1$, then this is true only on $L^s_{\mathbf{z}}=L^s_{\mathbf{z}}(\mathbb{P}):=\{X\in L^s(\mathbb{P}):\mathbb{E}X^k=z_k \text{ for } k=1,\dots,m\}$ for any $\mathbf{z}\in\mathbb{R}^m$. Recall that a pseudo-metric has the same properties as a metric with one exception: $\zeta_s(X,Y)=0$ does not necessarily imply $X=Y$ (here not even with probability one: just take two iid X,Y which are not a.s. constant).

A pseudo-metric ρ on a set of random variables is called *simple* if it depends only on the marginals of the random variables being compared, and *compound* otherwise.

It is called *ideal of order s* if

$$\rho(cX,cY) \;=\; |c|^s \rho(X,Y) \tag{6.6}$$

for all $c \in \mathbb{R}$ and

$$\rho(X+Z,Y+Z) \;\le\; \rho(X,Y) \tag{6.7}$$

for any Z independent of X,Y and with well-defined $\rho(X+Z,Y+Z)$. Obviously, ζ_s is simple, namely

$$\zeta_s(X,Y) \;=\; \zeta_s(F,G)$$

for any random variables X,Y with respective laws F,G, whereas the L^p-pseudo-metrics ℓ_p encountered in the previous chapter [☞ (5.30)] are compound. It will be shown in Proposition 6.1(a) below that ζ_s is also ideal of order s on any $L^s_{\mathbf{z}}$ for $\mathbf{z} \in \mathbb{R}^m$.

Another example of a compound pseudo-metric on L^s is given by

$$\kappa_s(X,Y) \;:=\; \mathbb{E}\,|\mathrm{sign}(X)|X|^s - \mathrm{sign}(Y)|Y|^s| \;=\; \mathbb{E}\left|X\,|X|^{s-1} - Y\,|Y|^{s-1}\right|,$$

and, in the same manner as ℓ_s, it induces a metric on $\mathscr{P}^s(\mathbb{R})$, called *minimal metric associated with* κ_s, namely

$$\kappa_s(F,G) \;:=\; \inf_{(X,Y)\sim(F,G)} \kappa_s(X,Y),$$

where the infimum is taken over all possible (F,G)-couplings defined on an *arbitrary* probability space. Proposition 6.1(c) will show that $\zeta_s(X,Y)$ can be bounded in terms of $\kappa_s(X,Y)$ for random variables $X,Y \in L^s$ having equal moments of integral order $k=1,\ldots,m$. Part (d) will show such a bound in terms of $\ell_s(X,Y)$.

In the following , it is stipulated that $\mathscr{P}^s_*(\mathbb{R})$, L^s_* stand for $\mathscr{P}^s(\mathbb{R})$, L^s if $0<p\le 1$, and for $\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$, $L^s_{\mathbf{z}}$ for arbitrary $\mathbf{z} \in \mathbb{R}^m$ if $s>1$.

Proposition 6.1. *Let $s=m+\alpha$ for some $m \in \mathbb{N}_0$ and $0<\alpha\le 1$. Then ζ_s, defined by* (6.4) *or* (6.5)*, has the following properties:*

(a) *ζ_s is an ideal pseudo-metric on L^s_*.*

(b) *For any $X,Y \in L^s$ and with $c_s := \prod_{k=1}^m (\alpha+k)$ [$:=1$ if $m=0$],*

$$c_s\,\zeta_s(X,Y) \;\ge\; \mathbb{E}|X|^s - \mathbb{E}|Y|^s. \tag{6.8}$$

(c) *For any $X,Y \in L^s_*$,*

$$\begin{aligned}\zeta_s(X,Y) \;&\le\; \frac{\Gamma(1+\alpha)}{\Gamma(1+s)}\,\Lambda_s(X,Y),\\ \Lambda_s(X,Y) \;&:=\; 2m\,\kappa_s(X,Y) + \big(2\kappa_s(X,Y)\big)^{\alpha}\big(\mathbb{E}|X|^s \wedge \mathbb{E}|Y|^s\big)^{1-\alpha},\end{aligned} \tag{6.9}$$

in particular

$$\zeta_s(X,Y) \;\le\; \frac{2}{(s-1)!}\,\kappa_s(X,Y) \tag{6.10}$$

if $\alpha = 1$ *and thus* $s = m+1 \in \mathbb{N}$.

(d) *For any* $X,Y \in L_*^s$,

$$\zeta_s(X,Y) \;\le\; \frac{\Gamma(1+\alpha)}{\Gamma(1+s)}\,\Theta_s(X,Y), \tag{6.11}$$

where $\Theta_s(X,Y) := \ell_s(X,Y)$ *if* $0 < s = \alpha \le 1$, *and*

$$\Theta_s(X,Y) \;:=\; \ell_s(X,Y)^{\alpha}\,\|X\|_s^m + m\,\ell_s(X,Y)\,(\ell_s(X,Y)+\|Y\|_s)^{m-1}$$

if $s \ge 1$.

Proof. (a) Property (6.6) follows directly from the fact that

$$\{|a|^{-s} f_{a,b} : f \in \mathfrak{F}_s\} \;=\; \mathfrak{F}_s$$

for any $(a,b) \in (\mathbb{R}\backslash\{0\}) \times \mathbb{R}$, where $f_{a,b}(x) := f(ax+b)$. This is also used in the following inequality giving the first half of property (6.7).

$$\begin{aligned}
\zeta_s(X+Z,Y+Z) \;&=\; \sup_{f\in\mathfrak{F}_s} \big|\mathbb{E}\big(f(X+Z)-f(Y+Z)\big)\big| \\
&\le\; \int \sup_{f\in\mathfrak{F}_s} \big|\mathbb{E}\big(f_{1,z}(X)-f_{1,z}(Y)\big)\big|\,\mathbb{P}(Z\in dz) \\
&=\; \int \sup_{f\in\mathfrak{F}_s} \big|\mathbb{E}\big(f(X)-f(Y)\big)\big|\,\mathbb{P}(Z\in dz) \\
&=\; \zeta_s(X,Y).
\end{aligned}$$

(b) This is immediate because $f(x) = |x|^s/c_s$ is in $\mathfrak{F}_s$ [☞ Problem 6.5].

(c) Suppose first that $m \ge 1$. Putting $\Delta_f(x) = f^{(m)}(x) - f^{(m)}(0)$, we then have the Taylor expansion

$$f(x) \;=\; \sum_{k=0}^{m} \frac{f^{(k)}(0)}{k!}x^k + \int_0^1 \frac{(1-t)^{m-1}}{(m-1)!}\Delta_f(tx)\,x^m\,dt, \quad x\in\mathbb{R},$$

with the help of which we infer

$$\begin{aligned}
&\big|\mathbb{E}\big(f(X)-f(Y)\big)\big| \\
&\qquad\le\; \int_0^1 \frac{(1-t)^{m-1}}{(m-1)!}\,\mathbb{E}\left|\Delta_f(tX)X^m - \Delta_f(tY)Y^m\right|\,dt
\end{aligned}$$

$$\begin{aligned}
&\le \int_0^1 \frac{(1-t)^{m-1}}{(m-1)!}\mathbb{E}\left|f^{(m)}(tX)-f^{(m)}(tY)\right| |X|^m\,dt\\
&\quad + \int_0^1 \frac{(1-t)^{m-1}}{(m-1)!}\mathbb{E}\left|\Delta_f(tY)\right| \left||X|^m-|Y|^m\right|\,dt\\
&\le \frac{\Gamma(\alpha+1)}{\Gamma(s+1)}\left(\mathbb{E}|X-Y|^\alpha\,|X|^m+\mathbb{E}|Y|^\alpha\left||X|^m-|Y|^m\right|\right) \qquad (6.12)
\end{aligned}$$

where the Hölder continuity of $f^{(m)}$ and

$$\int_0^1 \frac{(1-t)^{m-1}t^\alpha}{(m-1)!}\,dt \;=\; \frac{\Gamma(\alpha+1)}{\Gamma(s+1)}$$

have been utilized for the final estimate. Now observe that (6.12) remains obviously valid if $m=0$.

The rest of the proof consists in the derivation of bounds for $\mathbb{E}|X-Y|^\alpha\,|X|^m$ and $\mathbb{E}|Y|^\alpha\,||X|^m-|Y|^m|$. Let us start by noting that, if (w.l.o.g.) $|X|^{s-1}>|Y|^{s-1}$,

$$\begin{aligned}
|X-Y|\,|X|^{s-1} &= \left|X|X|^{s-1}-Y|X|^{s-1}\right|\\
&\le \left|X|X|^{s-1}-Y|Y|^{s-1}\right|+|Y|\left(|X|^{s-1}-|Y|^{s-1}\right)\\
&\le 2\left|X|X|^{s-1}-Y|Y|^{s-1}\right|,
\end{aligned}$$

where the last line follows as $|Y|\left(|X|^{s-1}-|Y|^{s-1}\right)\le |X|^s-|Y|^s\le \left|X|X|^{s-1}-Y|Y|^{s-1}\right|$. Then use Hölder's inequality and this estimate to obtain

$$\begin{aligned}
\mathbb{E}|X-Y|^\alpha\,|X|^m &= \mathbb{E}|X-Y|^\alpha\,|X|^{\alpha(s-1)}|X|^{s(1-\alpha)}\\
&\le \left(\mathbb{E}|X-Y|\,|X|^{s-1}\right)^\alpha\left(\mathbb{E}|X|^s\right)^{1-\alpha}\\
&\le \left(2\,\kappa_s(X,Y)\right)^\alpha\left(\mathbb{E}|X|^s\right)^{1-\alpha}.
\end{aligned}$$

Finally left with $\mathbb{E}|Y|^\alpha\,(|X|^m-|Y|^m)$ (only if $m\ge 1$), use

$$\begin{aligned}
|Y|^\alpha\,(|X|^m-|Y|^m) &= |Y|^\alpha\left||X|-|Y|\right|\sum_{k=0}^{m-1}|X|^k\,|Y|^{m-1-k}\\
&\le |X-Y|\sum_{k=0}^{m-1}|X|^k\,|Y|^{s-1-k}\\
&\le m\,|X-Y|\,|X|^{s-1}\\
&\le m\left|X\,|X|^{s-1}-Y\,|Y|^{s-1}\right|+m\,|Y|\left(|X|^{s-1}-|Y|^{s-1}\right)\\
&\le 2m\left|X\,|X|^{s-1}-Y\,|Y|^{s-1}\right|
\end{aligned}$$

on the event $\{|X|\ge|Y|\}$ and a similar estimate on $\{|X<|Y|\}$ to infer that

$$\mathbb{E}|Y|^\alpha\left||X|^m-|Y|^m\right| \;\le\; 2m\,\kappa_s(X,Y).$$

By combining the previous estimates, the asserted inequality (6.9) follows.

(d) If $0 < s = \alpha \le 1$, then the assertion is immediate from

$$\zeta_s(X,Y) = \sup_{f\in\mathfrak{F}_s} |\mathbb{E}(f(X)-f(Y))| \le \mathbb{E}|X-Y|^s = \ell_s(X,Y).$$

So let $s > 1$ and thus $m \ge 1$ hereafter. We will once again make use of (6.12) by providing suitable estimates for the two expectations there. First, by an appeal to Hölder's inequality,

$$\mathbb{E}|X-Y|^\alpha |X|^m \le (\mathbb{E}|X-Y|^s)^{\alpha/s} (\mathbb{E}|X|^s)^{m/s} = \ell_s(X,Y)^\alpha \|X\|_s^m$$

which equals the first term in the definition of $\Theta_s(X,Y)$. Second, put $Z = X\,\mathbf{1}_{\{|X|\le|Y|\}} + Y\,\mathbf{1}_{\{|X|>|Y|\}}$ and use

$$|Y|^\alpha \big||X|^m - |Y|^m\big| \le |Y|^\alpha \left((|X-Y|+|Z|)^m - |Z|^m\right)$$

to infer, again using Hölder's inequality,

$$\begin{aligned}
\mathbb{E}|Y|^\alpha \big||X|^m - |Y|^m\big| &= \mathbb{E}|Y|^\alpha \left(\sum_{k=1}^m \binom{m}{k} |X-Y|^k \, |Z|^{m-k}\right) \\
&\le \sum_{k=1}^m \binom{m}{k} \mathbb{E}|X-Y|^k \, |Y|^{s-k} \\
&\le \sum_{k=1}^m \binom{m}{k} \ell_s(X,Y)^m \, \|Y\|_s^{s-k} \\
&= (\ell_s(X,Y) + \|Y\|_s)^m - \|Y\|_s^m \\
&\le m\,\ell_s(X,Y)\,(\ell_s(X,Y) + \|Y\|_s)^{m-1}.
\end{aligned}$$

In combination with the first estimate above, the assertion follows. □

Convergence with respect to the Zolotarev metric is characterized by our second proposition.

Proposition 6.2. *Under the same assumptions as in the previous result, the following properties hold true for ζ_s:*

(a) *$\zeta_s(F_n,F) \to 0$ implies $\ell_s(F_n,F) \to 0$ and thus particularly $F_n \xrightarrow{w} F$ for any $F, F_1, F_2, \ldots \in \mathscr{P}_*^s(\mathbb{R})$.*

(b) *Conversely, $\ell_s(F_n,F) \to 0$ implies $\kappa_s(F_n,F) \to 0$ and therefore, by* (6.9)*, $\zeta_s(F_n,F) \to 0$ for any $F, F_1, F_2, \ldots \in \mathscr{P}_*^s(\mathbb{R})$.*

(c) *The metric space $(\mathscr{P}_*^s(\mathbb{R}), \zeta_s)$ is complete.*

Proof. (a) If $\zeta_s(F_n,F)\to 0$ for $F,F_1,F_2,...\in\mathscr{P}_*^s(\mathbb{R})$ and $X,X_1,X_2,...$ are such that $\mathscr{L}(X)=F$ and $\mathscr{L}(X_n)=F_n$ for $n\geq 1$, then these random variables are L^s-bounded because, by (6.8),

$$\sup_{n\geq 1}\mathbb{E}|X_n|^s\ \leq\ \sup_{n\geq 1}\zeta_s(F_n,F)\ +\ \mathbb{E}|X|^s$$

Hence the assertion follows from Proposition 6.1(d).

(b) Conversely, if $\ell_s(F_n,F)\to 0$, then $F_n^{-1}(U)\to F^{-1}(U)$ a.s. and thus also $X_n:=$ $sign(F_n^{-1}(U))|F_n^{-1}(U)|^s\to sign(F^{-1}(U))|F^{-1}(U)|^s=:X$ a.s. for any $Unif(0,1)$ random variable U. Furthermore, we have $\mathbb{E}|X_n|=\mathbb{E}|F_n^{-1}(U)|^s\to\mathbb{E}|F^{-1}(U)|^s=\mathbb{E}|X|$. But a combination of these facts implies $\kappa_s(F_n,F)\leq\|X_n-X\|_1\to 0$ due to a theorem by F. Riesz [☞ [12, Thm. 15.4], stated there for $s\geq 1$, but easily extended to all $s>0$].

(c) Now consider a Cauchy sequence $(F_n)_{n\geq 1}$ in $(\mathscr{P}_*^s(\mathbb{R}),\zeta_s)$ with associated sequence of chf's $(\varphi_n)_{n\geq 1}$. Since $x\mapsto\gamma_t\cos tx$ and $x\mapsto\gamma_t\sin tx$ are elements of $\mathfrak{F}_s$ for suitable $\gamma_t>0$, we infer that $(\varphi_n)_{n\geq 1}$ is (pointwise) Cauchy as well and thus converges to a function φ. It is further readily seen as that $\sup_{n\geq 1}\zeta_s(F_n,F_1)<\infty$ which, by (6.8), entails that the F_n have uniformly bounded absolute moments of order s and are therefore tight. Consequently, any subsequence contains a further subsequence which is weakly convergent. But their chf's converge pointwise to φ, implying that φ is a chf of a distribution F and $F_n\stackrel{w}{\to}F$. Then, by Fatou's lemma,

$$\mathbb{E}|F^{-1}(U)|^s\ \leq\ \liminf_{n\to\infty}\mathbb{E}|F_n^{-1}(U)|^s\ <\ \infty$$

and thus $F\in\mathscr{P}^s(\mathbb{R})$.

Finally, we must verify $\zeta_s(F_n,F)\to 0$. Put $X_n:=F_n^{-1}(U)$ and $X:=F^{-1}(U)$. In view of (6.9), it suffices to show $\kappa_s(X_n,X)\to 0$ and, for $p>1$, that X has the same moments up to order m as all X_n. But this follows if $(|X_n|^s)_{n\geq 1}$ is ui, for $X_n\to X$ a.s. and thus $sign(X_n)|X_n|^s\to sign(X)|X|^s$ a.s. We make use of the following technical result taken from [38, Lemmata 5.2 and 5.3]: There exists an even function $f\in\mathfrak{F}_s$ such that

(1) $f(x)=\psi(|x|)$ for some $\psi\in\mathscr{C}^\infty(\mathbb{R})$.
(2) $f(x)=0$ for $0\leq x\leq 1/2$ and $f(x)=ax^s$ for $x\geq 1$ and some $a\in\mathbb{R}_>$.

Putting $f_r(x):=r^sf(x/r)$ for $r\in\mathbb{R}_>$ and $C_f:=\sup_{x\in\mathbb{R}}|x|^{-s}|f(x)|$, we then easily find that, for each $r\in\mathbb{R}_>$,

(3) $f_r\in\mathfrak{F}_s$.
(4) $f_r(x)=0$ for $0\leq x\leq r/2$ and $f_r(x)=ax^s$ for $x\geq r$.
(5) $f_r(x)\leq C_fx^s$ for all $x\in\mathbb{R}_>$.
(6) $\lim_{r\to\infty}f_r(x)=0$ for all $x\in\mathbb{R}$.

Now fix any $\varepsilon>0$ and pick $n_0\in\mathbb{N}$ such that

$$\sup_{r>0}\big|\mathbb{E}\big(f_r(X_{n_0})-f_r(X_n)\big)\big|\ \leq\ \zeta_s(F_{n_0},F_n)\ <\ \varepsilon$$

for all $n \geq n_0$. Note that, by (5), (6), $X_{n_0} \in L^s$ and the dominated convergence theorem

$$\lim_{r\to\infty} \mathbb{E} f_r(X_{n_0}) \;=\; 0.$$

Then it follows that

$$a\mathbb{E}|X_n|^s \mathbf{1}_{\{|X_n|\geq r\}} \;\leq\; \mathbb{E} f_r(X_n) \;\leq\; \mathbb{E} f_r(X_{n_0}) \;+\; \left|\mathbb{E}\big(f_r(X_{n_0}) - f_r(X_n)\big)\right| \;< 2\varepsilon$$

for all $n \geq n_0$ and sufficiently large r, which clearly proves the uniform integrability of $(|X_n|^s)_{n\geq 1}$. □

We now give a proposition that shows how the property of ζ_s to be an ideal metric helps to provide contraction bounds for the distance between random weighted sums of iid random variables as they appear in the definition of the smoothing transform and in the analysis of distributional recursions like (6.2).

Proposition 6.3. *Let $(C,T) = (C,T_1,T_2,\ldots)$, $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be independent sequences of real-valued random variables in L^s such that*

(C-1) $X_1, X_2, \ldots$ are independent with $\mathscr{L}(X_n) = F_n$ for $n \geq 1$.
(C-2) $Y_1, Y_2, \ldots$ are independent with $\mathscr{L}(Y_n) = G_n$ for $n \geq 1$.
(C-3) For each $n \geq 1$, $F_n, G_n \in \mathscr{P}^s_*(\mathbb{R})$.
(C-4) $\sum_{i\geq 1} T_i X_i + C, \sum_{i\geq 1} T_i Y_i + C \in L^s$.

Then

$$\zeta_s\left(\sum_{i\geq 1} T_i X_i + C, \sum_{i\geq 1} T_i Y_i + C\right) \;\leq\; \sum_{i\geq 1} \mathbb{E}|T_i|^s\, \zeta_s(F_i, G_i). \tag{6.13}$$

If $\mathscr{S}$, the smoothing transform associated with (C,T), exists in L^s-sense, then in particular

$$\zeta_s(\mathscr{S}(F), \mathscr{S}(G)) \;\leq\; \mathbb{E}\left(\sum_{i\geq 1} |T_i|^s\right) \zeta_s(F,G). \tag{6.14}$$

for all $F, G \in \mathscr{P}^s_*(\mathbb{R})$.

Proof. First note that $\zeta_s(\sum_{i\geq 1} T_i X_i + C, \sum_{i\geq 1} T_i Y_i + C) < \infty$ because (C-3) and (C-4) ensure that $\sum_{i\geq 1} T_i X_i + C, \sum_{i\geq 1} T_i Y_i + C \in L^s_*$. Denote by Γ the distribution of (C,T) and let $t = (t_1, t_2, \ldots)$ in the subsequent integration with respect to Γ. Then, by repeated use of properties (6.6) and (6.7) of ζ_s, we infer for each $n \in \mathbb{N}$ that

$$\zeta_s\left(\sum_{i=1}^n T_i X_i + C, \sum_{i=1}^n T_i Y_i + C\right)$$

$$
\begin{aligned}
&= \sup_{f\in\mathfrak{F}_s}\left|\mathbb{E}\left(f\left(\sum_{i=1}^n T_iX_i+C\right)-f\left(\sum_{i=1}^n T_iY_i+C\right)\right)\right| \\
&\le \sup_{f\in\mathfrak{F}_s}\left|\int\mathbb{E}\left(f\left(\sum_{i=1}^n t_iX_i+c\right)-f\left(\sum_{i=1}^n t_iY_i+c\right)\right)\,\Lambda(dc,dt)\right| \\
&\le \int\sup_{f\in\mathfrak{F}_s}\left|\mathbb{E}\left(f\left(\sum_{i=1}^n t_iX_i+c\right)-f\left(\sum_{i=1}^n t_iY_i+c\right)\right)\right|\,\Lambda(dc,dt) \\
&= \int\zeta_s\left(\sum_{i=1}^n t_iX_i+c,\sum_{i=1}^n t_iY_i+c\right)\,\Lambda(dc,dt) \\
&= \int\zeta_s\left(\sum_{i=1}^n t_iX_i,\sum_{i=1}^n t_iY_i\right)\,\Lambda(dc,dt) \\
&\le \int\sum_{k=1}^n\zeta_s\left(\sum_{i=k}^n t_iX_i+\sum_{j=1}^{k-1}t_jY_j,\sum_{i=k+1}^n t_iX_i+\sum_{j=1}^{k}t_jY_j\right)\,\Lambda(dc,dt) \\
&= \int\sum_{k=1}^n\zeta_s\left(t_kX_k+S_k,t_kY_k+S_k\right)\,\Lambda(dc,dt) \\
&\left[\text{where } S_k:=\sum_{i=k+1}^n t_iX_i+\sum_{j=1}^{k-1}t_jY_j \text{ and is independent of } X_k,Y_k\right] \\
&= \int\sum_{k=1}^n\zeta_s\left(t_kX_k,t_kY_k\right)\,\Lambda(dc,dt) \\
&= \int\sum_{k=1}^n|t_k|^s\,\zeta_s\left(X_k,Y_k\right)\,\Lambda(dc,dt) \\
&= \sum_{i=1}^n\mathbb{E}|T_i|^s\,\zeta_s(F_i,G_i)
\end{aligned}
$$

which proves the first asserted inequality by letting n tend to infinity and using

$$
\lim_{n\to\infty}\zeta_s\left(\sum_{i=1}^n T_iX_i+C,\sum_{i=1}^n T_iY_i+C\right) \;=\; \zeta_s\left(\sum_{i\ge1}T_iX_i+C,\sum_{i\ge1}T_iY_i+C\right)
$$

[☞ Problem 6.10]. But the second inequality follows from the first one when choosing $F_i=F$ and $G_i=G$ for all $i\ge1$. □

The final assertion of the lemma has the following direct consequence.

Corollary 6.4. *Let $(C,T)=(C,T_1,T_2,\ldots)$ be a sequence of random variables in L^s such that the associated smoothing transform exists in L^s-sense and*

$$\mathbb{E}\left(\sum_{i\geq 1}|T_i|^s\right) < 1.$$

Then $\mathscr{S}$ is a ζ_s-contraction on $\mathscr{P}^s(\mathbb{R})$ if $0 < s \leq 1$, and on $\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$, if $s > 1$, $s = m+\alpha$ for $(m,\alpha) \in \mathbb{N}\times(0,1]$ and $\mathbf{z}\in\mathbb{R}^m$ is such that $\mathscr{S}$ is a self-map of $\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$.

Problems

In the subsequent problems, it is always assumed that $s = m+\alpha$ for $(m,\alpha) \in \mathbb{N}_0 \times (0,1]$ and that $\mathfrak{F}_s$ is given by (6.3).

Problem 6.5. Prove that $x \mapsto sign(x)|x|^s/c_s$ and $x \mapsto |x|^s/c_s$ with c_s as in Proposition 6.1(a) are both elements of $\mathfrak{F}_s$.

Problem 6.6. Prove for arbitrary $\mathbf{z}\in\mathbb{R}^m$ that $F,G \in \mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$ implies $\zeta_s(F,G) < \infty$.

Problem 6.7. Prove that $\zeta_s(X,Y) = \infty$ if $s > 1$ and X,Y do not have the same integral moments up to order m.

Problem 6.8. Recall that $\mathscr{C}^m_b(\mathbb{R}) = \{f \in \mathscr{C}^m(\mathbb{R}) : f^{(m)} \text{ is bounded}\}$ for $m \in \mathbb{N}_0$ and prove that, for any two random variables X,Y,

$$\mathbb{E}f(X) = \mathbb{E}f(Y) \quad \text{for all } f \in \mathscr{C}^m_b(\mathbb{R})$$

implies $\mathscr{L}(X) = \mathscr{L}(Y)$. Use this fact to verify that ζ_s, as defined in (6.4), is a metric on $\mathscr{P}^s_{\mathbf{z}}(\mathbb{R})$ for any $\mathbf{z}\in\mathbb{R}^m$.

Problem 6.9. Show that ℓ_p is an ideal metric of order $p \wedge 1$ on L^p for any $p > 0$.

Problem 6.10. Prove that $\zeta_s(X_n,Y_n) \to \zeta_s(X,Y)$ whenever $X_n \stackrel{L^s}{\to} X$, $Y_n \stackrel{L^s}{\to} Y$, and $X_n,Y_n \in L^s_*$ for each $n\in\mathbb{N}$.

6.3 Asymptotic normality: Zolotarev versus minimal L^p-metrics

The analysis of `Quicksort` is only one of various similar applications – and probably the most prominent one – where the minimal L^p-metric ℓ_p provides an effective tool to prove contraction of the random quantity of interest after normalization to a limit law which is characterized by a SFPE related to a smoothing transform [☞ also [92, 99] and the references stated therein]. On the other hand, in situations where the limit law is normal, it appears to be inferior to other metrics like the Zolotarev metric ζ_s just introduced. The following discussion, taken from [92, Section 2], is intended to shed some light on this.

Suppose that the limit law is in $\mathscr{P}^2_{0,1}(\mathbb{R})$, the set of distributions with mean zero and variance one, and characterized by the SFPE

$$X \stackrel{d}{=} \frac{X_1+X_2}{2^{1/2}} \tag{6.15}$$

already encountered in Section 1.1 [☞ (1.3)], where X_1, X_2 are independent copies of X. As argued there with the help of the CLT, its unique solution in $\mathscr{P}^2_{0,1}(\mathbb{R})$ is the standard normal law. Moreover, if $\mathscr{S}$ denotes the related smoothing transform, then the CLT further shows that

$$\mathscr{S}^n(F) \stackrel{w}{\to} Normal(0,1)$$

for any $F \in \mathscr{P}^2_{0,1}(\mathbb{R})$. On the other hand, Lemma 5.50(a) provides us with

$$\ell_2^2(\mathscr{S}(F),\mathscr{S}(G)) \le \left(\left(\frac{1}{2^{1/2}}\right)^2 + \left(\frac{1}{2^{1/2}}\right)^2\right) \ell_2^2(F,G) = \ell_2^2(F,G)$$

for all $F, G \in \mathscr{P}^2_0(\mathbb{R})$ which suggests that $\mathscr{S}$ is not an ℓ_2-contraction on $\mathscr{P}^2_0(\mathbb{R})$, nor on its restriction $\mathscr{P}^2_{0,1}(\mathbb{R})$. In fact, since every normal law with mean 0 and finite variance solves (6.15), $\mathscr{S}$ cannot be a contraction (or quasi-contraction) with respect to ℓ_2 on $\mathscr{P}^2_0(\mathbb{R})$, and the same holds true with respect to any ℓ_p for $p > 2$.

The problem is that, even when restricting to the class $\mathscr{P}^2_{0,1}(\mathbb{R})$, where the fixed point is unique, ℓ_2 does not reflect this restriction in a way to yield contraction, and here is exactly the point where the Zolotarev metric turns out to be more sensitive. To see this, fix any $\alpha \in (0,1]$, put $s = 2+\alpha$ and choose arbitrary $F, G \in \mathscr{P}^s_{0,1}(\mathbb{R})$. Then, with independent X_1, X_2, Y_1, Y_2 such that $(X_i, Y_i) \sim (F, G)$ for $i = 1, 2$, we obtain by invoking Proposition 6.3 that

$$\begin{aligned}\zeta_s(\mathscr{S}(F),\mathscr{S}(G)) &= \zeta_s\left(\frac{X_1+X_2}{2^{1/2}}, \frac{Y_1+Y_2}{2^{1/2}}\right) \\ &\le \left(\left(\frac{1}{2^{1/2}}\right)^s + \left(\frac{1}{2^{1/2}}\right)^s\right) \zeta_s(F,G)\end{aligned}$$

and thus contraction, for $2(1/2^{1/2})^s < 1$. Consequently, by Banach's fixed point theorem, the standard normal law is the unique fixed point of $\mathscr{S}$ in $\bigcap_{s>2} \mathscr{P}^s_{0,1}(\mathbb{R})$.

Of course, the preceding discussion has only exemplified the advantage of the Zolotarev metric for one special case. Nevertheless, it should have provided some evidence that this advantage prevails in other applications as well, most notably those where a normal distribution appears as the limit law.

6.4 Back to recurrence equation (6.2): a general convergence theorem

We return to the general recurrence equation

$$X_n \stackrel{d}{=} \sum_{i=1}^{N_n} A_{n,i} X_{i,\tau_{n,i}} + B_n, \quad n \geq n_0 \tag{6.1}$$

and its equivalent form

$$\widehat{X}_n \stackrel{d}{=} \sum_{i=1}^{N_n} T_{n,i} \widehat{X}_{i,\tau_{n,i}} + C_n, \quad n \geq n_0 \tag{6.2}$$

for the normalized sequence

$$\widehat{X}_n := \frac{X_n - \mu_n}{\sigma_n},$$

where $(\mu_n, \sigma_n^2) \in \mathbb{R} \times \mathbb{R}_>$ and

$$C_n := \frac{1}{\sigma_n}\left(B_n - \mu_n + \sum_{i=1}^{N_n} T_{n,i}\, \mu_{\tau_{n,i}}\right).$$

The general standing assumptions on $B_n, A_{n,i}, \tau_{n,i}, X_{n,i}$ are listed under (A1)-(A4). Under additional conditions on the input parameters $B_n, T_{n,i}$ and $\tau_{n,i}$, our goal is to show convergence of $\widehat{X}_n$ to a limit in L^s with respect to the Zolotarev metric ζ_s. Since normalization can only scale first and second moments, it is natural to use ζ_s for $0 < s \leq 3$.

Plainly, the X_n must be elements of L^s when using ζ_s, and in the case $2 < s \leq 3$ we further assume that $\mathbb{V}\mathrm{ar}\, X_n > 0$ for all $n \geq n_1 \geq n_0$. Our settings for the normalizing constants μ_n, σ_n is displayed in the following table, where an asterisk indicates an arbitrary positive number:

	μ_n	σ_n^2	
$0 < s \leq 1$	0	$*$	$n \geq 0$
$1 < s \leq 2$	$\mathbb{E}X_n$	$*$	$n \geq 0$
$2 < s \leq 3$	$\mathbb{E}X_n$	1	$0 \leq n < n_1$
	$\mathbb{E}X_n$	$\mathbb{V}\mathrm{ar}\, X_n$	$n \geq n_1$

This guarantees that all X_n have expectation zero if $1 < s \leq 3$ and further unit variance for $n \geq n_1$ if $2 < s \leq 3$. We are now in the position to state the main convergence theorem which slightly generalizes the corresponding results by NEININGER & RÜSCHENDORF [92, Thms. 4.1 and 4.6]

Theorem 6.11. *Let* $0 < s = m + \alpha \le 3$ *for* $(m, \alpha) \in \mathbb{N}_0 \times (0, 1]$ *and* $(X_n)_{n\ge 0}$ *be a sequence of random variables in* L^s *satisfying* (6.1) *under (A1)-(A4). Let* F_n *denote the law of the normalization* $\widehat{X}_n$. *Suppose further there exist random variables* $C, T_1, T_2, \ldots \in L^s$ *with* $\sum_{i\ge 1} T_i^2 \in L^{s/2}$ *and associated smoothing transform* $\mathscr{S}$, *such that*

$$\sum_{i\ge 1} \mathbb{E}|T_i|^s < 1, \tag{6.16}$$

$$\begin{aligned} \textstyle\sum_{i\ge 1} \mathbb{E}|T_{n,i} - T_i|^s \to 0 \quad &\textit{if} \quad 0 < s \le 2, \\ \mathbb{E}\left(\textstyle\sum_{i\ge 1} (T_{n,i} - T_i)^2\right)^{s/2} \to 0 \quad &\textit{if} \quad 2 < s \le 3, \end{aligned} \tag{6.17}$$

$$\mathbb{E}|C_n - C|^s \to 0. \tag{6.18}$$

Finally, suppose that, for some $\mathbb{N} \ni m_n \to \infty$, $m_n \le n$,

$$\mathbb{E}\left(\sum_{i\ge 1} \mathbf{1}_{\{\tau_{n,i} < m_n \ or\ = n\}} |T_{n,i}|^s\right) \to 0. \tag{6.19}$$

Then $\zeta_s(F_n, F) \to 0$, *where* F *denotes the unique fixed point of* $\mathscr{S}$ *in* $\mathscr{P}^s(\mathbb{R})$ *if* $0 < s \le 1$, *in* $\mathscr{P}_0^s(\mathbb{R})$ *if* $1 < s \le 2$, *and in* $\mathscr{P}_{0,1}^s(\mathbb{R})$ *if* $2 < s \le 3$ *[☞ Cor. 6.4].*

Remark 6.12. We note in passing that, under the conditions of the theorem,

$$\sum_{i\ge 1} T_{n,i} \in L^s \quad \text{and} \quad C_n \in L^s$$

for all $n \in \mathbb{N}_0$, and that $\mathscr{S}$ exists in L^s-sense if $0 < p \le 1$, and as a self-map of $\mathscr{P}_0^s(\mathbb{R})$ if $1 < s \le 3$ [☞ Corollary 5.35]. In the last case, one should also observe that $C, C_1, C_2, \ldots$ are centered. The condition $\sum_{i\ge 1} T_i^2 \in L^{s/2}$ is a real proviso only when $2 < s \le 3$, but follows from (6.16) otherwise, for then $(\sum_{i\ge 1} T_i^2)^{s/2} \le \sum_{i\ge 1} |T_i|^s$ by subadditivity. Moreover, (6.19) is easily seen to be equivalent to

$$\mathbb{E}\left(\sum_{i\ge 1} \mathbf{1}_{\{\tau_{n,i} < m \text{ or } = n\}} |T_{n,i}|^s\right) \to 0 \quad \text{for each } m \in \mathbb{N},$$

which is the corresponding condition stated in [92]. Also, the condition

$$\left\| \sum_{i\ge 1} |T_{n,i} - T_i| \right\|_s \to 0$$

is used there instead of (6.17) which appears to be stronger in the case when the N_n are unbounded [☞ [92, Thm. 4.6]].

Proof. The following argument shows that it is no loss of generality to assume $n_0 = n_1 = 0$, thus guaranteeing that *all* F_n are in $\mathscr{P}_0^s(\mathbb{R})$ if $1 < p \le 2$, and in $\mathscr{P}_{0,1}^s(\mathbb{R})$ if

$2 < s \le 3$: Defining

$$\begin{aligned}
X_n^* &:= \widehat{X}_{n_1+n}, \quad X_{i,n}^* := \widehat{X}_{i,n_1+n}, \\
(T_{n,i}^*, \tau_{n,i}^*) &:= \left(T_{n_1+n,i}\,\mathbf{1}_{\{\tau_{n_1+n,i}\ge n_1\}}, \tau_{n_1+n,i}\vee n_1\right), \\
N_n^* &:= (N_n - n_1)^+, \\
\text{and} \quad C_n^* &:= \sum_{i=1}^{N_{n_1+n}} T_{n_1+n,i}\widehat{X}_{i,\tau_{n_1+n,i}}\mathbf{1}_{\{\tau_{n_1+n,i}<n_1\}} + C_{n_1+n},
\end{aligned}$$

one can easily see that

$$X_n^* \overset{d}{=} \sum_{i=1}^{N_n^*} T_{n,i}^* X_{i,\tau_{n,i}^*}^* + C_n^* \quad (n \ge 0)$$

and that the $T_{n,i}^*, \tau_{n,i}^*, C_n^*, X_{i,n}^*$ satisfy the same model assumptions as the original variables $T_{n,i}, \tau_{n,i}, C_n, \widehat{X}_{i,n}$, including the conditions of the theorem: just note for (6.17) that

$$\begin{aligned}
&\textstyle\sum_{i\ge1} \mathbb{E}|T_{n,i}^* - T_{n_1+n,i}|^s \to 0 \quad \text{if} \quad 0 < s \le 2, \\
\text{and} \quad &\mathbb{E}\left(\textstyle\sum_{i\ge1}(T_{n,i}^* - T_{n_1+n,i})^2\right)^{s/2} \to 0 \quad \text{if} \quad 2 < s \le 3,
\end{aligned}$$

and for (6.18) that

$$\|C_n^* - C\|_s \;\le\; \|C_{n_1+n} - C\|_s + \left\|\max_{0\le k<n_1} \widehat{X}_k\right\|_s \left\|\sum_{i=1}^{N_n} T_{n,i}\,\mathbf{1}_{\{\tau_{n_1+n,i}<n_1\}}\right\|_s \;\to\; 0.$$

Hence, we will from now on assume that $n_0 = n_1 = 0$.

Let $(\widehat{X}_{i,\infty})_{i\ge1}$ be a family of independent random variables with common distribution F, generic copy $\widehat{X}_\infty$, and independent of the $T_{n,i}, T_i$ and $\tau_{n,i}$. Setting

$$\begin{aligned}
Y_n &:= \sum_{i\ge1} T_{n,i}\widehat{X}_{i,\tau_{n,i}} + C_n \quad [\text{with law } F_n], \\
Y_n^{(1)} &:= \sum_{i\ge1} T_{n,i}\widehat{X}_{i,\tau_{n,i}} + C \quad [\text{with law } F_n^{(1)},\ \text{say}], \\
Y_n^{(2)} &:= \sum_{i\ge1} T_{n,i}\widehat{X}_{i,\infty} + C \quad [\text{with law } F_n^{(2)},\ \text{say}], \\
Y_n^* &:= \sum_{i\ge1} T_i\widehat{X}_{i,\infty} + C \quad [\text{with law } F]
\end{aligned}$$

for $n \ge n_1$, we must show $\zeta_s(Y_n, Y_n^*) \to 0$ and will do so by “inserting” the auxiliary random variables $Y_n^{(i)}$ for $i = 1,2$, viz.

$$a_n \;:=\; \zeta_s(Y_n,Y_n^*) \;\le\; \zeta_s(Y_n,Y_n^{(1)})+\zeta_s(Y_n^{(1)},Y_n^{(2)})+\zeta_s(Y_n^{(2)},Y_n^*), \tag{6.20}$$

and proving convergence to zero of each term on the right-hand side individually.

It is vital to observe here that F and all $F_n,F_n^{(1)},F_n^{(2)}$ for $n\ge 0$ are in $\mathscr{P}_0^s(\mathbb{R})$ if $1<s\le 2$, and in $\mathscr{P}_{0,1}^s(\mathbb{R})$ if $2<s\le 3$, for this ensures that all terms on the right-hand side of (6.20) are finite. Moreover, by Proposition 6.2(b), we may (and will) replace ζ_s with ℓ_s where appropriate, in fact for the first and the last of the three terms. Using assumption (6.18), we obtain

$$\ell_s^{s\vee 1}(Y_n,Y_n^{(1)}) \;=\; \ell_s^{s\vee 1}\left(\sum_{i\ge 1}T_{n,i}\widehat{X}_{i,\tau_{n,i}}+C_n,\sum_{i\ge 1}T_{n,i}\widehat{X}_{i,\tau_{n,i}}+C\right) \;=\; \mathbb{E}|C_n-C|^s \;\to\; 0,$$

while (6.17) in connection with subadditivity $(0<s\le 1)$ or a suitable martingale inequality $(1<s\le 3)$ provides us with

$$\begin{aligned}\ell_s^s(Y_n^{(2)},Y_n^*) \;&\le\; \begin{cases}\sum_{i\ge 1}\mathbb{E}|T_{n,i}-T_i|^s\,\mathbb{E}|\widehat{X}_\infty|^s, & \text{if } 0<s\le 1\\ 2\sum_{i\ge 1}\mathbb{E}|T_{n,i}-T_i|^s\,\mathbb{E}|\widehat{X}_\infty|^s, & \text{if } 1<s\le 2\\ c_s\,\mathbb{E}(\sum_{i\ge 1}(T_{n,i}-T_i)^2)^{s/2}\left[\left(\mathbb{E}\widehat{X}_\infty^2\right)^{s/2}+\mathbb{E}|\widehat{X}_\infty|^s\right], & \text{if } 2<s\le 3\end{cases}\\ &\to\; 0\end{aligned} \tag{6.21}$$

where the constant c_s depends only on s. The details are very similar to those given at various places in Section 5.5 and therefore left to the reader [☞ Problem 6.15]. We just note that inequality (B.4) in the Appendix together with $\sum_{i\ge 1}\mathbb{E}|T_{n,i}-T_i|^s\le \mathbb{E}(\sum_{i\ge 1}(T_{n,i}-T_i)^2)^{s/2}$ should be utilized in the case $2<s\le 3$.

Turning to $\zeta_s(Y_n^{(1)},Y_n^{(2)})$, Proposition 6.3 after conditioning with respect to $(T_{n,i},\tau_{n,i})_{i\ge 1}$ and C provides us with the estimate

$$\zeta_s(Y_n^{(1)},Y_n^{(2)}) \;\le\; \sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s a_{\tau_{n,i}}$$

when keeping in mind that $\zeta_s(\widehat{X}_{i,k},\widehat{X}_{i,\infty})=\zeta_s(F_k,F)=a_k$ for any $i,k\in\mathbb{N}$. Then, by combining the previous estimates in (6.20), we arrive at

$$a_n \;\le\; \sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s a_{\tau_{n,i}}+b_n$$

for some nonnegative b_n converging to zero. Next, use (6.17), (6.16) and (6.19) to infer first that

$$\lim_{n\to\infty}\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s \;=\; \sum_{i\ge 1}\mathbb{E}|T_i|^s$$

[☞ Problem 6.16] and then that, for some $\varepsilon,\gamma\in(0,1)$ with $\gamma/(1-\varepsilon)<1$, $n_2\in\mathbb{N}$ and all $n\ge n_2$,

$$\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s\mathbf{1}_{\{\tau_{n,i}=n\}}<\varepsilon\quad\text{and}\quad\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s<\gamma(1-\varepsilon).$$

As a consequence,

$$(1-\varepsilon)a_n\ \le\ \sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s a_{\tau_{n,i}}\mathbf{1}_{\{\tau_{n,i}<n\}}+b_n\ <\ \gamma(1-\varepsilon)a^*_{n-1}+b_n$$

or, equivalently,

$$a_n\ \le\ \frac{1}{1-\varepsilon}\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s a_{\tau_{n,i}}\mathbf{1}_{\{\tau_{n,i}<n\}}+\frac{b_n}{1-\varepsilon}\ <\ \gamma a^*_{n-1}+\frac{b_n}{1-\varepsilon}$$

for all $n\ge n_2$, where $a^*_n=\max_{0\le k\le n}a_k$. The second inequality immediately implies the boundedness of the a_n. Then, by another use of (6.19),

$$\varepsilon_n\ :=\ \sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s a_{\tau_{n,i}}\mathbf{1}_{\{\tau_{n,i}<m_n\}}\ \le\ \sup_{k\ge 0}a_k\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s\mathbf{1}_{\{\tau_{n,i}<m_n\}}\ \to\ 0,$$

so that

$$a_n\ \le\ \frac{1}{1-\varepsilon}\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s a_{\tau_{n,i}}\mathbf{1}_{\{m_n\le\tau_{n,i}<n\}}+\frac{\varepsilon_n+b_n}{1-\varepsilon}\tag{6.22}$$

for all $n\ge n_2$. Now $a_n\to 0$ is easily deduced and details are thus left to the reader [☞ Problem 6.17]. □

We close this section with two corollaries [☞ [92, Thm. 5.1 and Cor. 5.2]] which are easily inferred from the previous theorem by checking its assumptions.

Corollary 6.13. *Let $(X_n)_{n\ge 0}$ be a sequence in L^s, $0<s\le 3$, satisfying the distributional recursion* (6.1) *with $A_{n,i}=1$ for $1\le i\le N_n$. Suppose that, for suitable $f(n)\in\mathbb{R}$ and $g(n)\in\mathbb{R}_\ge$ (positive for sufficiently large n),*

$$\mathbb{E}X_n=f(n)+o(g(n)),\quad \textit{if } 1<s\le 3$$

and, furthermore,

$$\mathbb{V}\text{ar}\,X_n=g(n)^2+o(g(n)^2),\quad \textit{if } 2<s\le 3.$$

Finally suppose that, for suitable $C,T_1,T_2,\ldots\in L^s$ with $\sum_{i\ge 1}T_i^2\in L^{s/2}$,

$$\begin{aligned}\sum_{i\ge 1}\mathbb{E}\left|\frac{g(\tau_{n,i})}{g(n)}-T_i\right|^s\ \to\ 0\quad \textit{if}\quad 0<s\le 2,\\ \mathbb{E}\left(\sum_{i\ge 1}\left(\frac{g(\tau_{n,i})}{g(n)}-T_i\right)^2\right)^{s/2}\ \to\ 0\quad \textit{if}\quad 2<s\le 3,\end{aligned}\tag{6.23}$$

$$\mathbb{E}\left|\frac{1}{g(n)}\left(B_n - f(n) + \sum_{i=1}^{N_n} f(\tau_{n,i})\right) - C\right|^s \to 0, \tag{6.24}$$

$$\mathbb{E}\left(\sum_{i=1}^{N_n} \mathbf{1}_{\{\tau_{n,i}=n\}}\right) \to 0 \tag{6.25}$$

and the contraction condition (6.16)*, i.e.* $\sum_{i\geq 1}\mathbb{E}|T_i|^s < 1$*, hold true. Then*

$$\frac{X_n - f(n)}{g(n)} \xrightarrow{d} F,$$

where F *is the unique solution of the smoothing transform associated with* $(C, T_1, T_2, \ldots)$ *in* $\mathscr{P}^s(\mathbb{R})$ *if* $0 < s \leq 1$*, in* $\mathscr{P}^s_0(\mathbb{R})$ *if* $1 < s \leq 2$*, and in* $\mathscr{P}^s_{0,1}(\mathbb{R})$ *if* $2 < s \leq 3$*.*

Proof. First check that $(X_n - f(n))/g(n) \xrightarrow{d} F$ iff $\widehat{X}_n = (X_n - \mu_n)/\sigma_n \xrightarrow{d} F$ with μ_n, σ_n^2 chosen as in Theorem 6.11. Then show that all assumptions of the last result are valid. Details are left to the reader [☞ Problem 6.18]. □

Specializing to the case when $s > 2$, the following central limit theorem can be stated.

Corollary 6.14. *Let* $s > 2$ *and* $C = 0$ *in the previous corollary and suppose additionally that*

$$\sum_{i\geq 1} T_i^2 = 1 \quad \textit{and} \quad \mathbb{P}(T_i = 1 \textit{ for some } i \geq 1) < 1. \tag{6.26}$$

Then

$$\frac{X_n - f(n)}{g(n)} \xrightarrow{d} \textit{Normal}\,(0,1).$$

Proof. ☞ Problem 6.19 □

Problems

Problem 6.15. Verify (6.21) by proving the inequalities stated there.

Problem 6.16. Prove that (6.17) and (6.16) imply

$$\lim_{n\to\infty} \sum_{i\geq 1} \mathbb{E}\big||T_{n,i}|^s - |T_i|^s\big| \;=\; 0$$

and thus in particular

$$\lim_{n\to\infty}\sum_{i\ge 1}\mathbb{E}|T_{n,i}|^s \;=\; \sum_{i\ge 1}\mathbb{E}|T_i|^s$$

Problem 6.17. Use (6.20) and the boundedness of the a_n to show that $a_n \to 0$ [☞ also Problem 5.65].

Problem 6.18. Prove Corollary 6.13.

Problem 6.19. Prove Corollary 6.14.

Problem 6.20. Use Corollary 6.13 to reprove Theorem 5.61, the main limit result for `Quicksort`, for any $0 < p \le 3$.

6.5 The degenerate case: recursion (6.2) with tautological limit equation

There is an important class of recursions of the form (6.1) which, albeit contracting to a limit distribution after normalization, are not covered by the results of the previous section because the distributional limit equation of the normalized recursion (6.2) is tautological and thus not providing any information about that limit. This situation, also studied and referred to as the *degenerate case* by NEININGER & RÜSCHENDORF [93], occurs when

$$\lim_{n\to\infty}(T_{n,1},T_{n,2},\ldots) \;=\; (1,0,0,\ldots) \quad\text{and}\quad \lim_{n\to\infty} C_n \;=\; 0$$

(assuming w.l.o.g. that $T_{n,1} \ge T_{n,2} \ge \ldots$ a.s.) and, consequently, the limit SFPE resulting from (6.2) becomes

$$\widehat{X} \;\stackrel{d}{=}\; \widehat{X}.$$

Under the basic assumption that all X_n are in L^s for some $s > 2$, the results proved in this section will provide conditions that ensure asymptotic normality of $\widehat{X}_n$. The normalizing constants μ_n, σ_n^2 are chosen as before, thus $\mu_n = \mathbb{E}X_n$ for $n \ge 0$, $\sigma_n^2 = \mathbb{V}\mathrm{ar}X_n$ if the variance is positive, and $\sigma_n^2 = 1$ otherwise.

As in [93], we will first focus on the case $N_n = 1$ hereafter, that is, on recursions of the form

$$X_n \;\stackrel{d}{=}\; X_{\tau_n} + B_n, \quad n \ge n_0, \tag{6.27}$$

which after normalization becomes

$$\widehat{X}_n \;\stackrel{d}{=}\; T_n\widehat{X}_{\tau_n} + C_n, \quad n \ge n_0, \tag{6.28}$$

where

$$T_n \;:=\; \frac{\sigma_{\tau_n}}{\sigma_n} \quad\text{and}\quad C_n \;:=\; \frac{B_n - \mu_n + \mu_{\tau_n}}{\sigma_n}.$$

Since there is no branching ($N_n = 1$), we have dropped the double index on the right-hand sides of these equations (writing X_n, T_n, τ_n for $X_{1,n}, T_{n,1}, \tau_{n,1}$). We make the usual assumptions that $X_0, X_1, ...$ are mutually independent and also independent of $(B_n, \tau_n)_{n\ge n_0}$ (and thus of $(T_n)_{n\ge n_0}$ as well), that each τ_n takes values in $\{0,...,n\}$, and that $\mathbb{P}(\tau_n = n) < 1$ for $n \ge n_0$. The recursions (6.27) and (6.28) can be extended to all $n \ge 0$ by setting

$$\tau_n := n, \quad T_n := 1 \quad \text{and} \quad B_n = C_n := 0$$

for $n = 0,...,n_0 - 1$. This is tacitly assumed hereafter.

The following CLT provides an extension of the corresponding result by NEININGER & RÜSCHENDORF [93, Thm. 2.1] in that it holds for any $s \in (2,3]$ instead of $s = 3$ only. Moreover, (6.33) in the latter case appears to be slightly stronger due to a somewhat different approach chosen here for the proof [☞ especially Lemma 6.23 below].

Theorem 6.21. *Let $s = 2+\alpha$ for $\alpha \in (0,1]$ and $(X_n)_{n\ge 0}$ be a sequence of random variables in L^s satisfying (6.27) under the usual conditions. Suppose further that, for some $\beta, \theta \in \mathbb{R}_>$ and $0 \le \gamma < \beta$*

$$\sigma_n^2 \;=\; \theta \log^{2\beta} n + O(\log^{2\gamma} n), \tag{6.29}$$

$$\|B_n - \mu_n + \mu_{\tau_n}\|_s \;=\; O(\log^{\gamma} n), \tag{6.30}$$

as $n \to \infty$, and

$$\limsup_{n\to\infty} \mathbb{E}\log\left(\frac{\tau_n \vee 1}{n}\right) < 0, \tag{6.31}$$

$$\sup_{n\ge 1}\left\|\log\left(\frac{\tau_n \vee 1}{n}\right)\right\|_s < \infty. \tag{6.32}$$

Then $\widehat{X}_n \xrightarrow{d} Normal(0,1)$, in fact

$$\zeta_s(\mathscr{L}(\widehat{X}_n), Normal(0,1)) \;=\; O(\log^{-\alpha\delta} n) \tag{6.33}$$

as $n \to \infty$, where $\delta := (\beta - \gamma) \wedge 1$.

It is not difficult to show that (6.29) and (6.32) imply the degeneracy conditions

$$\frac{\tau_n}{n} \to 1 \quad \text{and} \quad T_n = \frac{\sigma_{\tau_n}}{\sigma_n} \to 1 \quad \text{a.s.}$$

[☞ Problem 6.25].

Proof. W.l.o.g. [☞ Problem 6.26] let $\mathbb{V}\mathrm{ar}X_n > 0$ for all $n \ge 0$ which ensures that all $F_n := \mathscr{L}(\widehat{X}_n)$ are elements of $\mathscr{P}^s_{0,1}(\mathbb{R})$. Based on the sequence $(\tau_n)_{n\ge 0}$, define

$$\varphi : \Omega \times \mathbb{N}_0 \to \mathbb{N}_0, \quad (\omega, n) \mapsto \tau_n(\omega),$$

and let $(\varphi_n, B_{n,1}, B_{n,2}, ...)$ for $n \geq 1$ be iid copies of $(\varphi, B_1, B_2, ...)$. Notice that $(\varphi_n)_{n\geq 1}$ forms an IFS of iid Lipschitz maps on $\mathbb{N}_0$ and put $\varphi_{n:1} := \varphi_n \circ ... \circ \varphi_1$ for $n \geq 1$ [as in Chapter 3]. Then it is easily verified via iteration that

$$\widehat{X}_n \stackrel{d}{=} \frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n} \widehat{X}_{\varphi_{k:1}(n)} + \sum_{j=1}^{k} \frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n} C_{j,\varphi_{j-1:1}(n)} =: \widehat{X}_{n,k} \tag{6.34}$$

holds true for each $n \geq 0$ and $k \geq 1$, where $\varphi_{0:1}(n) := n$ and

$$C_{j,k} := \frac{B_{j,k} - \mu_k + \mu_{\varphi_j(k)}}{\sigma_k}$$

for $j,k \geq 1$. Note that $(\varphi_j, C_{j,1}, C_{j,2}, ...)$ for $j \geq 1$ are iid copies of $(\varphi, C_1, C_2, ...)$.

Pick $m \in \mathbb{N}$ so large that

$$\nu_m := \sup_{n\geq m} \mathbb{P}(\varphi(n) = n) < 1 \tag{6.35}$$

$$\text{and} \quad \xi := 1 + \frac{1}{\log m} \sup_{n\geq m} \mathbb{E} \log\left(\frac{\tau_n \vee 1}{n}\right) < 1, \tag{6.36}$$

which is possible by (6.31). Observe also that, by (6.29),

$$\sup_{n\geq 1} \frac{\sigma_n^*}{\sigma_n} < \infty,$$

where $\sigma_n^* := \max_{0\leq k\leq n} \sigma_k$.

Let $Z_1, Z_2, ...$ be iid standard normal random variables independent of the "rest of the world" and define

$$Y_{n,k} := \frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n} Z_{k+1} + \sum_{j=1}^{k} \frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n} C_{j,\varphi_{j-1:1}(n)}, \tag{6.37}$$

$$Y_{n,k}^* := \frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n} Z_{k+1} + \sum_{j=1}^{k} \frac{\left(\sigma^2_{\varphi_{j-1:1}(n)} - \sigma^2_{\varphi_{j:1}(n)}\right)^{1/2}}{\sigma_n} Z_j \tag{6.38}$$

for $n \geq 0$ and $1 \leq k \leq n$. Put further $a_n := \zeta_s(F_n, Normal(0,1))$. It is easily verified and stated as part of Lemma 6.23 below that each $Y_{n,k}^*$ has a standard normal law. Therefore,

$$a_n = \zeta_s(\widehat{X}_{n,k}, Y_{n,k}^*) \leq \zeta_s(\widehat{X}_{n,k}, Y_{n,k}) + \zeta_s(Y_{n,k}, Y_{n,k}^*) \tag{6.39}$$

for all $n \geq 0$ and $1 \leq k \leq n$.

Since $\widehat{X}_{n,k}, Z_k$ have equal first and second moments, we infer by an appeal to Proposition 6.3 that

$$\begin{aligned}\zeta_s(\widehat{X}_{n,k},Y_{n,k}) &\le \nu_m^k a_n + \mathbb{E}\left(\frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n}\right)^s \mathbf{1}_{\{\varphi_{k:1}(n)<n\}}\, a_{\varphi_{k:1}(n)} \\ &\le \nu_m^k a_n + \mathbb{E}\left(\frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n}\right)^s \mathbf{1}_{\{\varphi_{k:1}(n)<n\}}\, a^*_{n-1}\end{aligned} \tag{6.40}$$

for any $n \ge m$, where $a_n^* := \max_{0\le j\le n} a_j$. Put $p := s\beta \wedge 1$ and estimate further (with suitable constants $\theta_1, \theta_2, \theta_3 \in \mathbb{R}_>$)

$$\begin{aligned}\mathbb{E}\left(\frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n}\right)^s \mathbf{1}_{\{\varphi_{k:1}(n)<n\}} &\le \left(\frac{\sigma_m^*}{\sigma_n}\right)^s + \mathbb{E}\left(\frac{\theta_1 \log^\beta \varphi_{k:1}(n)}{\theta_2 \log^\beta n}\right)^s \mathbf{1}_{\{m\le\varphi_{k:1}(n)<n\}} \\ &\le \left(\frac{\sigma_m^*}{\sigma_n}\right)^s + \left(\frac{\theta_1}{\theta_2}\right)^s \xi^{kp} \\ &\le \theta_3\left(\frac{1}{\log^{s\beta} n} + \xi^{kp}\right)\end{aligned} \tag{6.41}$$

for all $k,n \in \mathbb{N}$ satisfying $n \ge k \ge m$, where Lemma 6.22 further below has been utilized for the last line.

As for the last term in (6.39), a combination of Proposition 6.1(d), viz. (6.11), and Lemma 6.23(d) (with $Z_{j,k}$ as defined there) provides us with

$$\begin{aligned}\zeta_s(Y_{n,k},Y^*_{n,k}) &\le K\,\ell_s(Y_{n,k},Y^*_{n,k})^\alpha \\ &\le K\left(\sum_{j=1}^k \left\| \frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n}\big(C_{j,\varphi_{j-1:1}(n)} - Z_{j,\varphi_{j-1:1}(n)}\big)\right\|_s\right)^\alpha \\ &\le \frac{M\big(1+k\log^{-(\beta-\delta)} n\big)^\alpha}{\log^{\alpha\delta} n}\end{aligned}$$

for all $k,n \ge 1$ and suitable constants $K,M \in \mathbb{R}_>$.

By combining the last result with (6.39)-(6.41), we arrive at

$$a_n \le \frac{\theta_3}{1-\nu_m^k}\left(\frac{1}{\log^{s\beta} n} + \xi^{kp}\right) a^*_{n-1} + \frac{M\big(1+k\log^{-(\beta-\delta)} n\big)^\alpha}{\log^{\alpha\delta} n}$$

for all $k,n \in \mathbb{N}$ with $n \ge k \ge m$ which, by fixing k large enough, particularly implies

$$a_n \le \zeta\, a^*_{n-1} + O(\log^{\alpha\delta} n)$$

for some $\zeta \in (0,1)$ and all $n \ge n_1$, n_1 sufficiently large. As an immediate consequence, we find that $a^* := \sup_{n\ge 0} a_n < \infty$ and thus

$$a_n \le \frac{\theta_3}{1-\nu_m^k}\left(\frac{1}{\log^{s\beta} n} + \xi^{kp}\right) a^* + \frac{M\big(1+k\log^{-(\beta-\delta)} n\big)^\alpha}{\log^{\alpha\delta} n}$$

for all $k,n \in \mathbb{N}$ with $n \geq k \geq m$. Finally, choose $k = k_n \simeq \log^{\varepsilon} n$ for some arbitrary $\varepsilon \in (0, \beta - \delta)$ to infer that

$$a_n \;\leq\; \frac{\theta_3}{1 - \nu_m^{k_n}} \left(\frac{1}{\log^{s\beta} n} + \xi^{k_n p} \right) a^* + \frac{M\big(1 + k_n \log^{-(\beta-\delta)} n\big)^{\alpha}}{\log^{\alpha\delta} n} \;=\; O(\log^{\alpha\delta} n)$$

as claimed. □

The two lemmata that have been used in the previous proof are next.

Lemma 6.22. *Given the assumptions of Theorem 6.21, let $m \in \mathbb{N}$ be such that* (6.36) *holds and $r > 0$. Then*

$$\mathbb{E}\left(\frac{\log \varphi_{k:1}(n)}{\log n}\right)^{r} \mathbf{1}_{\{\varphi_{k:1}(n) \geq m\}} \;\leq\; \xi^{kp}$$

for all $k \geq 1$, where $p = r \wedge 1$.

Proof. We start by pointing out that

$$\begin{aligned}
&\mathbb{E}\left(\frac{\log \varphi_{k:1}(n)}{\log n}\right)^{r} \mathbf{1}_{\{\varphi_{k:1}(n) \geq m\}} \\
&\quad \leq\; \mathbb{E}\left(\prod_{j=1}^{k} \frac{\log \varphi_j(\varphi_{j-1:1}(n))}{\log \varphi_{j-1:1}(n)}\right)^{p} \mathbf{1}_{\{\varphi_{i:1}(n) \geq m,\, i=1\ldots,k\}} \\
&\quad \leq\; \left[\mathbb{E}\left(\prod_{j=1}^{k} \frac{\log \varphi_j(\varphi_{j-1:1}(n))}{\log \varphi_{j-1:1}(n)} \mathbf{1}_{\{\varphi_{i:1}(n) \geq m,\, i=1\ldots,k\}}\right)\right]^{p},
\end{aligned}$$

where we have used Jensen's inequality for the last line and the fact that all factors under the expectation are ≤ 1 for the penultimate one. Putting

$$\mathfrak{A}_0 := \{\Omega, \emptyset\} \quad \text{and} \quad \mathfrak{A}_k := \sigma(\varphi_1, \ldots, \varphi_k) \quad \text{for } k \geq 1, \tag{6.42}$$

we have on the event $\{\varphi_{j-1:1}(n) \geq m\}$ that

$$\begin{aligned}
&\mathbb{E}\left(\frac{\log \varphi_j(\varphi_{j-1:1}(n))}{\log \varphi_{j-1:1}(n)} \mathbf{1}_{\{\varphi_j(\varphi_{j-1:1}(n)) \geq m\}} \,\middle|\, \mathfrak{A}_{j-1}\right) \\
&\quad \leq\; 1 + \mathbb{E}\left(\frac{\log\left(\frac{\varphi_j(\varphi_{j-1:1}(n))}{\varphi_{j-1:1}(n)}\right)}{\log \varphi_{j-1:1}(n)} \mathbf{1}_{\{\varphi_j(\varphi_{j-1:1}(n)) \geq m\}} \,\middle|\, \mathfrak{A}_{j-1}\right) \\
&\quad \leq\; 1 + \frac{1}{\log m} \sup_{k \geq m} \mathbb{E} \log\left(\frac{\varphi(k) \vee 1}{k}\right) \\
&\quad =\; \xi \;<\; 1 \quad \text{a.s.}
\end{aligned}$$

Hence, by successive conditioning with respect to $\mathfrak{A}_{k_n-1},...,\mathfrak{A}_0$, we find

$$\mathbb{E}\left(\prod_{j=1}^{k}\frac{\log\varphi_j(\varphi_{j-1:1}(n))}{\log\varphi_{j-1:1}(n)}\mathbf{1}_{\{\varphi_{j:1}(n)\ge m,\, j=1...,k\}}\right)\ \le\ \xi^k$$

for any $k\ge 1$. □

Lemma 6.23. *Under the assumptions of Theorem 6.21, let Y_n^* be given by* (6.38) *and define*

$$Z_{j,k}\ :=\ \left(\frac{\sigma_k^2-\sigma_{\varphi_j(k)}^2}{\sigma_k^2}\right)^{1/2}Z_j$$

for $j\ge 1$ and $k\ge 0$. Then

$$Y_{n,k}^*\ =\ \frac{\sigma_{\varphi_{k:1}(n)}}{\sigma_n}Z_{k+1}+\sum_{j=1}^{k}\frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n}Z_{j,\varphi_{j-1:1}(n)}$$

for each $n\ge n_0$, and the following assertions hold true:

(a) $\mathscr{L}(Y_{n,k}^*)=$ *Normal* $(0,1)$ *for each $n\ge n_0$, in fact this holds even true for the conditional law of $Y_{n,k}^*$ given $\mathfrak{A}_k=\sigma(\varphi_1,...,\varphi_k)$.*

(b) *For each $j\ge 1$ and $k\ge 0$, $Z_{j,k}$ and $C_{j,k}$ have identical first and second conditional moments given $\mathfrak{A}_{j-1}$ as defined in* (6.42)*, whence the same holds true for $Z_{j,\varphi_{j-1:1}(n)}$ and $C_{j,\varphi_{j-1:1}(n)}$ for any $n\ge n_0$.*

(c) *For each $j\ge 1$, $\|Z_{j,k}\|_s=O(\|C_k\|_s)=O(\log^{\gamma-\beta}k)$ as $k\to\infty$.*

(d) *For all $k,n\ge 1$,*

$$\sum_{j=1}^{k}\left\|\frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n}\left(C_{j,\varphi_{j-1:1}(n)}-Z_{j,\varphi_{j-1:1}(n)}\right)\right\|_s\ \le\ \frac{g(k)}{\log^\delta n}.$$

where $g(k):=M(1+k\log^{-(\beta-\delta)}n)$ for a suitable constant $M\in\mathbb{R}_>$.

Proof. We only prove the last two parts and leave the other assertions to the reader [☞ Problem 6.27].

(c) The assumptions (6.29), (6.30) of the theorem directly imply

$$\|C_{j,k}\|_s\ =\ \|C_k\|_s\ =\ \sigma_k^{-1}\|B_k-\mu_k+\mu_{\tau_k}\|_s\ =\ O(\log^{\gamma-\beta}k)$$

as $k\to\infty$, so that only $\|Z_{j,k}\|_s$ remains to be examined. In the following, we will make use of the facts that

$$0\ \le\ -\log\left(\frac{\varphi(k)\vee 1}{k}\right)\ \le\ \log k,$$

$$\mathbb{E}(Z_{j,k}^2|\mathfrak{A}_{j-1}) = \mathbb{E}(C_{j,k}^2|\mathfrak{A}_{j-1}) = \frac{\sigma_k^2 - \mathbb{E}\sigma_{\varphi(k)}^2}{\sigma_k^2} \quad \text{a.s.}$$

$$\text{and} \quad \mathbb{E}(Z_{j,k}^2|\mathfrak{A}_j) = \frac{\sigma_k^2 - \sigma_{\varphi(k)}^2}{\sigma_k^2} \quad \text{a.s.}$$

for all $j,k \geq 1$. First, we point out that

$$\begin{aligned}
\mathbb{E}\sigma_{\varphi(k)}^2 &= \theta\,\mathbb{E}\log^{2\beta}(\varphi(k)\vee 1) + O(\log^{2\gamma} k) \\
&= \theta\,\log^{2\beta} k + \theta\,\mathbb{E}\log^{2\beta}\left(\frac{\varphi(k)\vee 1}{k}\right) + O(\log^{2\gamma} k) \\
&= \theta\,\log^{2\beta} k + \theta\,\mathbb{E}\log^{2}\left(\frac{\varphi(k)\vee 1}{k}\right) O(1\vee\log^{2\beta-2} k) + O(\log^{2\gamma} k) \\
&= \theta\,\log^{2\beta} k + O(\log^{(2\beta-2)\vee 2\gamma} k)
\end{aligned}$$

where (6.30) has been used for the last line. Second, the previous expansion is utilized to give

$$\begin{aligned}
\left|\frac{\sigma_{\varphi(k)}^2 - \mathbb{E}\sigma_{\varphi(k)}^2}{\sigma_k^2}\right|^{s/2} &= \frac{1}{\log^{s\beta} k}\left|\log^{2\beta}\left(\frac{\varphi(k)\vee 1}{k}\right) + O(\log^{(2\beta-2)\vee 2\gamma} k)\right|^{s/2} \\
&= \frac{1}{\log^{s} k}\,\log^{s}\left(\frac{\varphi(k)\vee 1}{k}\right) + O\left(\frac{1}{\log^{s\delta} k}\right),
\end{aligned} \tag{6.43}$$

where $\delta = 1\wedge(\beta-\gamma)$ should be recalled. Third, we have that

$$\begin{aligned}
\mathbb{E}(|Z_{j,k}|^s|\mathfrak{A}_j) &= \left|\frac{\sigma_k^2 - \sigma_{\varphi_j(k)}^2}{\sigma_k^2}\right|^{s/2} \mathfrak{n}_s \\
&\leq 2^{s/2}\,\mathfrak{n}_s\left(\left|\mathbb{E}(C_{j,k}^2|\mathfrak{A}_{j-1})\right|^{s/2} + \left|\frac{\sigma_{\varphi_j(k)}^2 - \mathbb{E}\sigma_{\varphi_j(k)}^2}{\sigma_k^2}\right|^{s/2}\right) \\
&\leq 2^{s/2}\,\mathfrak{n}_s\left(\mathbb{E}(|C_{j,k}|^s|\mathfrak{A}_{j-1}) + \left|\frac{\sigma_{\varphi_j(k)}^2 - \mathbb{E}\sigma_{\varphi_j(k)}^2}{\sigma_k^2}\right|^{s/2}\right) \quad \text{a.s.}
\end{aligned}$$

where $\mathfrak{n}_s$ denotes the s^{th} absolute moment of a standard normal variable. Now take unconditional expectations and use (6.43) together with assumptions (6.29), (6.30) of the theorem to arrive at the desired conclusion for part (c), viz.

$$\begin{aligned}
\|Z_{j,k}\|_s^s &= \mathbb{E}\left(\mathbb{E}(|Z_{j,k}|^s|\mathfrak{A}_j)\right) \\
&\leq 2^{s/2}\,\mathfrak{n}_s\left(\|C_{j,k}\|_s^s + \mathbb{E}\left|\frac{\sigma_{\varphi_j(k)}^2 - \mathbb{E}\sigma_{\varphi_j(k)}^2}{\sigma_k^2}\right|^{s/2}\right)
\end{aligned}$$

$$= O\left(\frac{1}{\log^{s\delta} k}\right)$$

for all $j \geq 1$.

For the proof of (d), put $\Delta_{j,k} := C_{j,k} - Z_{j,k}$, let $m \in \mathbb{N}$ be such that (6.36) holds, and let K denote a generic positive constant which may differ from line to line. By the previous part, we have that, for any $j, n \geq 1$,

$$\begin{aligned}
\mathbb{E}\Big(\big|\Delta_{j,\varphi_{j-1:1}(n)}\big|^s \Big| \mathfrak{A}_{j-1}\Big) &= \sum_{k \geq n_0} \mathbf{1}_{\{\varphi_{j-1:1}(n)=k\}} \mathbb{E}\big|\Delta_{1,k}\big|^s \\
&\leq K \sum_{k \geq n_0} \mathbf{1}_{\{\varphi_{j-1:1}(n)=k\}} \log^{s\delta} k \\
&= K \log^{s\delta} \varphi_{j-1:1}(n)
\end{aligned}$$

a.s. on $\{\varphi_{j-1:1}(n) \geq n_0\}$. Consequently,

$$\begin{aligned}
&\left\|\frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n} \Delta_{j,\varphi_{j-1:1}(n)}\right\|_s^s \\
&\quad = \left\|\frac{\sigma_{\varphi_{j-1:1}(n)}}{\sigma_n} \Delta_{j,\varphi_{j-1:1}(n)} \mathbf{1}_{\{\varphi_{j-1:1}(n) \geq n_0\}}\right\|_s^s \\
&\quad \leq K \mathbb{E}\left(\frac{\log \varphi_{j-1:1}(n)}{\log n}\right)^{s\beta} \mathbb{E}\Big(\big|\Delta_{j,\varphi_{j-1:1}(n)}\big|^s \Big| \mathfrak{A}_{j-1}\Big) \mathbf{1}_{\{\varphi_{j-1:1}(n) \geq n_0\}} \\
&\quad \leq \frac{K}{\log^{s\delta} n} \mathbb{E}\left(\frac{\log \varphi_{j-1:1}(n)}{\log n}\right)^{s(\beta-\delta)} \mathbf{1}_{\{\varphi_{j-1:1}(n) \geq n_0\}} \\
&\quad \leq \frac{K}{\log^{s\delta} n} \left(\left(\frac{\log m}{\log n}\right)^{s(\beta-\delta)} + \xi^{k(s\gamma \wedge 1)}\right)
\end{aligned}$$

for all $j, n \geq 1$, where Lemma 6.22 has been used to provide the last estimate. The assertion now easily follows when taking the root of order s and summing over $j = 1, \ldots, k$. □

Results similar to Theorem 6.21 can also be stated in the branching case when (under the usual independence assumptions)

$$X_n \stackrel{d}{=} \sum_{i=1}^{N_n} X_{i,\tau_{n,i}} + B_n, \quad n \geq n_0, \tag{6.44}$$

with

$$1 \leq N_n < \infty \text{ a.s. for each } n \geq 1 \quad \text{and} \quad \limsup_{n \to \infty} \mathbb{E} N_n > 1. \tag{6.45}$$

We do not intend to embark on an extensive discussion of this case but confine ourselves to the statement of one such generalization, Theorem 6.24 below which appears in a similar form in [93, Thm. 5.1]. We also provide an appropriate setup for its proof, which follows similar lines as the one just given, but leave further details to the interested reader [☞ Problem 6.29]. Naturally, the previously used notation will be in force throughout unless stated otherwise.

Before stating the result, note that (6.44) after normalization becomes

$$\widehat{X}_n \stackrel{d}{=} \sum_{i=1}^{N_n} \frac{\sigma_{i,\tau_{n,i}}}{\sigma_n} \widehat{X}_{i,\tau_{n,i}} + C_n \;=\; \sum_{i\ge 1} T_{n,i} \widehat{X}_{i,\tau_{n,i}} + C_n, \tag{6.46}$$

where

$$C_n := \frac{1}{\sigma_n}\left(B_n - \mu_n + \sum_{i=1}^{N} \mu_{\tau_{n,i}}\right) \quad \text{and} \quad T_{n,i} := \frac{\sigma_{i,\tau_{n,i}}}{\sigma_n} \mathbf{1}_{\{N\ge i\}}$$

for $i,n \ge 1$.

Theorem 6.24. *Let $s = 2+\alpha$ for $\alpha \in (0,1]$ and $(X_n)_{n\ge 0}$ a sequence of random variables in L^s satisfying* (6.27) *under the usual conditions and* (6.45)*. Suppose further that* (6.29) *and* (6.30) *are valid for some $\beta,\theta \in \mathbb{R}_>$ and $0 \le \gamma < \beta$. Finally, assume that*

$$\limsup_{n\to\infty} \mathbb{E}\log\left(\frac{1}{n}\prod_{i=1}^{N_n}(\tau_{n,i}\vee 1)\right) < 0, \tag{6.47}$$

$$\limsup_{n\to\infty} \mathbb{E}\log\left(\sum_{i=1}^{N_n}\mathbf{1}_{\{\tau_{n,i}=n\}}\right) = 0, \tag{6.48}$$

$$\sup_{n\ge 1}\left\|\log\left(\frac{\tau_{n,1}\vee 1}{n}\right)\right\|_s < \infty, \tag{6.49}$$

$$\left\|\sum_{i=2}^{N_n}\log(\tau_{n,i}\vee 1)\right\|_s \;=\; O(\log^{\gamma} n). \tag{6.50}$$

Then $\widehat{X}_n \xrightarrow{d} Normal(0,1)$, in fact

$$\zeta_s(\mathscr{L}(\widehat{X}_n), Normal(0,1)) \;=\; O(\log^{-\alpha\delta} n)$$

as $n\to\infty$, where $\delta := (\beta-\gamma)\wedge 1$.

Proof (setup). Based on the family $(\tau_{n,i})_{n\ge 0, i\ge 1}$, define random maps $\phi_1,\phi_2,\dots$ by

$$\phi_i : \Omega\times\mathbb{N}_0 \to \mathbb{N}_0, \quad (\omega,n)\mapsto \tau_{n,i}(\omega),$$

and let $(N_i(\mathsf{v}),\phi_{\mathsf{v}i},B_i(\mathsf{v}),C_i(\mathsf{v}))_{i\ge 1}$ for $\mathsf{v}\in\mathbb{T}$ be iid copies of $(N_i,\phi_i,B_i,C_i)_{i\ge 1}$. Put further $\varphi_\varnothing(n) := n$ and

$$\varphi_{\mathsf{v}}(n) := \phi_{\mathsf{v}_1\dots\mathsf{v}_k}\circ\phi_{\mathsf{v}_1\dots\mathsf{v}_{k-1}}\circ\dots\circ\phi_{\mathsf{v}_1}(n),$$

$$I_n(\mathsf{v}) := \mathbf{1}_{\{\mathsf{v}_1 \le N_n(\varnothing), \mathsf{v}_2 \le N_n(\mathsf{v}_1), \ldots, \mathsf{v}_k \le N_n(\mathsf{v}_1 \ldots \mathsf{v}_{k-1})\}},$$
$$J_n(\mathsf{v}) := I_n(\mathsf{v}) \prod_{j=1}^{k} \mathbf{1}_{\{i : \mathbb{V}\mathrm{ar} X_i > 0\}}(\varphi_{\mathsf{v}_1 \ldots \mathsf{v}_j}(n))$$

for $k, n \ge 1$ and $\mathsf{v} = \mathsf{v}_1 \ldots \mathsf{v}_k \in \mathbb{T} \backslash \{\varnothing\}$. Then it is not dificult to verify that

$$\widehat{X}_n \stackrel{d}{=} \sum_{|\mathsf{v}|=k} J_n(\mathsf{v}) \frac{\sigma_{\varphi_\mathsf{v}(n)}}{\sigma_n} \widehat{X}_{\varphi_\mathsf{v}(n)}(\mathsf{v}) + \sum_{j=0}^{k-1} \sum_{|\mathsf{v}|=j} I_n(\mathsf{v}) \frac{\sigma_{\varphi_\mathsf{v}(n)}}{\sigma_n} C_{\varphi_\mathsf{v}(n)}(\mathsf{v}) \;=:\; \widehat{X}_{n,k}$$

for any $k \ge 1$ and $n \ge n_0$, where $\widehat{X}_j(\mathsf{v})$ denotes a copy of $\widehat{X}_j$ independent of all other occurring variables, for each $j \ge 0$ and $\mathsf{v} \in \mathbb{T}$. Notice that $\widehat{X}_k$ is either in $L^s_{0,1}$ or equal to 0. Finally, put

$$Y_{n,k} \;:=\; \sum_{|\mathsf{v}|=k} J_n(\mathsf{v}) \frac{\sigma_{\varphi_\mathsf{v}(n)}}{\sigma_n} Z(\mathsf{v}) + \sum_{j=0}^{k-1} \sum_{|\mathsf{v}|=j} J_n(\mathsf{v}) \frac{\sigma_{\varphi_\mathsf{v}(n)}}{\sigma_n} C_{\varphi_\mathsf{v}(n)}(\mathsf{v}),$$
$$Y^*_{n,k} \;:=\; \sum_{|\mathsf{v}|=k} J_n(\mathsf{v}) \frac{\sigma_{\varphi_\mathsf{v}(n)}}{\sigma_n} Z(\mathsf{v}) + \sum_{j=0}^{k-1} \sum_{|\mathsf{v}|=j} J_n(\mathsf{v}) \frac{\sigma_{\varphi_\mathsf{v}(n)}}{\sigma_n} S_n(\mathsf{v}) Z(\mathsf{v}),$$

where $(Z(\mathsf{v}))_{\mathsf{v} \in \mathbb{T}}$ is a family of iid standard normals independent of all other occurring variables and

$$S_n(\mathsf{v}) := \frac{1}{\sigma_{\varphi_\mathsf{v}(n)}} \left(\sigma^2_{\varphi_\mathsf{v}(n)} - \sum_{i \ge 1} J_n(\mathsf{v}i) \sigma^2_{\varphi_{\mathsf{v}i}(n)} \right)^{1/2}.$$

The reader should check that $\mathscr{L}(Y^*_{n,k}) = \mathit{Normal}(0,1)$. With the help of these settings, the reader should be able to prove the result by proceeding in a similar manner as for Theorem 6.21 [☞ Problem 6.29]. □

Problems

Problem 6.25. Given the assumptions (6.29) and (6.32) of Theorem 6.21, prove that

$$\sum_{n \ge 1} \mathbb{P}\left(\frac{\tau_n}{n} < 1 - \frac{1}{n} \right) < \infty$$

and use this to infer

$$\frac{\tau_n}{n} \to 1 \quad \text{and} \quad T_n = \frac{\sigma_{\tau_n}}{\sigma_n} \to 1 \quad \text{a.s.}$$

Problem 6.26. Given a recursion of the form (6.27) and the assumptions of Theorem 6.21, let U be a $\mathit{Unif}(-1,1)$ variable independent of all other occurring random variables and define $X_n' := X_n + U$ for $n \geq 0$, which has the same mean μ_n as X_n and positive variance ${\sigma_n'}^2 := \mathbb{V}\mathrm{ar} X_n + \mathbb{V}\mathrm{ar} U$ $(= \sigma_n^2 + \mathbb{V}\mathrm{ar} U$ if $\mathbb{V}\mathrm{ar} X_n > 0)$. Prove the following assertions which together provide a justification for the assumption that all X_n have positive variances in the proof of the afore-mentioned theorem:

(a) $X_n' \stackrel{d}{=} X_{\tau_n}' + B_n$ for each $n \in \mathbb{N}_0$.
(b) Theorem 6.21 applies to $(X_n')_{n\geq 0}$ with the same parameters θ, β and γ, thus (with $\delta = (\beta - \gamma) \wedge 1$)

$$\zeta_s(\mathscr{L}(\widehat{X}_n'), Normal(0,1)) \;=\; O(\log^{-\alpha\delta} n)$$

(c) $c := \sup_{n\geq 0} \|\widehat{X}_n\|_s < \infty$.
(d) $\ell_s(\widehat{X}_n, \widehat{X}_n') \leq {\sigma_n'}^{-1}(c + \|U\|_s) = O(\log^{-\beta} n)$ as $n \to \infty$.
(e) $\zeta_s(\widehat{X}_n, \widehat{X}_n') = O(\log^{-\alpha\beta} n) = O(\log^{-\alpha\delta} n)$ as $n \to \infty$.

Problem 6.27. Give a proof of Lemma 6.23(a) and (b).

Problem 6.28. In [93], the normal limit law of $\widehat{X}_n$ in Theorem 6.21 is identified in a different manner than here, namely via the following result [☞ Lemma 3.3 there]: Let $X, W \in L_{0,1}^2$ be independent random variables such that

$$X \;\stackrel{d}{=}\; qX + (1-q^2)^{1/2} W$$

for all $q \in (0,1)$. Then $\mathscr{L}(X) = Normal(0,1)$. Give a proof of this assertion.

Problem 6.29. Give a proof of Theorem 6.24 based on the setup described above.

6.6 Applications

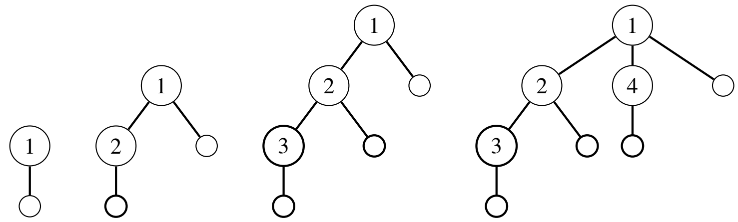
6.6.1 The total path length in a random recursive tree

A *recursive tree* of size n is a rooted unordered tree with labels $1,...,n$ such that the root is given label 1 and, for any $2 \leq k \leq n$, the labels of the vertices on the unique path from the root to the vertex k are increasing [☞ Figure 6.1 below]. It is unordered in the sense that we do not care about the order in which labels are assigned to the children of any vertex.

The following *dynamic construction* particularly shows that there are $(n-1)!$ distinct recursive trees of size n:

- Start with the root and label it by 1.
- At any stage $2 \leq k \leq n$ with $k-1$ vertices assigned and labeled $1,...,k-1$, pick anyone of these and attach a new node with label k to it.

For the general step one may also think of one external node attached to each of the $k-1$ assigned vertices of which one is picked and given label k [☞ top part



The first four steps in the dynamic construction of the following recursive tree.

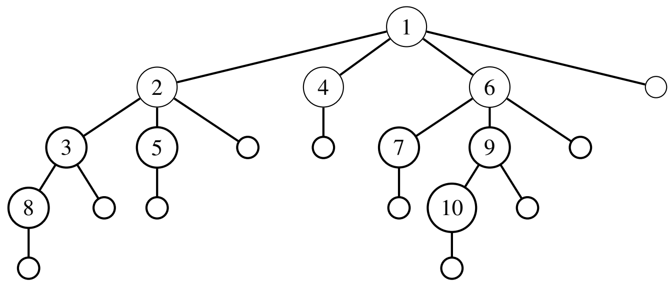


Fig. 6.1 A recursive tree with 10 (internal) nodes. In the uniform model, each of the external nodes, shown as empty circles, is equally likely to be picked for the next key 11.

of Figure 6.1 for an illustration]. Doing so uniformly at random, the construction provides a simple procedure to generate a so-called *random* or *uniform recursive tree* of size n. This is sometimes called the *uniform model* because it obviously induces a uniform distribution on the set of all $(n-1)!$ recursive trees of size n. In fact, by proceeding indefinitely as described, we obtain a nested sequence $\mathscr{T}_1, \mathscr{T}_2, ...$ of random trees such that, for each $n \geq 1$, $\mathscr{T}_n$ is a random recursive tree of size n. The associated natural filtration will be denoted $(\mathscr{G}_n)_{n\geq 1}$ hereafter.

Based on the uniform model, we will take a look at some interesting functionals of recursive trees with special attention to the *total path length* studied by DOBROW & FILL [34]. Let D_k denote the depth of the node k in $\mathscr{T}_n$ for any $k \geq n$, i.e., its distance from the root. Plainly, $D_1 = 0$ and $D_2 = 1$. Since node k for $k \geq 2$ is attached to any of $j = 1, ..., k-1$ with equal probability, we see that

$$D_k \;=\; 1 + D_{Z_{k-1}} \;=\; 1 + \sum_{j=1}^{k-1} \mathbf{1}_{\{Z_{k-1}=j\}} D_j, \tag{6.51}$$

where Z_{k-1} is independent of $D_1, ..., D_{k-1}$ with $\mathscr{L}(Z_{k-1}) = Unif\{1, ..., k-1\}$, and thus

$$\mathbb{E}D_k \;=\; 1 + \frac{1}{k-1} \sum_{j=1}^{k-1} \mathbb{E}D_j \;=\; H_{k-1}, \tag{6.52}$$

where, as in Section 1.4, H_k denotes the k^{th} harmonic sum. Second moment and variance of D_k can also be explicitly computed with the help of (6.51) and (6.52). The reader is asked to do so in Problem 6.39, the result for the variance being

$$\mathbb{V}\mathrm{ar}\, D_k \;=\; H_{k-1} - H_{k-1}^{(2)} \tag{6.53}$$

for each $k \geq 1$, where $H_n^{(2)} := \sum_{k=1}^n k^{-2}$. This was first shown by MOON [87] who more generally studied the distributional properties of the distance between two arbitrary nodes in $\mathscr{T}_n$. This and other interesting formulae like for the discrete pdf $\mathbb{P}(D_n = k)$ are discussed in the Problem Section. As for the asymptotic distribution of D_n, we see from the recursion (6.51) in combination with

$$\mathbb{E}D_n = \log n + O(1) \quad \text{and} \quad \mathbb{V}\mathrm{ar}\, D_n = \log n + O(1)$$

that Theorem 6.21 applies to yield asymptotic normality, viz.

$$\frac{D_n - \log n}{\log^{1/2} n} \overset{d}{\to} \mathit{Normal}(0,1). \tag{6.54}$$

This was earlier obtained by other means by DEVROYE [31] and MAHMOUD [80].

As already mentioned, we are particularly interested in the asymptotic behavior of the total path length in $\mathscr{T}_n$, defined as

$$\mathsf{TPL}_n \;:=\; \sum_{k=1}^n D_k$$

for $n \geq 1$. By (6.52), its mean is given by

$$\mu_n \;:=\; \mathbb{E}\mathsf{TPL}_n \;=\; \sum_{k=1}^{n-1} H_k \;=\; n(H_n - 1).$$

If $L_{n,k}$ denotes the number of nodes at level k in $\mathscr{T}_n$ for $k = 0, \dots, n$ (thus $L_{n,0} = 1$, $L_{n,n} = 0$ and $\sum_{k=0}^{n-1} L_{n,k} = n$), then we obviously also have the identity

$$\mathsf{TPL}_n \;=\; \sum_{k=1}^{n-1} k L_{n,k},$$

which will be useful for the proof of the following result, again taken from [80, Thm. 2] [☞ also [34, Thm. 2.1]].

Proposition 6.30. *The sequence* $(W_n, \mathscr{G}_n)_{n \geq 1}$, *where*

$$W_n \;:=\; \frac{\mathsf{TPL}_n - \mu_n}{n},$$

forms an L^2-bounded zero-mean martingale, viz.

$$\mathbb{E}W_n^2 \;=\; 2-\frac{H_n}{n}-H_n^{(2)} \tag{6.55}$$

for all $n \geq 1$. It thus converges a.s. and in L^2 to a zero-mean random variable $W \in L^2$ with variance

$$\mathbb{E}W^2 \;=\; \lim_{n\to\infty}\mathbb{E}W_n^2 \;=\; 2-\frac{\pi^2}{6} \;=:\; \aleph^2. \tag{6.56}$$

Proof. Since

$$\begin{aligned}\mathbb{E}(D_n|\mathscr{G}_{n-1}) \;&=\; \frac{1}{n-1}\sum_{k=1}^{n-1} k\,L_{n-1,k-1} \;=\; \frac{L_{n-1}}{n-1}+1\\ &=\; W_{n-1}+\frac{\mu_{n-1}}{n-1}+1 \;=\; W_{n-1}+H_{n-1} \quad \text{a.s.}\end{aligned}$$

and

$$\frac{\mu_{n-1}-\mu_n}{n} \;=\; \frac{(n-1)H_{n-1}-nH_n-1}{n} \;=\; -\frac{H_{n-1}}{n},$$

for $n \geq 2$, the martingale property follows from

$$\begin{aligned}\mathbb{E}(W_n|\mathscr{G}_{n-1}) \;&=\; \left(1-\frac{1}{n}\right)W_{n-1}+\frac{\mathbb{E}(D_n|\mathscr{T}_{n-1})}{n}+\frac{\mu_{n-1}-\mu_n}{n}\\ &=\; \left(1-\frac{1}{n}\right)W_{n-1}+\frac{W_{n-1}+H_{n-1}}{n}-\frac{H_{n-1}}{n}\\ &=\; W_{n-1} \quad \text{a.s.}\end{aligned}$$

Turning to the proof of L^2-boundedness and formula (6.55), we first note that

$$W_n-W_{n-1} \;=\; \frac{D_n}{n}+\frac{\mu_{n-1}-\mu_n}{n}-\frac{W_{n-1}}{n} \;=\; \frac{(D_n-\mathbb{E}D_n)-W_{n-1}}{n}$$

for each $n \geq 2$. It then follows with the help of the previous results that

$$\begin{aligned}\mathbb{E}W_n^2 \;&=\; \mathbb{E}\left(\left(1-\frac{1}{n}\right)W_{n-1}+\frac{1}{n}(D_n-\mathbb{E}D_n)\right)^2\\ &=\; \left(1-\frac{1}{n}\right)^2\mathbb{E}W_{n-1}^2+\frac{1}{n^2}\mathbb{V}\mathrm{ar}\,D_n+\frac{2(n-1)}{n^2}\,\mathbb{E}W_{n-1}D_n\\ &=\; \left(1-\frac{1}{n}\right)^2\mathbb{E}W_{n-1}^2+\frac{1}{n^2}\mathbb{V}\mathrm{ar}\,D_n+\frac{2(n-1)}{n^2}\,\mathbb{E}W_{n-1}\mathbb{E}(D_n|\mathscr{G}_{n-1})\end{aligned}$$

$$= \left(1-\frac{1}{n}\right)^2 \mathbb{E}W_{n-1}^2 + \frac{1}{n^2}\,\mathbb{V}\mathrm{ar}\,D_n + \frac{2(n-1)}{n^2}\,\mathbb{E}W_{n-1}(W_{n-1}+H_{n-1})$$

$$= \left(1-\frac{1}{n^2}\right)\mathbb{E}W_{n-1}^2 + \frac{1}{n^2}\,(H_{n-1}-H_{n-1}^{(2)}) \tag{6.57}$$

and therefore

$$\left(1-\frac{1}{n+1}\right)\mathbb{E}W_n^2 \;=\; \left(1-\frac{1}{n}\right)\mathbb{E}W_{n-1}^2 + \frac{1}{n(n+1)}\,(H_{n-1}-H_{n-1}^{(2)})$$

for each $n \ge 2$. We leave it as an exercise [☞ Problem 6.42] to derive from this the asserted formula (6.55) which particularly implies the L^2-boundedness of $(W_n)_{n\ge 1}$. All remaining assertions of the proposition then follow by an appeal to the sharpened L^2-version of the martingale convergence theorem. ☐

In view of the previous result, it is natural to ask next about further information on the limit distribution of W_n, i.e., of $\mathscr{L}(W)$. In order to make use of the contraction method, we want to study the normalization of TPL_n, given by

$$\widehat{\mathsf{TPL}}_n \;=\; \frac{\mathsf{TPL}_n - \mu_n}{\sigma_n}, \quad \sigma_n^2 := \mathbb{V}\mathrm{ar}\,\mathsf{TPL}_n.$$

Up to a scalar, it has the same almost sure limit as W_n because, by (6.56),

$$\widehat{\mathsf{TPL}}_n \;=\; \widehat{W}_n \;=\; \frac{W_n}{(\mathbb{E}W_n^2)^{1/2}} \;\to\; \frac{W}{\aleph} \quad \text{a.s.} \tag{6.58}$$

By using the main convergence theorem 6.11 proved in Section 6.4, we will now derive the SFPE valid for this limit and also convergence with respect to the Zolotarev metric ζ_3. To this end, we first need a distributional recursion of type (6.1) for TPL_n and therefore begin with the following basic lemma taken from [34].

Lemma 6.31. *For $n \ge 2$,*

$$\mathsf{TPL}_n \stackrel{d}{=} \mathsf{TPL}_{1,n-N_n} + \mathsf{TPL}_{2,N_n} + N_n \tag{6.59}$$

where $(\mathsf{TPL}_{1,k})_{k\ge 1}, (\mathsf{TPL}_{2,k})_{k\ge 1}$ and N_n are independent with $\mathscr{L}(N_n) = \mathit{Uniform}\{1,...,n-1\}$ and $\mathscr{L}(\mathsf{TPL}_{i,k}) = \mathscr{L}(\mathsf{TPL}_k)$ for each $k \ge 1$ and $i = 1,2$.

Proof. Every random recursive tree $\mathscr{T}_n$ of size $n \ge 2$ can be decomposed into two subtrees $\mathscr{T}_{1,n}^*, \mathscr{T}_{2,n}^*$ rooted at $1,2$ and of sizes $n-N_n, N_n$, respectively, where a node belongs to $\mathscr{T}_{2,n}^*$ if it has node 2 as an ancestor, and to $\mathscr{T}_{1,n}^*$ if not. As one can easily see, these two subtrees are conditionally independent random recursive trees given N_n, that is

$$\mathbb{P}((\mathscr{T}_{1,n}^*, \mathscr{T}_{2,n}^*) \in \cdot | N_n = k) = \mathbb{P}(\mathscr{T}_{n-k} \in \cdot) \otimes \mathbb{P}(\mathscr{T}_k \in \cdot)$$

for $k = 1, ..., n-1$. Regarding the total path length, one should observe that the depth of any node in $\mathscr{T}_{1,n}^*$ coincides with its depth within $\mathscr{T}_n$, whereas the depth of a node from $\mathscr{T}_{2,n}^*$ within $\mathscr{T}_n$ is one plus its depth within the subtree itself. Therefore, (6.59) holds true. It remains to prove that N_n is uniform on $\{1, ..., n-1\}$, which is a trivial fact in the case $n = 2$. Using the dynamic construction of recursive trees, we hence find by induction that

$$\begin{aligned}
\mathbb{P}(N_{n+1} = k) &= \frac{1}{n-1}\,(\mathbb{P}(N_{n+1} = k|N_n = k) + \mathbb{P}(N_{n+1} = k|N_n = k-1)) \\
&= \frac{1}{n-1}\Big(\mathbb{P}\big(\text{node } n+1 \text{ is appended to } \mathscr{T}_{1,n}^*\big||\mathscr{T}_{1,n}^*| = n-k\big) \\
&\qquad + \mathbb{P}\big(\text{node } n+1 \text{ is appended to } \mathscr{T}_{2,n}^*\big||\mathscr{T}_{2,n}^*| = k-1\big)\Big) \\
&= \frac{1}{n-1}\left(\frac{n-k}{n} + \frac{k-1}{n}\right) \\
&= \frac{1}{n-1}
\end{aligned}$$

as required. □

Rewriting equation (6.59) for $\widehat{\mathsf{TPL}}_n$, we obtain

$$\widehat{\mathsf{TPL}}_n \stackrel{d}{=} \frac{\sigma_{n-N_n}}{\sigma_n}\,\widehat{\mathsf{TPL}}_{1,n-N_n} + \frac{\sigma_{N_n}}{\sigma_n}\,\widehat{\mathsf{TPL}}_{2,N_n} + C_n, \quad n \geq 2, \tag{6.60}$$

where

$$C_n := \frac{N_n - \mu_n + \mu_{n-N_n} + \mu_{N_n}}{\sigma_n}.$$

The reader should notice the great similarity of this equation to the corresponding one in the `Quicksort` example [☞ (5.64)]. As a consequence, it could indeed also be further analyzed along the same lines as in 5.6 by using the minimal L^2-metric ℓ_2. On the other hand, it provides a good example where the general convergence theorem from Section 6.4 applies which only leaves us with the verification of its assumptions, with $s = 3$.

Let us begin with the statement of the following second order approximations of μ_n and σ_n^2 which are easily obtained from the previous results in combination with the expansion $H_n = \log n + \gamma + (2n)^{-1} + O(n^{-2})$ for the harmonic sum, where γ denotes Euler's constant. As $n \to \infty$,

$$\begin{aligned}
\mu_n &= n(H_n - 1) = n\log n + (\gamma - 1)n + 2 + O(n^{-1}), \\
\sigma_n^2 &= n^2\,\mathbb{E}W_n^2 = \aleph^2 n^2 - n\log n + (1-\gamma)n + O(1),
\end{aligned}$$

the last result being part of Problem 6.42. As in 5.6, we may further choose the N_n in such a way that $n^{-1}N_n \to U$ a.s. for some *Uniform*$(0,1)$ variable U. Using this and the expansion for μ_n, we find that

$$
\begin{aligned}
C_n \;&=\; \frac{n}{\sigma_n}\left(\frac{N_n}{n}+\frac{N_n}{n}\log\left(\frac{N_n}{n}\right)+\left(1-\frac{N_n}{n}\right)\log\left(1-\frac{N_n}{n}\right)\right) \\
\to\; C \;&:=\; \frac{U+U\log U+(1-U)\log(1-U)}{\aleph} \quad \text{a.s.}
\end{aligned}
$$

By the dominated convergence theorem, $\|n^{-1}N_n - U\|_3 \to 0$ as well and therefore

$$\|C_n - C\|_3 \to 0$$

which proves condition (6.18) of Theorem 6.11. In the notation of that result, we have further that

$$T_{n,1} = \frac{\sigma_{n-N_n}}{\sigma_n}, \quad T_{n,2} = \frac{\sigma_{N_n}}{\sigma_n} \quad \text{and} \quad T_{n,i} = 0 \text{ otherwise},$$

so that

$$(T_{n,1}, T_{n,2}, T_{n,3}, \ldots) \to (T_1, T_2, T_3, \ldots) := (1-U, U, 0, \ldots) \quad \text{a.s.}$$

It is now an easy task left for the reader as Problem 6.43 to check that all remaining assumptions of Theorem 6.11 are valid, so that we can state the following result about the total path length TPL_n.

Theorem 6.32. *If F_n denotes the distribution of the normalized total path length $\widehat{\mathsf{TPL}}_n$ and U, C are as stated above, then*

$$\zeta_3(F_n, F) \to 0,$$

where F, by (6.58) *the law of $W/\aleph$, is the unique solution in $\mathscr{P}^2_{0,1}(\mathbb{R})$ of the SFPE*

$$X \overset{d}{=} UX_1 + (1-U)X_2 + C \tag{6.61}$$

with X_1, X_2 being independent copies of X and also independent of U, C.

DOBROW & FILL [34] give a more detailed analysis of the limit distribution F in the previous theorem by studying its higher order moments and its Lebesgue density f, say. They state that f is everywhere positive, which can be obtained by adapting a result by TAN & HADJICOSTAS [107], and furthermore that f satisfies the functional equation

$$f(t) \;=\; \int_0^1 \int_{-\infty}^{\infty} \frac{1}{u} f\left(\frac{t-u+\mathscr{E}(u)-(1-u)w}{u}\right) f(w)\, dw\, du \tag{6.62}$$

for $\lambda\!\!\lambda$-almost all $t \in \mathbb{R}$, where $\mathscr{E}(u) := -u\log u - (1-u)\log(1-u)$ denotes the so-called *binary entropy function*. The reader is asked in Problem 6.44 to derive this equation from the SFPE (6.61) for F.

6.6.2 The number of leaves of a random recursive tree

There are of course many other interesting functionals that can be studied for a random recursive tree, for instance, its profile (= number of nodes at each level), its height, its number of nodes of a given degree, or the total internal and external path length. SMYTHE & MAHMOUD [102] provide an excellent survey of results in this direction and the relevant literature until 1995. For a more recent exposition of recursive trees including some further references, the reader may consult the book by DRMOTA [37, Ch. 6]. Here we confine ourselves to one more application, viz. the number of leaves (= nodes with no descendant) of $\mathscr{T}_n$, denoted as L_n. Note that $L_1 = L_2 = 1$ and that $n - L_n = |\mathscr{I}_n|$, where $\mathscr{I}_n$ is the (random) set of internal nodes of $\mathscr{T}_n$.

Lemma 6.33. *For each $n \ge 2$,*

$$L_n \;=\; L_{n-1} + \sum_{k \in \mathscr{I}_{n-1}} \mathbf{1}_{\{k \prec n\}}, \tag{6.63}$$

$$L_n \;\stackrel{d}{=}\; \mathbf{1}_{\{N_n \le n-2\}} L_{1,n-N_n} + L_{2,N_n}, \tag{6.64}$$

where $k \prec n$ means that node n is attached to node k. Moreover, $(L_{1,k})_{k\ge1}$, $(L_{2,k})_{k\ge1}$ and N_n are independent with $\mathscr{L}(N_n) = \mathit{Uniform}\{1,\dots,n-1\}$ and $\mathscr{L}(L_{i,k}) = \mathscr{L}(L_k)$ for each $k \ge 1$ and $i = 1,2$.

Proof. The first identity follows because, when node n is added to $\mathscr{T}_{n-1}$, the number of leaves increases by one only if it is attached to an internal node of $\mathscr{T}_{n-1}$ but remains unchanged otherwise. For (6.64), we recall the proof of Lemma 6.31 and let $L_{1,n}^*, L_{2,n}^*$ denote the number of leaves of the subtrees $\mathscr{T}_1^*, \mathscr{T}_2^*$ rooted at $1,2$ and of sizes $n - N_n, N_n$, respectively. Then it is clear that

$$\mathbb{P}((L_{1,n}^*, L_{2,n}^*) \in \cdot | N_n = k) = \mathbb{P}(L_{n-k} \in \cdot) \otimes \mathbb{P}(L_k \in \cdot)$$

for $k = 1,\dots,n-1$. Moreover, if $\mathscr{T}_{1,n}^*$ consists of node 1 only, thus $N_n = n-1$, then 1 is a leaf in this subtree but *not* in $\mathscr{T}_n$ (having descendant 2 there). In this case $L_{1,n}^* = L_{1,n-N_n}$ equals 1 but does not contribute to L_n. After these observations, (6.64) easily follows. □

(6.63) is very useful to derive recursions for the first and second moment of L_n that can be solved to give explicit formulae. By taking expectations and using

$$\mathbb{E}\left(\sum_{k \in \mathscr{I}_{n-1}} \mathbf{1}_{\{k \prec n\}} \middle| \mathscr{I}_{n-1} \right) \;=\; \frac{|\mathscr{I}_{n-1}|}{n-1} \;=\; 1 - \frac{L_{n-1}}{n-1},$$

we find as in [89, eq. (26)] that

$$\mathbb{E}L_n \;=\; \mathbb{E}L_{n-1} + \mathbb{E}\left(\sum_{k\in\mathscr{I}_{n-1}} \mathbf{1}_{\{k\prec n\}}\right) \;=\; \left(1-\frac{1}{n-1}\right)\mathbb{E}L_{n-1}+1$$

for $n \geq 2$, which upon straightforward calculations shows that

$$\mathbb{E}L_n \;=\; \frac{n}{2} \tag{6.65}$$

for each $n \geq 2$. As for the second moment, (6.63) provides us with

$$\begin{aligned}
\mathbb{E}L_m^2 \;&=\; \mathbb{E}L_{m-1}^2 + \mathbb{E}\left(\sum_{k\in\mathscr{I}_{m-1}} \mathbf{1}_{\{k\prec n\}}\right)^2 + 2\,\mathbb{E}\left(L_{m-1}\sum_{k\in\mathscr{I}_{m-1}} \mathbf{1}_{\{k\prec n\}}\right)\\
&=\; \mathbb{E}L_{m-1}^2 + \mathbb{E}\left(\sum_{k\in\mathscr{I}_{m-1}} \mathbf{1}_{\{k\prec n\}}\right) + 2\,\mathbb{E}\left(L_{m-1}\,\mathbb{E}\left(\sum_{k\in\mathscr{I}_{m-1}} \mathbf{1}_{\{k\prec n\}}\middle|L_{m-1}\right)\right)\\
&=\; \mathbb{E}L_{m-1}^2 + 1 - \frac{\mathbb{E}L_{m-1}}{m-1} + 2\,\mathbb{E}L_{m-1}\left(1-\frac{L_{m-1}}{m-1}\right)\\
&=\; \left(1-\frac{2}{m-1}\right)\mathbb{E}L_{m-1}^2 + n - \frac{1}{2}
\end{aligned}$$

which in turn is equivalent to

$$(m-2)(m-1)\,\mathbb{E}L_m^2 \;=\; (m-3)(m-2)\,\mathbb{E}L_{m-1}^2 + \frac{(m-2)(m-1)(2m-1)}{2}$$

for each $m \geq 3$. Summation over $m = 3,\ldots,n$ and using well-known formulae for $\sum_m m, \sum_m m^2$ and $\sum_m m^3$ then yields

$$\begin{aligned}
(n-2)(n-1)\,\mathbb{E}L_n^2 \;&=\; \sum_{m=3}^{n}\frac{(m-2)(m-1)(2m-1)}{2}\\
&=\; \sum_{m=1}^{n-2}\frac{m(m+1)(2m+3)}{2}\\
&=\; \sum_{m=1}^{n-2}\left(m^3+\frac{5}{2}m^2+\frac{3}{2}m\right)\\
&=\; \frac{(n-2)(n-1)}{4}\left((n-2)(n-1)+\frac{5(n-2)}{3}+\frac{3}{4}\right)
\end{aligned}$$

and thereupon

$$\mathbb{E}L_n^2 \;=\; \frac{n^2}{4}+\frac{n}{12}, \quad \text{thus} \quad \mathbb{V}\mathrm{ar}\,L_n \;=\; \frac{n}{12}. \tag{6.66}$$

This is the result also stated by NAJOCK & HEYDE [90, p. 677], but some confusion may arise when consulting other sources. Indeed, the reader will find a different formula for $\mathbb{E}L_n^2$ in [83, 102] which – after a short calculation – turns out to be the same as ours but for $\mathbb{E}L_{n+1}^2$. Yet another, fully mistaken formula appears in [37].

One may also derive the exact distribution of L_n from the recursion

$$\begin{aligned}\mathbb{P}(L_n = k) &= \mathbb{E}\left(\sum_{j\in\mathscr{I}_{n-1}} \mathbf{1}_{\{L_{n-1}=k-1,\, j\prec n\}}\right) + \mathbb{E}\left(\sum_{j\in\mathscr{T}_{n-1}\setminus\mathscr{I}_{n-1}} \mathbf{1}_{\{L_{n-1}=k,\, j\prec n\}}\right)\\ &= \frac{n-k}{n-1}\,\mathbb{P}(L_{n-1}=k-1) + \frac{k}{n-1}\,\mathbb{P}(L_{n-1}=k)\end{aligned}$$

for $n \ge 2$ and $k = 1,...,n-1$, which is again an immediate consequence of (6.63) and may be restated as

$$p_{n,k} = (n-k)\,p_{n-1,k-1} + k\,p_{n-1,k}$$

for $p_{n,k} := (n-1)!\,\mathbb{P}(L_n = k)$, where $p_{n,0} := 0$. Since $p_{1,1} = 1$, this is exactly the recursion defining the *Eulerian number of the first kind* $\left\langle {n-1 \atop k-1} \right\rangle$ for $k \ge 1$, where

$$\left\langle {n \atop k} \right\rangle = \sum_{j=0}^{k} (-1)^j \binom{n+1}{j} (k+1-j)^n$$

counts the number of permutations of $1,...,n$ with exactly k ascents, that is, elements which are strictly larger than the previous element [☞ for example COMTET [29, p. 51, 243-244] or KNUTH [72, p. 34-40]]. Consequently,

$$\mathbb{P}(L_n = k) = \frac{1}{(n-1)!}\left\langle {n-1 \atop k-1} \right\rangle \tag{6.67}$$

for $n \ge 2$ and $k = 1,...,n-1$.

Turning to our principal goal of finding the asymptotic law of

$$\widehat{L}_n = \frac{L_n - (n/2)}{(n/12)^{1/2}},$$

we first note that (6.64) provides us with the distributional recursion

$$\widehat{L}_n \stackrel{d}{=} \mathbf{1}_{\{N_n \le n-2\}} \left(1 - \frac{N_n}{n}\right)^{1/2} \widehat{L}_{1,n-N_n} + \left(\frac{N_n}{n}\right)^{1/2} \widehat{L}_{2,N_n} - \frac{\mathbf{1}_{\{N_n = n-1\}}}{(n/3)^{1/2}} \tag{6.68}$$

for $n \ge 2$, where $n^{-1}N_n \to U$ a.s. for some *Uniform*$(0,1)$ variable U. Hence, the limiting SFPE equals

$$X \stackrel{d}{=} (1-U)^{1/2}X_1 + U^{1/2}X_2$$

with iid X, X_1, X_2 independent of U, which uniquely characterizes the standard normal law. Verifying the conditions of Theorem 6.11 is now an easy task left to the reader [☞ Problem 6.45] and leads to the following CLT for L_n first obtained by NAJOCK & HEYDE [90] who were motivated by an application in philology, namely stemma (= family tree of preserved copies of ancient manuscripts) reconstruction.

Theorem 6.34. *The normalized leaf number $\widehat{L}_n$ with law F_n, say, converges in distribution to a standard normal distribution, in fact*

$$\zeta_3(F_n, Normal(0,1)) \to 0$$

as $n \to \infty$.

A connection of the number of leaves distribution with a certain urn model is discussed in Problem 6.46

6.6.3 Size and total path length of a random m-ary search tree

At the end of Section 1.4, we have briefly described how to generate a *binary search tree (BST)* from a permutation $\pi = (\pi_1, ..., \pi_n)$ of $1, ..., n$, or in fact any finite totally ordered set of n distinct elements (keys). The BST forms a fundamental data structure in computer science and has the following three properties:

(BST-1) Keys stored in the left subtree of a node are always less than the node's key.
(BST-2) Keys stored in the right subtree of a node are always larger than the node's key.
(BST-3) The left and right subtrees are both also BST's.

Its main advantage over other data structures is that related sorting algorithms (like `Quicksort`) and search algorithms can be very efficient due to the fact that on average, when assuming random input, the maximal path length (= the height of the BST) grows only logarithmically in the size of the data set it stores.

Search trees of higher branching degrees were first suggested by MUNTZ & UZGALIS [88] to solve internal memory problems with large data sets. The *m-ary search tree* generalizes the BST and is generated as follows from a permutation or data set $\pi = (\pi_1, ..., \pi_n)$ of length n:

(mST-1) Each node has maximal capacity $m-1$ and the keys it contains are stored in increasing order.
(mST-2) If $n < m$, all keys are in the root.
(mST-3) If $n \geq m$, then $\pi_1, ..., \pi_{m-1}$ are stored in the root and $\pi_m, ..., \pi_n$ go into one of the m subtrees subject to the condition that, when $(\pi_{(1)}, ..., \pi_{(m-1)})$ denotes the increasing order statistics of $\pi_1, ..., \pi_{m-1}$ and $\pi_{(m)} = -\pi_{(0)} := \infty$, then x goes into the j^{th} subtree if $\pi_{(j-1)} < x < \pi_{(j)}$ for $j = 1, ..., m-1$.
(msT-4) All subtrees are also m-ary search trees.

The construction is illustrated in Figure 6.2 for a ternary search tree ($m = 3$). For furtrher background information we refer to the monography by MAHMOUD [81, Ch. 3].

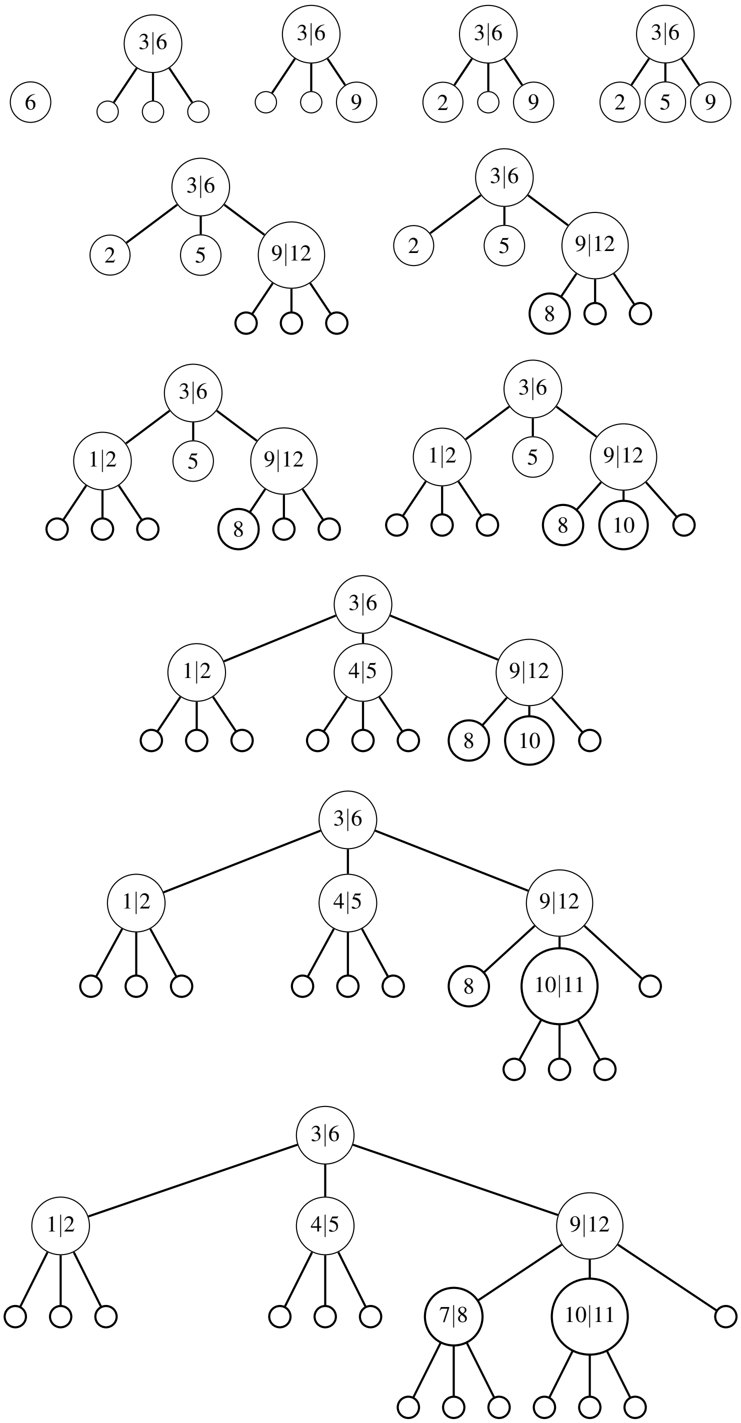


Fig. 6.2 The permutation $(6,3,9,2,5,12,8,1,10,4,11,7)$ from Example 1.12 stored in a ternary search tree of size 6 [☞ also Figure 1.2].

In the case when the input π is chosen uniformly at random from the set of permutations of length n, the obtained random tree $\mathscr{T}_{m,n}$, say, is called a *random m-ary search tree*. The case $m = 2$ leads back to a *random BST*.

Total path length of $\mathscr{T}_{m,n}$

As for a random recursive tree, we are interested in the total path length of $\mathscr{T}_{m,n}$, which is defined as the total sum of depths of nodes containing at least one key and again denoted by TPL_n. Then $\mathsf{TPL}_i = i$ for $i = 0,\ldots,m-1$ and

$$\mathsf{TPL}_n \stackrel{d}{=} \sum_{i=1}^{m} \mathsf{TPL}_{i,Z_{n,i}} + n - m + 1, \quad n \geq m, \tag{6.69}$$

where $(\mathsf{TPL}_{i,k})_{k\geq 0}$ for $i = 1,\ldots,m$ and $(Z_{n,1},\ldots,Z_{n,m})$ are mutually independent with $\mathscr{L}(\mathsf{TPL}_{i,k}) = \mathsf{TPL}_k$. Moreover, the $Z_{n,i}$ denote the random sizes (= number of nonempty nodes) of the subtrees rooted at level one and satisfy $Z_{n,1} + \ldots + Z_{n,m} = n - m + 1$. It is easily verified [☞ Problem 6.47] that

$$\mathbb{P}(Z_{n,1} = n_1,\ldots,Z_{n,m} = n_m) = \frac{(m-1)!}{n(n-1)\cdot\ldots\cdot(n-m+2)} = \frac{1}{\binom{n}{m-1}} \tag{6.70}$$

for all $(n_1,\ldots,n_m) \in \mathbb{N}_0^m$ with $\sum_{i=1}^m n_i = n - m + 1$ and, as $n \to \infty$,

$$\left(\frac{Z_{n,1}}{n-m+1},\ldots,\frac{Z_{n,m-1}}{n-m+1}\right) \to (U_{(1)} - U_{(0)},\ldots,U_{(m)} - U_{(m-1)}) \tag{6.71}$$

a.s. and in L^s for any $s > 0$, where $U_{(0)} := 0$, $U_{(m)} = 1$ and $(U_{(1)},\ldots,U_{(m-1)})$ denotes the order statistics of iid *Unif*$(0,1)$ variables $U_1,\ldots,U_{m-1}$.

Let us point out once again [☞ also at the end of Section 1.4] that in the binary case $m = 2$, the distributional recursion (6.69) for the total path length in a random BST equals exactly the number of key comparisons to sort a random list of n items by `Quicksort`.

In order to determine the asymptotic behavior of TPL_n after normalization, we first need information about its mean $\mu_n = \mathbb{E}\mathsf{TPL}_n$ and variance $\sigma_n^2 = \mathbb{V}\mathrm{ar}\,\mathsf{TPL}_n$ in the case $m \geq 3$. The recursions these quantities satisfy are far more difficult to solve than in the binary case, and we therefore contend ourselves with a statement of the following result taken from [81, Section 3.5] which also provides the details of its (analytic) derivation.

Lemma 6.35. *As $n \to \infty$, the mean and the variance of the total path length TPL_n in a random m-ary search tree on n keys $(m \geq 2)$ satisfy*

$$\mu_n = \frac{nH_n}{H_m - 1} + b_m n + O(n^\lambda) = \frac{n\log n}{H_m - 1} + c_m n + o(n^\lambda) \tag{6.72}$$

for some $\lambda < 1$ *and* $b_m, c_m \in \mathbb{R}$, *and*

$$\sigma_n^2 = \frac{1}{(H_m-1)^2}\left(\frac{(m+1)H_m^{(2)}-2}{m-1}-\frac{\pi^2}{6}\right)n^2+o(n^2). \tag{6.73}$$

The reader may want to assess that (6.73) in the case $m=2$ matches the asymptotic `Quicksort` variance (1.27) to be derived in Problem 1.16. In the ternary case $m=3$, MAHMOUD [81, p. 143] even provides exact formulae for μ_n and σ_n^2 which presumably discourage anyone who want to strive for the same for general m.

It is now easily checked that Corollary 6.13 with $s=3$ applies to

$$\mathsf{TPL}_n^* := \frac{\mathsf{TPL}_n - f(n)}{g(n)}$$

(notice that this is not the "exact" normalization $\widehat{\mathsf{TPL}}_n$) upon setting

$$f(n) := \frac{n\log n}{H_m-1}+c_m n,$$
$$g(n) := \sigma_m^* n, \quad \sigma_m^* := \frac{1}{H_m-1}\left(\frac{(m+1)H_m^{(2)}-2}{m-1}-\frac{\pi^2}{6}\right)^{1/2}$$

and further observing that, by (6.71),

$$\frac{g(Z_{n,i})}{g(n)} \to V_i := U_{(i)}-U_{(i-1)}, \quad i=1,...,m,$$
$$\frac{1}{g(n)}\left(n-m+1-f(n)+\sum_{i=1}^m f(Z_{n,i})\right) \to 1+\frac{1}{\sigma_m^*}\sum_{i=1}^m V_i\log V_i$$

in L^p for any $p>0$. Here is the result for $\widehat{\mathsf{TPL}}_n$:

Theorem 6.36. *The normalized total path length* $\widehat{\mathsf{TPL}}_n$ *of a random m-ary search tree on n keys converges in distribution to the unique solution in* $\mathscr{P}_{0,1}^2(\mathbb{R})$ *of the SFPE*

$$X \stackrel{d}{=} \sum_{i=1}^m V_i X_i + 1 + \frac{1}{\sigma_m^*}\sum_{i=1}^m V_i \log V_i \tag{6.74}$$

where $X_1,...,X_m$ *are independent of* $V_1,...,V_m$ *and iid copies of* X.

Size of $\mathscr{T}_{m,n}$
Fixing any $m \geq 3$, let S_n denote the size of $\mathscr{T}_{m,n}$, that is the number of nodes containing at least one key. Then $S_0 = 0$, $S_1 = ... = S_{m-1} = 1$, and

$$S_n \stackrel{d}{=} 1 + \sum_{i=1}^{m} S_{i,Z_{n,i}}, \quad n \geq m, \tag{6.75}$$

where $\mathscr{L}(S_{i,k}) = \mathscr{L}(S_k)$ for $i = 1, ..., m$ and $k \geq 0$, the $Z_{n,i}$ are defined as in (6.69), and the usual independence assumptions are made.

An analysis very similar to the one given for TPL_n enables us to derive a CLT for S_n, however, only for $3 \leq m \leq 26$. As before, the most difficult part is to provide approximations for $\mu_n := \mathbb{E}S_n$ and $\sigma_n^2 := \mathbb{V}\mathrm{ar}\, S_n$. The following result is due to MAHMOUD & PITTEL [82, Thm. 1] [☞ also [81, Thm. 3.1]], although earlier approximations for μ_n had already been obtained by KNUTH [72] and BAEZA-YATES [11].

Lemma 6.37. *As $n \to \infty$, the mean and the variance of the size S_n in a random m-ary search tree on n keys $(m \geq 3)$ satisfy for some $\alpha = \alpha_m < 2$:*

(a) *[72, 11, 82]*

$$\mu_n = \frac{n}{2(H_m - 1)} - \frac{1}{m-1} + O(n^{\alpha - 1}). \tag{6.76}$$

(b) *[82] If $3 \leq m \leq 26$, then $\alpha < 3/2$ and*

$$\sigma_n^2 = \gamma n + o(n) \tag{6.77}$$

for some positive $\gamma = \gamma_m$ only depending on m.

(c) *[82] If $m > 26$, then $\alpha > 3/2$ and*

$$\sigma_n^2 = \phi_{m,2}(\beta \log n)\, n^{2(\alpha - 1)} + o(n^{2(\alpha-1)}), \tag{6.78}$$

where $\phi_{m,2}$ is a positive bounded 2π-periodic function and $\beta = \beta_m \in \mathbb{R}_>$ a suitable constant.

A table of the values of α_m, γ_m and $\inf_n \phi_{m,2}(\beta \log n)$ for $3 \leq m \leq 85$ may also be found in both, [82] and [81].

So we see that the variance of S_n exhibits an unexpected phase transition from an asymptotic linear behavior if $3 \leq m \leq 26$ to a superlinear growth if $m \geq 27$. As a consequence, the following result, which may easily be inferred from Corollary 6.14 by checking its conditions, does only hold for the case $3 \leq m \leq 26$. It has also been derived by other means in [82, Thm. 2] for $3 \leq m \leq 15$, extended by LEW & MAHMOUD [74] to $m \leq 26$, and shown in full generality by CHERN & HWANG [26, Thm. 1] who also show that for $m > 26$ all integral moments of $n^{-(\alpha-1)}(S_n - \mu n)$,

$\mu := (2H_m - 2)^{-1}$, exhibit periodic behavior of the above kind, more precisely: if $m > 26$, then

$$\mathbb{E}\left(\frac{S_n - \mu n}{n^{\alpha-1}}\right)^k \;=\; \phi_{m,k}(\beta_m \log n)(1+o(1))$$

for all $k \in \mathbb{N}$ and some 2π-periodic functions $\phi_{m,k}$. They further point out that $n^{-(\alpha-1)}(S_n - \mu n)$ does not have a limit law and provide some further discussion of this fact in connection with the periodicity phenomenon [☞ [26, Section 4]].

Theorem 6.38. *If $3 \le m \le 26$, the normalized size $\widehat{S}_n = \sigma_n^{-1/2}(S_n - \mu_n)$ of a random m-ary search tree on n keys converges in distribution to a standard normal law.*

Proof. In order to use Corollary 6.14, put $f(n) = (2H_m - 2)^{-1}n$, $g(n) := \gamma n^{1/2}$ and consider $S_n^* := g(n)^{-1/2}(S_n - f(n))$ as in the previous example. Notice that, for $i = 1,\ldots,m-1$ and with $(V_1,\ldots,V_m)$ as before,

$$\frac{g(Z_{n,i})}{g(n)} \;\to\; V_i^{1/2}$$

in L^p for any $p > 0$, and that $\sum_{i=1}^m V_i = 1$. We leave it to the reader as Problem 6.48 to fill in all remaining details of the proof. □

Problems

In all subsequent problems, $(\mathscr{T}_n)_{n\ge1}$ denotes a random recursive tree sequence as described in Subsection 6.6.1 and D_k the depth of node k in $\mathscr{T}_n$ for $n \ge k$.

Problem 6.39. Prove the following assertions:

(a) For each $k \ge 3$,

$$\mathbb{E}D_k^2 \;=\; c_k + \sum_{j=1}^{k-1}\frac{c_j}{j} \;=\; H_{k-1} + \sum_{i=2}^{k-1}\frac{2H_{i-1}}{i}, \tag{6.79}$$

where $c_1 := 0$, $c_2 := 1$ and $c_k := 1 + \frac{2}{k-1}\sum_{j=1}^{k-2} H_j$ for $k \ge 3$.

(b) The variance of D_k is given by (6.53), that is

$$\mathbb{V}\mathrm{ar}\, D_k \;=\; H_{k-1} - H_{k-1}^{(2)}$$

for each $k \ge 1$.

(c) For each $p \ge 1$, there exists $\theta_p \in \mathbb{R}_>$ such that

$$\mathbb{E}D_k^p \;\le\; \theta_p \log^p k \tag{6.80}$$

for all $k \ge 1$.

Problem 6.40. Under the same assumptions as in the previous problem, prove the following assertions:

(a) If $L_{n,k}$ denotes the number of nodes of $\mathscr{T}_n$ at level k and $\ell_{n,k}$ its expectation, then

$$\mathbb{P}(D_n = k) \;=\; \frac{\mathbb{E}L_{n-1,k-1}}{n-1} \tag{6.81}$$

for $n \ge 2$ and $k = 1, \ldots, n-1$.

(b) [MEIR & MOON [86]] For each $n \ge 1$ and $0 \le k \le n$,

$$\ell_{n,k} \;=\; \ell_{n-1,k} + \frac{\ell_{n-1,k-1}}{n},$$

where $\ell_{0,k} = \delta_{0,k}$ for $k \ge 0$. Equivalently, $a_{n,k} := (n-1)!\,\ell_{n,k}$ satisfies

$$a_{n,k} \;=\; a_{n-1,k} + n\,a_{n-1,k-1}. \tag{6.82}$$

(c) [DONDAJEWSKI & SZYMAŃSKI [36]] Let $\left[{n \atop k}\right]$ denote the k^{th} *signless Stirling number of the first kind of order n*, which is defined as the coefficient of the x^k of the rising factorial $\langle x\rangle_n := x(x+1)\cdot\ldots\cdot(x+n-1)$, thus

$$\langle x\rangle_n \;=\; \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k.$$

Then

$$\ell_{n,k} \;=\; \frac{1}{(n-1)!}\begin{bmatrix} n \\ k \end{bmatrix} \tag{6.83}$$

for each $n \ge 1$ and $0 \le k \le n$. [Hint: Verify that $\left[{n \atop k}\right]$ and $a_{n,k}$ both satisfy (6.82) with the same boundary conditions.]

(d) [SZYMAŃSKI [106]] Use (6.81) and (6.83) to conclude

$$\mathbb{P}(D_n = k) \;=\; \frac{1}{(n-1)!}\begin{bmatrix} n-1 \\ k \end{bmatrix}$$

for each $n \ge 1$ and $k = 1, \ldots, n-1$.

Problem 6.41. [MOON [87]] Let $D_{i,j}$ denote the distance (= length of the shortest path) between nodes i, j in $\mathscr{T}_n$ for $n \ge i \vee j$, thus $D_{1,j} = D_j$ and $D_{j,j} = 0$. Prove the following generalizations of (6.52) and (6.53):

$$\mathbb{E}D_{i,j} \;=\; H_i + H_{j-1} - 2 + \frac{1}{i},$$

$$\mathbb{V}\mathrm{ar}\, D_{i,j} \;=\; H_i + H_{j-1} + 4 - 3H_i^{(2)} - H_{j-1}^{(2)} - \frac{4H_i}{i} + 3i - \frac{1}{i^2}.$$

[Hint: Use that $D_{i,j} = 1 + \sum_{k=1}^{j-1} D_{i,k} \mathbf{1}_{\{k \prec j\}}$ for arbitrary $1 \le i < j$, where $k \prec j$ means that j is attached to node k.]

Problem 6.42. Complete the proof of Proposition (6.30) by deriving (6.55) from (6.57). Show further that

$$\mathbb{E}W_n^2 \;=\; 2 - \frac{\pi^2}{6} - \frac{\log n}{n} + \frac{1-\gamma}{n} + O\left(\frac{1}{n^2}\right) \tag{6.84}$$

where γ denotes Euler's constant.

Problem 6.43. Complete the proof of Theorem 6.32 and show further, by mimicking the arguments from Section 5.6, that $\ell_p(F_n, F) \to 0$ for all $p > 0$.

Problem 6.44. Use the SFPE (6.61) to prove the functional equation (6.62) for the density of the asymptotic total path length in a random recursive tree.

Problem 6.45. Based on (6.68), give a proof of Theorem 6.34 with the help of Theorem 6.11.

Problem 6.46. BERNARD FRIEDMAN [55] proposed the following urn model: An urn initially contains W_n white and B_n black balls at time n. Each time a ball is drawn at random and then replaced together with α balls of the same and β balls of the opposite color. The famous Pólya urn is obtained if $\beta = 0$. Make a connection of Friedman's urn for suitable α, β with the distribution of the number of leaves L_n in a random recursive tree.

Problem 6.47. Letting $Z_{n,1}, \ldots, Z_{n,m}$ denote the level-one subtree sizes of a random m-ary search tree on n keys, prove that (6.70) and (6.71) hold true.

Problem 6.48. Complete the proof of Theorem 6.38.

Chapter 7
The smoothing transform: a stochastic linear recursion with branching
Part II: Fixed points

7.1 The smoothing transform with deterministic weights

References

1. Aldous, D.J., Bandyopadhyay, A.: A survey of max-type recursive distributional equations. Ann. Appl. Probab. **15**(2), 1047–1110 (2005)
2. Alsmeyer, G.: Renewal, recurrence and regeneration (in preparation). Springer-Verlag, New York (2010+)
3. Alsmeyer, G., Fuh, C.D.: Limit theorems for iterated random functions by regenerative methods. Stochastic Process. Appl. **96**, 123–142 (2001). Corrigendum in 97:341-345 (2002)
4. Alsmeyer, G., Iksanov, A., Rösler, U.: On distributional properties of perpetuities. J. Theoret. Probab. **22**(3), 666–682 (2009)
5. Alsmeyer, G., Rösler, U.: The best constant in the Topchii-Vatutin inequality for martingales. Statist. Probab. Lett. **65**(3), 199–206 (2003)
6. Asmussen, S.: Applied probability and queues, 2^{nd} edn. Springer-Verlag, New York (2003)
7. Athreya, K., McDonald, D., Ney, P.: Coupling and the renewal theorem. Amer. Math. Monthly **85**(10), 809–814 (1978)
8. Athreya, K.B., Dai, J.: Random logistic maps. I. J. Theoret. Probab. **13**(2), 595–608 (2000)
9. Athreya, K.B., Dai, J.J.: On the nonuniqueness of the invariant probability for i.i.d. random logistic maps. Ann. Probab. **30**(1), 437–442 (2002)
10. Athreya, K.B., Schuh, H.J.: Random logistic maps. II. The critical case. J. Theoret. Probab. **16**(4), 813–830 (2004) (2003)
11. Baeza-Yates, R.A.: Some average measures in *m*-ary search trees. Inform. Process. Lett. **25**(6), 375–381 (1987). DOI 10.1016/0020-0190(87)90215-8. URL `http://dx.doi.org/10.1016/0020-0190(87)90215-8`
12. Bauer, H.: Measure and integration theory, *de Gruyter Studies in Mathematics*, vol. 26. Walter de Gruyter & Co., Berlin (2001). Translated from the German by Robert B. Burckel
13. Benda, M.: A central limit theorem for contractive stochastic dynamical systems. J. Appl. Probab. **35**(1), 200–205 (1998)
14. Bender, E.A.: An introduction to mathematical modeling. John Wiley & Sons, New York-Chichester-Brisbane (1978). A Wiley-Interscience Publication
15. Billingsley, P.: The Lindeberg-Lévy theorem for martingales. Proc. Amer. Math. Soc. **12**, 788–792 (1961)
16. Billingsley, P.: Probability and measure, 2^{nd} edn. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York (1986)

17. Billingsley, P.: Convergence of probability measures, 2^{nd} edn. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York (1999). A Wiley-Interscience Publication
18. Birkhoff, G.D.: Proof of the ergodic theorem. Proc. Natl. Acad. Sci. USA **17**, 656–660 (1931)
19. Birkhoff, G.D.: What is the ergodic theorem? Amer. Math. Monthly **49**, 222–226 (1942)
20. Blackwell, D.: A renewal theorem. Duke Math. J. **15**, 145–150 (1948)
21. Borkovec, M., Klüppelberg, C.: The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. Ann. Appl. Probab. **11**(4), 1220–1241 (2001)
22. Breiman, L.: Probability. Addison-Wesley Publishing Company, Reading, Mass. (1968)
23. Burkholder, D.L.: Martingale transforms. Ann. Math. Statist. **37**, 1494–1504 (1966)
24. Burkholder, D.L., Davis, B.J., Gundy, R.F.: Integral inequalities for convex functions of operators on martingales. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 223–240. Univ. California Press, Berkeley, Calif. (1972)
25. Chamayou, J.F., Letac, G.: Explicit stationary distributions for compositions of random functions and products of random matrices. J. Theoret. Probab. **4**(1), 3–36 (1991)
26. Chern, H.H., Hwang, H.K.: Phase changes in random m-ary search trees and generalized quicksort. Random Structures Algorithms **19**(3-4), 316–358 (2001)
27. Choquet, G., Deny, J.: Sur l'équation de convolution $\mu = \mu * \sigma$. C. R. Acad. Sci. Paris **250**, 799–801 (1960)
28. Chow, Y.S., Teicher, H.: Probability theory. Independence, interchangeability, martingales, 3^{rd} edn. Springer Texts in Statistics. Springer-Verlag, New York (1997)
29. Comtet, L.: Advanced combinatorics. The art of finite and infinite expansions, enlarged edn. D. Reidel Publishing Co., Dordrecht (1974)
30. Dai, J.J.: A result regarding convergence of random logistic maps. Statist. Probab. Lett. **47**(1), 11–14 (2000)
31. Devroye, L.: Applications of the theory of records in the study of random trees. Acta Inform. **26**(1-2), 123–130 (1988)
32. Devroye, L., Kruszewski, P.: A note on the Horton-Strahler number for random trees. Inform. Process. Lett. **52**(3), 155–159 (1994)
33. Diaconis, P., Freedman, D.: Iterated random functions. SIAM Rev. **41**(1), 45–76 (1999)
34. Dobrow, R.P., Fill, J.A.: Total path length for random recursive trees. Combin. Probab. Comput. **8**(4), 317–333 (1999). Random graphs and combinatorial structures (Oberwolfach, 1997)
35. Dobrushin, R.: Prescribing a system of random variables by conditional distributions. Theor. Probab. Appl. **15**, 458–486 (1970)
36. Dondajewski, M., Szymański, J.: On the distribution of vertex-degrees in a strata of a random recursive tree. Bull. Acad. Polon. Sci. Sér. Sci. Math. **30**(5-6), 205–209 (1982)
37. Drmota, M.: Random trees. Springer-Verlag, Vienna, New York (2009). An interplay between combinatorics and probability
38. Drmota, M., Janson, S., Neininger, R.: A functional limit theorem for the profile of search trees. Ann. Appl. Probab. **18**(1), 288–333 (2008)
39. Durrett, R.: Probability: theory and examples, 2^{nd} edn. Duxbury Press, Belmont, CA (1996)
40. Dwass, M.: The total progeny in a branching process and a related random walk. J. Appl. Probability **6**, 682–686 (1969)
41. Edgar, G.A.: Integral, probability, and fractal measures. Springer-Verlag, New York (1998)
42. Edgar, G.A.: Measure, topology, and fractal geometry, 2^{nd} edn. Undergraduate Texts in Mathematics. Springer-Verlag, New York (2008)
43. Elstrodt, J.: Maß- und Integrationstheorie, 4^{th} edn. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin (2005). Grundwissen Mathematik. [Basic Knowledge in Mathematics]
44. Elton, J.H.: A multiplicative ergodic theorem for Lipschitz maps. Stochastic Process. Appl. **34**(1), 39–47 (1990)
45. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica **50**(4), 987–1007 (1982)

46. Erdös, P., Feller, W., Pollard, H.: A property of power series with positive coefficients. Bull. Am. Math. Soc. **55**, 201–204 (1949)
47. Falconer, K.J.: Random fractals. Math. Proc. Cambridge Philos. Soc. **100**(3), 559–582 (1986)
48. Falconer, K.J.: Fractal geometry. Mathematical foundations and applications. John Wiley & Sons Ltd., Chichester (1990)
49. Fama, E.: The behavior of stock market prices. J. Business **38**, 34–105 (1965)
50. Fekete, M.: Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. Math. Z. **17**(1), 228–249 (1923)
51. Feller, W.: An introduction to probability theory and its applications. Vol. II. Second edition. John Wiley & Sons Inc., New York (1971)
52. Fill, J.A., Janson, S.: A characterization of the set of fixed points of the Quicksort transformation. Electron. Comm. Probab. **5**, 77–84 (electronic) (2000)
53. Fill, J.A., Janson, S.: Quicksort asymptotics. J. Algorithms **44**(1), 4–28 (2002)
54. Fill, J.A., Janson, S.: The number of bit comparisons used by Quicksort: an average-case analysis. In: Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 300–307 (electronic). ACM, New York (2004)
55. Friedman, B.: A simple urn model. Comm. Pure Appl. Math. **2**, 59–70 (1949)
56. Furstenberg, H., Kesten, H.: Products of random matrices. Ann. Math. Statist. **31**, 457–469 (1960)
57. Gade, H.: Deep water exchanges in a sill fjord: a stochastic process. J. Phys. Oceanography **3**, 213–219 (1973)
58. Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**(1), 126–166 (1991)
59. Graf, S.: Statistically self-similar fractals. Probab. Theory Related Fields **74**(3), 357–392 (1987)
60. Grincevičius, A.: Products of random affine transformations. Lith. Math. J. **20**, 279–282 (1981)
61. Guégan, D., Diebolt, J.: Probabilistic properties of the β-ARCH-model. Statist. Sinica **4**(1), 71–87 (1994)
62. Hall, P., Heyde, C.C.: Martingale limit theory and its application. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York (1980). Probability and Mathematical Statistics
63. Helland, I.S.: A random exchange model with constant decrements (1974). Report No. 52, Mat. inst., adv. B. University of Bergen
64. Helland, I.S., Nilsen, T.S.: On a general random exchange model. J. Appl. Probability **13**(4), 781–790 (1976)
65. Hoare, C.A.R.: Algorithm 64: Quicksort. Comm. of the Assoc. for Computing Machinery **4**, 321 (1961)
66. Hoare, C.A.R.: Quicksort. Comput. J. **5**, 10–15 (1962)
67. Horton, R.E.: Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology. Geological Soc. of America Bulletin **56**(3), 275–370 (1945)
68. Hutchinson, J.E.: Fractals and self-similarity. Indiana Univ. Math. J. **30**(5), 713–747 (1981)
69. Kesten, H.: Random difference equations and renewal theory for products of random matrices. Acta Math. **131**, 207–248 (1973)
70. Kesten, H., Stigum, B.P.: A limit theorem for multidimensional Galton-Watson processes. Ann. Math. Statist. **37**, 1211–1223 (1966)
71. Kingman, J.F.C.: Subadditive ergodic theory. Ann. Probability **1**, 883–909 (1973). With discussion by D. L. Burkholder, Daryl Daley, H. Kesten, P. Ney, Frank Spitzer and J. M. Hammersley, and a reply by the author
72. Knuth, D.E.: The art of computer programming. Volume 3. Sorting and searching. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont. (1973). Addison-Wesley Series in Computer Science and Information Processing

73. Letac, G.: A contraction principle for certain Markov chains and its applications. In: Random matrices and their applications (Brunswick, Maine, 1984), *Contemp. Math.*, vol. 50, pp. 263–273. Amer. Math. Soc., Providence, RI (1986)
74. Lew, W., Mahmoud, H.M.: The joint distribution of elastic buckets in multiway search trees. SIAM J. Comput. **23**(5), 1050–1074 (1994)
75. Liggett, T.M.: An improved subadditive ergodic theorem. Ann. Probab. **13**(4), 1279–1285 (1985)
76. Lindley, D.V.: The theory of queues with a single server. Proc. Cambridge Philos Soc. **48**, 277–289 (1952)
77. Lindvall, T.: A probabilistic proof of Blackwell's renewal theorem. Ann. Probability **5**(3), 482–485 (1977)
78. Lindvall, T., Rogers, L.C.G.: On coupling of random walks and renewal processes. J. Appl. Probab. **33**(1), 122–126 (1996)
79. Maercker, G.: Statistical inference in conditional heteroscedastic autoregressive models. Ph.D. thesis, Technische Universität Braunschweig (1997)
80. Mahmoud, H.M.: Limiting distributions for path lengths in recursive trees. Probab. Engrg. Inform. Sci. **5**(1), 53–59 (1991)
81. Mahmoud, H.M.: Evolution of random search trees. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York (1992)
82. Mahmoud, H.M., Pittel, B.: Analysis of the space of search trees under the random insertion algorithm. J. Algorithms **10**(1), 52–75 (1989)
83. Mahmoud, H.M., Smythe, R.T.: Asymptotic joint normality of outdegrees of nodes in random recursive trees. Random Structures Algorithms **3**(3), 255–266 (1992)
84. Mandelbrot, B.: The variation of certain speculative prices. J. Business **36**, 394–419 (1963)
85. Mauldin, R.D., Williams, S.C.: Random recursive constructions: asymptotic geometric and topological properties. Trans. Amer. Math. Soc. **295**(1), 325–346 (1986)
86. Meir, A., Moon, J.W.: On the altitude of nodes in random trees. Canad. J. Math. **30**(5), 997–1015 (1978)
87. Moon, J.W.: The distance between nodes in recursive trees. In: Combinatorics (Proc. British Combinatorial Conf., Univ. Coll. Wales, Aberystwyth, 1973), pp. 125–132. London Math. Soc. Lecture Note Ser., No. 13. Cambridge Univ. Press, London (1974)
88. Muntz, R., Uzgalis, R.: Dynamic storage allocation for binary search trees in a two-level memory. In: Proceedings of Princeton Conference on Information Sciences and Systems, Vol. 4, pp. 345–349. Princeton, New Jersey (1971)
89. Na, H.S., Rapoport, A.: Distribution of nodes of a tree by degree. Math. Biosci. **6**, 313–329 (1970)
90. Najock, D., Heyde, C.C.: On the number of terminal vertices in certain random trees with an application to stemma construction in philology. J. Appl. Probab. **19**(3), 675–680 (1982)
91. Neininger, R.: On a multivariate contraction method for random recursive structures with applications to Quicksort. Random Structures Algorithms **19**(3-4), 498–524 (2001). Analysis of algorithms (Krynica Morska, 2000)
92. Neininger, R., Rüschendorf, L.: A general limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Probab. **14**(1), 378–418 (2004)
93. Neininger, R., Rüschendorf, L.: On the contraction method with degenerate limit equation. Ann. Probab. **32**(3B), 2838–2856 (2004)
94. Peres, Y., Schlag, W., Solomyak, B.: Sixty years of Bernoulli convolutions. In: Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), *Progr. Probab.*, vol. 46, pp. 39–65. Birkhäuser, Basel (2000)
95. Rachev, S.T., Rüschendorf, L.: Probability metrics and recursive algorithms. Adv. in Appl. Probab. **27**(3), 770–799 (1995)
96. Rösler, U.: A limit theorem for "Quicksort". RAIRO Inform. Théor. Appl. **25**(1), 85–100 (1991)
97. Rösler, U.: A fixed point theorem for distributions. Stochastic Process. Appl. **42**(2), 195–214 (1992)

98. Rösler, U.: On the analysis of stochastic divide and conquer algorithms. Algorithmica **29**(1-2), 238–261 (2001). Average-case analysis of algorithms (Princeton, NJ, 1998)
99. Rösler, U., Rüschendorf, L.: The contraction method for recursive algorithms. Algorithmica **29**(1-2), 3–33 (2001). Average-case analysis of algorithms (Princeton, NJ, 1998)
100. Rüschendorf, L.: Wasserstein metric. In: M. Hazewinkel (ed.) Encyclopaedia of mathematics. Supplement. Vol. II, pp. vi+631. Kluwer Academic Publishers, Dordrecht (2000)
101. Smith, W.L.: Renewal theory and its ramifications. J. Roy. Statist. Soc. Ser. B **20**, 243–302 (1958)
102. Smythe, R.T., Mahmoud, H.M.: A survey of recursive trees. Theory Probab. Math. Statist. (51), 1–27 (1995)
103. Steinsaltz, D.: Random logistic maps and Lyapunov exponents. Indag. Math. (N.S.) **12**(4), 557–584 (2001)
104. Strahler, A.N.: Hypsometric (area-altitude) analysis of erosional topology. Geological Soc. of America Bulletin **63**(11), 1117–1142 (1952)
105. Strahler, A.N.: Quantitative analysis of watershed geomorphology. Trans. of the American Geophysical Union **8**(6), 913–920 (1957)
106. Szymański, J.: On the complexity of algorithms on recursive trees. Theoret. Comput. Sci. **74**(3), 355–361 (1990)
107. Tan, K.H., Hadjicostas, P.: Some properties of a limiting distribution in Quicksort. Statist. Probab. Lett. **25**(1), 87–94 (1995)
108. Thorisson, H.: A complete coupling proof of Blackwell's renewal theorem. Stochastic Process. Appl. **26**(1), 87–97 (1987)
109. Vasershtein, L.N.: Markov processes over denumerable products of spaces describing large system of automata. Problemy Peredači Informacii **5**(3), 64–72 (1969)
110. Vatutin, V.A., Topchiĭ, V.A.: The maximum of critical Galton-Watson processes, and left-continuous random walks. Theory Probab. Appl. **42**(1), 17–27 (1998)
111. Vervaat, W.: On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. Adv. in Appl. Probab. **11**(4), 750–783 (1979)
112. Wald, A.: Sequential Analysis. John Wiley & Sons Inc., New York (1947)
113. Weiss, A.: ARMA models with ARCH errors. J. Time Ser. Anal. **5**(2), 129–143 (1984)
114. Wu, W.B., Shao, X.: Limit theorems for iterated random functions. J. Appl. Probab. **41**(2), 425–436 (2004)
115. Wu, W.B., Woodroofe, M.: A central limit theorem for iterated random functions. J. Appl. Probab. **37**(3), 748–755 (2000)
116. Zolotarev, V.M.: Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. Theory Probab. Appl. **21**(4), 721–737 (1976)

Appendix A
A quick look at some ergodic theory and theorems

A.1 Measure-preserving transformations and ergodicity

Given a probability space $(\mathbb{Y},\mathscr{A},\mathbf{P})$, a measurable mapping $T:\mathbb{Y}\to\mathbb{Y}$ is called *measure-preserving transformation of* $(\mathbb{Y},\mathscr{A},\mathbf{P})$ if $\mathbf{P}(T\in\cdot)=\mathbf{P}$. The iterations $(T^n(y))_{n\ge 0}$ for any initial value $y\in\mathbb{Y}$ provide an *orbit* or *trajectory* of the *dynamical system* generated by T. If y is picked according to $\mathbf{P}$ and thus formally replaced with a random element $Y_0:(\Omega,\mathfrak{A},\mathbb{P})\to(\mathbb{Y},\mathscr{A})$ having law $\mathbf{P}$, then

$$Y_n := T^n(Y_0),\quad n\ge 0$$

forms a *stationary sequence*, the defining property being that

$$(Y_n,...,Y_{n+m})\ \stackrel{d}{=}\ (Y_0,...,Y_m)\quad\text{for all } m,n\in\mathbb{N}_0. \tag{A.1}$$

When studying the statistical properties of an orbit $(T^n(y))_{n\ge 0}$, for instance by looking at absolute or relative frequencies

$$N_{T,n}(y,A) := \sum_{k=0}^{n}\mathbf{1}_A(T^k(y))\quad\text{or}\quad h_{T,n}(y,A) := \frac{1}{n+1}\sum_{k=0}^{n}\mathbf{1}_A(T^k(y)),$$

respectively, for $A\in\mathscr{A}$, the notion of T*-invariance* arises quite naturally. A set $A\in\mathscr{A}$ is called T*-invariant* or *invariant under* T if $T^{-1}(A)=A$. Their collection $\mathscr{I}_T^*$ forms a σ-field, called *σ-field of T-invariant sets* or just *invariant σ-field of T*. Its completion $\mathscr{I}_T$, say, within $\mathscr{A}$ consists of all sets $A\in\mathscr{A}$ for which T-invariance holds $\mathbf{P}$-a.s., thus

$$\mathscr{I}_T = \{A\in\mathscr{A} : T^{-1}(A)=A\ \mathbf{P}\text{-a.s.}\}$$

Obviously, $y\in A$ for a [$\mathbf{P}$-a.s.] T-invariant set A entails $T^n(y)\in A$ [$\mathbf{P}$-a.s.] for all $n\in\mathbb{N}$. As a consequence , if $\mathscr{I}_T$ contains a set A having $0<\mathbf{P}(A)<1$, then the distribution of $\boldsymbol{Y}=(Y_n)_{n\ge 0}$ under $\mathbf{P}$ may be decomposed as

$$\mathbf{P}(\boldsymbol{Y}\in\cdot)\ =\ \mathbf{P}(A)\,\mathbb{P}(\boldsymbol{Y}\in\cdot|Y_0\in A)+(1-\mathbf{P}(A))\,\mathbb{P}(\boldsymbol{Y}\in\cdot|Y_0\in A^c) \tag{A.2}$$

and $\boldsymbol{Y}$ remains stationary under both, $\mathbb{P}(\cdot|Y_0 \in A)$ and $\mathbb{P}(\cdot|Y_0 \in A^c)$ [☞ Problem A.6]. If no such decomposition exists or, equivalently, $\mathscr{I}_T$ is $\mathbf{P}$-trivial, then the sequence $\boldsymbol{Y}$ as well as the associated transformation T are called *ergodic*.

In probability theory, a stationary sequence $\boldsymbol{Y} = (Y_n)_{n\ge 0}$ of $\mathbb{Y}$-valued random variables on a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ is simply defined by property (A.1) and hence does not require the ergodic-theoretic setting of a measure-preserving transformation. On the other hand, when considering the associated coordinate model $(\mathbb{Y}^{\mathbb{N}_0}, \mathscr{A}^{\mathbb{N}_0}, \Lambda)$ with $\Lambda := \mathbb{P}(\boldsymbol{Y} \in \cdot)$ and $\boldsymbol{X} = (X_n)_{n\ge 0}$ denoting the identical mapping on this space, so that $\Lambda(\boldsymbol{X} \in \cdot) = \mathbb{P}(\boldsymbol{Y} \in \cdot)$, the stationarity of $\boldsymbol{Y}$ is equivalent to the property that the *shift* S on $\mathbb{Y}^{\mathbb{N}_0}$, defined by

$$S(y_0, y_1, \ldots) := (y_1, y_2, \ldots),$$

is measure-preserving. Therefore it appears to be natural to call $\boldsymbol{Y}$ ergodic if $\mathscr{S}$ is ergodic. Defining the invariant σ-field associated with $\boldsymbol{Y}$ by

$$\mathscr{I}_{\boldsymbol{Y}} := \left\{ B \in \mathscr{A}^{\mathbb{N}_0} : \mathbf{1}_B(\boldsymbol{Y}) = \mathbf{1}_B(S \circ \boldsymbol{Y}) \ \mathbb{P}\text{-a.s.} \right\},$$

we further see that $\mathscr{I}_{\boldsymbol{Y}} = \boldsymbol{Y}^{-1}(\mathscr{I}_S)$ and that ergodicity of $\boldsymbol{Y}$ holds iff $\mathscr{I}_{\boldsymbol{Y}}$ is $\mathbb{P}$-trivial.

By Kolmogorov's consistency theorem, any stationary sequence $\boldsymbol{Y} = (Y_n)_{n\ge 0}$ has a *doubly infinite extension* $\boldsymbol{Y}^* = (Y_n)_{n\in\mathbb{Z}}$ with distribution Γ^*, say, which in turn is associated with the measure-preserving shift map S^* on the doubly infinite product space $(\mathbb{Y}^{\mathbb{Z}}, \mathscr{A}^{\mathbb{Z}}, \Gamma^*)$, defined by

$$S^*(\ldots, y_{-1} y_0, y_1, \ldots) := (\ldots y_0, y_1, y_2, \ldots).$$

Plainly, S^* is invertible, and the inverse S^{*-1} is also measure-preserving. It should not take one by surprise that both transformations are further ergodic if this is true for S. The following lemma shows that sequences of iid random variables are ergodic.

Proposition A.1. *Any sequence $(Y_n)_{n\in\mathbb{T}}$ of iid random variables, where $\mathbb{T} = \mathbb{N}$ or $= \mathbb{Z}$, is ergodic.*

Proof. This follows from Kolmogorov's zero-one law if $\mathbb{T} = \mathbb{N}$ and extends to $\mathbb{T} = \mathbb{Z}$ by the above remark about the ergodicity of S^* and S^{*-1}. ◻

As an example of a non-ergodic stationary sequence one can take any stationary positive recurrent (and thus irreducible) periodic discrete Markov chain.

Let us finally introduce the ergodic theoretic notion of a *factor* which appeared in the proof of Prop. 3.18. Given two measure-preserving transformations T_1, T_2 of probability spaces $(\mathbb{Y}_1, \mathscr{A}_1, \mathbf{P}_1)$ and $(\mathbb{Y}_2, \mathscr{A}_2, \mathbf{P}_2)$, respectively, T_2 is called a *factor of* T_1 if there exists a measure-preserving map $\varphi : \mathbb{Y}_1 \to \mathbb{Y}_2$ (i.e. $\mathbf{P}_1(\varphi \in \cdot) = \mathbf{P}_2$) such that $\varphi \circ T_1 = T_2 \circ \varphi$ $\mathbf{P}_1$-a.s.

Proposition A.2. *If T_2 is a factor of an ergodic transformation T_1, then T_2 is ergodic as well.*

Proof. With the notation introduced above, choose any $A_2 \in \mathscr{I}_{T_2}$ and put $A_1 := \varphi^{-1}(A_2)$. Then $\varphi \circ T_1 = T_2 \circ \varphi$ $\mathbf{P}_1$-a.s. implies

$$T_1^{-1}(A_1) = (\varphi \circ T_1)^{-1}(A_2) = (T_2 \circ \varphi)^{-1}(A_2) = \varphi^{-1}(A_2) = A_1 \quad \mathbf{P}_1\text{-a.s.},$$

that is $A_1 \in \mathscr{I}_{T_1}$. Since T_1 is ergodic and $\mathbf{P}_1(\varphi \in \cdot) = \mathbf{P}_2$, we hence infer

$$\mathbf{P}_2(A_2) = \mathbf{P}_2(\varphi^{-1}(A_1)) = \mathbf{P}_1(A_1) \in \{0,1\},$$

which proves that T_2 is ergodic. □

A.2 Birkhoff's ergodic theorem

The following theorem due to BIRKHOFF [18] [☞ also [19] by the same author] is one of the fundamental results in ergodic theory and may also be viewed as the extension of the classical SLLN for sums of iid random variables to stationary sequences. We provide two versions of the result, the first one formulated in terms of a measure-preserving transformation as in [18], the second more probabilistic one in terms of a stationary sequence.

Theorem A.3. [Birkhoff's ergodic theorem for measure-preserving transformations] *Let T be a measure-preserving transformation of a probability space $(\mathbb{Y}, \mathscr{A}, \mathbf{P})$ and $g : \mathbb{Y} \to \mathbb{R}$ be a $\mathbf{P}$-integrable function, i.e. $g \in L^1(\mathbf{P})$. Then*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} g \circ T^k = \mathbf{E}(g|\mathscr{I}_T) \quad \mathbf{P}\text{-a.s. and in } L^1(\mathbf{P}), \tag{A.3}$$

and the a.s. convergence ramins valid if g is quasi-$\mathbf{P}$-integrable. As a particular consequence,

$$\lim_{n\to\infty} h_{T,n}(\cdot, A) = \mathbf{P}(A|\mathscr{I}_T) \quad \mathbf{P}\text{-a.s.} \tag{A.4}$$

for any $A \in \mathscr{A}$.

Clearly, the conditional expectations in (A.3) and (A.4) reduce to unconditional ones if T is ergodic and thus $\mathscr{I}_T$ $\mathbf{P}$-trivial.

Theorem A.4. [Birkhoff's ergodic theorem for stationary sequences] *Let $\boldsymbol{Y} = (Y_n)_{n\ge 0}$ be a stationary sequence of $\mathbb{Y}$-valued random variables on a probability space $(\Omega,\mathfrak{A},\mathbb{P})$ and $g:\mathbb{Y}\to\mathbb{R}$ be such that $g\circ Y_0$ is integrable. Then*

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n} g\circ Y_k = \mathbb{E}(g\circ Y_0|\mathscr{I}_{\boldsymbol{Y}})\quad \mathbb{P}\text{-a.s. and in } L^1(\mathbb{P}), \tag{A.5}$$

*and the a.s. convergence remains valid if g is quasi-**P**-integrable. As a particular consequence,*

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n} \mathbf{1}_A(Y_k) = \mathbb{P}(Y_0\in A|\mathscr{I}_{\boldsymbol{Y}})\quad \mathbb{P}\text{-a.s.} \tag{A.6}$$

for any $A\in\mathscr{A}$.

Excellent introductions to the theory of stationary sequences from a probabilist's viewpoint, including a proof of Theorem A.4, may be found in the textbooks by BREIMAN [22] and DURRETT [39].

A.3 Kingman's subadditive ergodic theorem

A sequence of real numbers $(c_n)_{n\ge 1}$ is called *subadditive* if

$$c_{m+n}\le c_m+c_n$$

for all $m,n\in\mathbb{N}$. An old lemma by FEKETE [50] states that every such sequence converges, viz.

$$\lim_{n\to\infty} c_n \;=\; \inf_{n\ge 1}\frac{c_n}{n}\;\in\;\mathbb{R}\cup\{-\infty\}.$$

The subadditive ergodic theorem for triangular schemes $(X_{k,n})_{n\ge 1}^{0\le k<n}$ of real-valued random variables, first obtained by KINGMAN [71] and later improved by LIGGETT [75], builds upon this property together with a certain type of stationarity. Here we present the more general version by LIGGETT.

Theorem A.5. [Subadditive ergodic theorem] *Let $(X_{k,n})_{n\ge 1}^{0\le k<n}$ be a family of real-valued random variables which satisfies the following conditions:*

(SA-1) $X_{0,n}\le X_{0,m}+X_{m,n}$ *a.s. for all* $0\le m<n$.
(SA-2) $(X_{nk,(n+1)k})_{n\ge 1}$ *is a stationary sequence for each* $k\ge 1$.
(SA-3) *The distribution of* $(X_{m,m+n})_{n\ge 1}$ *does not depend on* $m\ge 0$.

(SA-4) $\mathbb{E}X_{0,1}^{+} < \infty$ *and* $\mu := \inf_{n\geq 1} n^{-1}\mathbb{E}X_{0,n} > -\infty$.

Then

(a) $\lim_{n\to\infty} n^{-1}\mathbb{E}X_{0,n} = \mu$.
(b) $n^{-1}X_{0,n}$ *converges a.s. and in* L^1 *to a random variable* X *with mean* μ.
(c) *If all stationary sequences in (SA-2) are ergodic, then* $X = \mu$ *a.s.*
(d) *If* $\mu = -\infty$ *in (SA-4), then* $n^{-1}X_{0,n} \to -\infty$ *a.s.*

We note that Kingman assumed also (SA-4), but instead of (SA-1)-(SA-3) the stronger conditions

(SA-5) $X_{k,n} \leq X_{k,m} + X_{m,n}$ a.s. for all $0 \leq k < m < n$.
(SA-6) The distribution of $(X_{m+k,n+k})_{0\leq m<n}$ does not depend on $k \geq 0$.

A proof of the result may be found in the original article [75] or in the textbook by DURRETT [39, Ch. 6], the latter also containing a good collection of interesting applications including the Furstenberg-Kesten theorem for products of random matrices [☞ Theorem 3.4]. The reader is asked in Problem A.7 to deduce Birkhoff's ergodic theorem A.4 from the result.

Problems

Problem A.6. Let T be a measure-preserving transformation of a probability space $(\mathbb{Y},\mathscr{A},\mathbf{P})$ with associated σ-field $\mathscr{I}_T$ of $\mathbf{P}$-a.s. T-invariant sets. Let further Y_0 be a random element in $\mathbb{Y}$ with $\mathscr{L}(Y_0) = \mathbf{P}$ and $(Y_n)_{n\geq 0}$ be the stationary sequence defined by $Y_n := T^n(Y_0)$ for $n \geq 1$. Suppose there exists $A \in \mathscr{I}_T$ with $0 < \mathbf{P}(A) < 1$. Prove that $(Y_n)_{n\geq 0}$ is stationary under both, $\mathbf{P}(\cdot|Y_0 \in A)$ and $\mathbf{P}(\cdot|Y_0 \in A^c)$.

Problem A.7. Give a proof of Birkhoff's ergodic theorem A.4 with the help of the subadditive ergodic theorem.

Appendix B
Convex function inequalities for martingales and their maxima

Let $(M_n)_{n\geq 0}$ be a martingale with natural filtration $(\mathscr{F}_n)_{n\geq 0}$ and increments $D_n = M_n - M_{n-1}$ for $n \geq 1$. In the following, we collect (without proofs) some powerful martingale inequalities that provide bounds for the ϕ-*moments* $\mathbb{E}\phi(M_n)$, when $\phi : \mathbb{R} \to \mathbb{R}_\geq$ denotes an even convex function with $\phi(0) = 0$ and some additional properties. This includes the standard class $\phi(x) = |x|^p$ for $p \geq 1$. Setting $M_\infty := \liminf_{n\to\infty} M_n$, all provided upper bounds remain valid for $n = \infty$ when observing that Fatou's lemma implies

$$\mathbb{E}\phi(M_\infty) \ \leq \ \liminf_{n\to\infty} \mathbb{E}\phi(M_n).$$

We begin with the class of ϕ that have a concave derivative in $\mathbb{R}_>$ which encompasses $\phi(x) = |x|^p$ for $1 \leq p \leq 2$. The subsequent result is cited from [5] and an improvement (with regard to the appearing constant) of a version due to TOPCHIĬ & VATUTIN [110].

Theorem B.1. [Topchiĭ-Vatutin inequality] *Let $\phi : \mathbb{R} \to \mathbb{R}_\geq$ be an even convex function with concave derivative on $\mathbb{R}_>$ and $\phi(0) = 0$. Then*

$$\mathbb{E}\phi(M_n) - \mathbb{E}\phi(M_0) \ \leq \ c \sum_{k=1}^{n} \mathbb{E}\phi(D_k), \tag{B.1}$$

for all $n \in \overline{\mathbb{N}}_0$ and $c = 2$. The constant may be chosen as $c = 1$ if $(M_n)_{n\geq 0}$ is nonnegative or has a.s. symmetric conditional increment distributions, and the same holds generally true, if $\phi(x) = |x|$ or $\phi(x) = x^2$, in the last case even with equality sign in (B.1).

We continue with two famous convex function inequalities by BURKHOLDER, DAVIS & GUNDY [24] which are valid for a much larger class of convex functions ϕ. For proofs the reader may consult the textbooks [28, Thms. 11.3.1 & 11.3.2], [62, Section 2.4], or the afore-mentioned original work.

Theorem B.2. [Burkholder-Davis-Gundy inequalities] *Let $\phi:\mathbb{R}\to\mathbb{R}_\ge$ be an even convex function satisfying $\phi(0)=0$ and $\phi(2t)\le\gamma\phi(t)$ for all $t\ge 0$ and some $\gamma>0$. Put $E_n(\phi):=\mathbb{E}(\max_{0\le k\le n}\phi(M_k))$. Then*

$$a_\gamma\mathbb{E}\phi\left(\left(\sum_{k=1}^n D_k^2\right)^{1/2}\right)\ \le\ E(\phi)\ \le\ b_\gamma\mathbb{E}\phi\left(\left(\sum_{k=1}^n D_k^2\right)^{1/2}\right) \tag{B.2}$$

and

$$E_n(\phi)\ \le\ c_\gamma\left[\mathbb{E}\phi\left(\left(\sum_{k=1}^n\mathbb{E}\big(D_k^2|\mathscr{F}_{k-1}\big)\right)^{1/2}\right)+\mathbb{E}\left(\max_{0\le k\le n}\phi(D_k)\right)\right] \tag{B.3}$$

for all $n\in\overline{\mathbb{N}}_0$ and constants $a_\gamma,b_\gamma,c_\gamma\in\mathbb{R}_>$ depending only on γ. The last inequality actually remains valid if, ceteris paribus, ϕ is merely nondecreasing instead of convex on $\mathbb{R}_\ge$.

Of special importance for our purposes is the case when M_n is a weighted sum of iid zero-mean random variables and $\phi(x)=|x|^p$ for some $p>0$. We therefore note:

Corollary B.3. *If $\phi(x)=|x|^p$ (thus $\gamma=2^p$) for some $p>0$ and $M_n=\sum_{k=1}^n t_kX_k$ for $t_1,t_2,...\in\mathbb{R}$ and iid $X_1,X_2,...\in L_0^p$, then* (B.3) *takes the form*

$$E_n(\phi)\ \le\ c_p\left[\|X_1\|_2^p\left(\sum_{k=1}^n t_k^2\right)^{p/2}+\mathbb{E}\left(\max_{1\le k\le n}|t_kX_k|^p\right)\right], \tag{B.4}$$

for all $n\in\overline{\mathbb{N}}_0$ and a constant c_p only depending on p, giving in particular

$$\mathbb{E}|M_n|^p\ \le\ c_p\left[\|X_1\|_2^p\left(\sum_{k=1}^n t_k^2\right)^{p/2}+\|X_1\|_p^p\sum_{k=1}^n|t_k|^p\right]. \tag{B.5}$$

We close this section with a statement of the classical L^p-inequality by BURKHOLDER, a proof of which may again be found in [28, Thm. 11.2.1], [62, Thm. 2.10], or in the original work [23]. It is important to note that it holds for $p>1$ only. The case $p=1$ is different but will not be considered here.

Theorem B.4. [Burkholder inequality] *Let $p > 1$. Then*

$$a_p \left\| \left(\sum_{k=1}^{n} D_k^2 \right)^{1/2} \right\|_p^p \leq \|M_n\|_p \leq b_p \left\| \left(\sum_{k=1}^{n} D_k^2 \right)^{1/2} \right\|_p^p \tag{B.6}$$

for $n \in \overline{\mathbb{N}}_0$ and constants $a_p, B_p \in \mathbb{R}_>$ only depending on p.

Appendix C
Banach's fixed point theorem

We consider an arbitrary continuous self-map $f:\mathbb{X}\to\mathbb{X}$ of a metric space $(\mathbb{X},\rho)$ and denote by $f^n = f\circ\ldots\circ f$ (n-times) its n-fold composition for $n\ge 1$. If there exists an initial value $x_0\in\mathbb{X}$ such that the sequence $x_n := f(x_{n-1}) = f^n(x_0)$, $n\ge 1$, converges to some $x_\infty\in\mathbb{X}$, then the continuity of f implies that x_∞ is a fixed point of f, for

$$x_\infty \;=\; \lim_{n\to\infty} x_n \;=\; f\left(\lim_{n\to\infty} x_{n-1}\right) \;=\; f(x_\infty).$$

Generally, one cannot say anything about the existence and number of fixed points of a continuous map f, but there are situations where it has a unique fixed point ξ and every iteration sequence $(f^n(x))_{n\ge 1}$, $x\in\mathbb{X}$, converges to ξ. An assertion of this kind is the content of *Banach's fixed-point theorem* for a special class of maps that will be defined first.

Definition C.1. A self-map $f:\mathbb{X}\to\mathbb{X}$ of a metric space $(\mathbb{X},\rho)$ is called *contraction* or more specifically *α-contraction* if there exists $\alpha\in[0,1)$ such that

$$\rho(f(x),f(y)) \le \alpha\,\rho(x,y) \tag{C.1}$$

for all $x,y\in\mathbb{X}$. If (C.1) holds true when replacing f with f^n for some $n\ge 2$, then f is called *quasi-contraction* or *α-quasi-contraction*.

Under a contraction, the distance between two iteration sequences $(f^n(x))_{n\ge 1}$ and $(f^n(y))_{n\ge 1}$ is therefore decreasing geometrically fast, viz.

$$\rho(f^n(x),f^n(y)) \;\le\; \alpha^n\,\rho(x,y) \tag{C.2}$$

for all $n\ge 1$. The following theorem shows that this entails convergence to a unique fixed point of f if the space $(\mathbb{X},\rho)$ is complete. Notice that the contraction property trivially implies continuity.

Theorem C.2. [Banach's fixed-point theorem] *Every contraction $f:\mathbb{X}\to\mathbb{X}$ on a complete metric space $(\mathbb{X},\rho)$ possesses a unique fixed point $\xi\in\mathbb{X}$.*

The example $f(x)=x/2$ on $\mathbb{X}=(0,1]$ shows that one cannot dispense with the completeness of $\mathbb{X}$.

Proof. We first prove existence of a fixed point ξ. Pick an arbitrary $x_0\in\mathbb{X}$ and put $x_n=f(x_{n-1})$ for $n\geq 1$. Then the contraction property implies

$$\rho(x_{k+1},x_k) \;=\; \rho(f(x_k),f(x_{k-1})) \;\leq\; \alpha\,\rho(x_k,x_{k-1})$$

and therefore upon k-fold iteration

$$\rho(x_{k+1},x_k) \;\leq\; \alpha^k\rho(x_1,x_0)$$

for some $\alpha\in[0,1)$ and all $k\geq 1$. As a consequence,

$$\begin{aligned}\rho(x_{m+n},x_m) \;&\leq\; \sum_{k=m}^{m+n-1}\rho(x_{k+1},x_k)\\ &\leq\; \rho(x_1,x_0)\sum_{k=m}^{m+n-1}\alpha^k \;\leq\; \frac{\alpha^m}{1-\alpha}\,\rho(x_1,x_0)\end{aligned} \tag{C.3}$$

for all $m,n\geq 1$, that is, $(x_n)_{n\geq 0}$ is a Cauchy sequence in $\mathbb{X}$ and hence, by completeness, convergent to some $\xi\in\mathbb{X}$ which is also a fixed point of f because f is continuous.

Turning to uniqueness of ξ, suppose that ζ is a second fixed point of f. By another use of the contraction property, we then infer

$$\rho(\xi,\zeta) \;=\; \rho(f(\xi),f(\zeta)) \;\leq\; \alpha\,\rho(\xi,\zeta)$$

and thus $\rho(\xi,\zeta)=0$. □

By combining (C.3) with the fact that $\rho:\mathbb{X}\times\mathbb{X}\to\mathbb{R}_{\geq}$ is continuous in each variable, we easily conclude the geometric convergence of any iteration sequence to the fixed point ξ.

Corollary C.3. *In the situation of Theorem C.2, furthermore*

$$\rho(\xi,f^n(x)) \;\leq\; \frac{\alpha^n}{1-\alpha}\,\rho(f(x),x)$$

holds true for all $x\in\mathbb{X}$ and $n\geq 1$, where α denotes the contraction parameter of f.

Proof. Obvious in view of $\lim_{n\to\infty}\rho(x_{m+n},x_m)=\rho(\xi,x_m)$ and (C.3). □

The next result shows that Banach's fixed point theorem essentially remains valid for quasi-contractions.

Theorem C.4. [Banach's fixed-point theorem for quasi-contractions] *Every quasi-contraction $f:\mathbb{X}\to\mathbb{X}$ on a complete metric space $(\mathbb{X},\rho)$ possesses a unique fixed point $\xi\in\mathbb{X}$, and*

$$\rho(\xi,f^n(x)) \;\le\; \frac{\alpha^n}{1-\alpha}\max_{0\le r<m}\rho(f^{m+r}(x),f^m(x)) \tag{C.4}$$

for some $m\ge 1$, $\alpha\in[0,1)$ and all $x\in\mathbb{X}$, $n\ge 1$.

Proof. Pick m,α such that f^m forms an α-contraction on $(\mathbb{X},\rho)$ with unique fixed point ξ. Writing $n\in\mathbb{N}$ in the form $km+r$ with unique $k\in\mathbb{N}_0$ and $r\in\{0,\dots,m-1\}$, we infer with the help of Corollary C.3

$$\rho(\xi,f^n(x)) \;\le\; \max_{0\le j<m}\rho(\xi,f^{km+j}(x)) \;\le\; \frac{\alpha}{1-\alpha}\max_{0\le j<m}\rho(f^{m+j}(x),f^j(x))$$

and thus (C.4), in particular $\rho(\xi,f^n(x))\to 0$. Since f is continuous, the latter implies that ξ is also the (necessarily unique) fixed point of f. □

The proof of Banach's fixed point theorem C.2 shows even more: Replacing the global by a local contraction property along an iteration sequence, existence of a fixed point still follows, but it needs no longer be unique.

Theorem C.5. *Let $(\mathbb{X},\rho)$ be a complete metric space and $f:\mathbb{X}\to\mathbb{X}$ an arbitrary self-map. Suppose there exist $x_0\in\mathbb{X}$ and constants $c\ge 0$ and $\alpha\in[0,1)$ such that*

$$\rho(f^{n+1}(x_0),f^n(x_0)) \;\le\; c\alpha^n \tag{C.5}$$

for all $n\ge 1$. Then $\xi=\lim_{n\to\infty}f^n(x_0)$ exists and it is a fixed point of f if the map is continuous. Moreover,

$$\rho(\xi,f^n(x_0)) \;\le\; \frac{c\alpha^n}{1-\alpha} \tag{C.6}$$

for all $n\ge 1$.

Proof. Putting once again $x_n:=f^n(x_0)$ and using (C.5), we obtain in (C.3)

$$\rho(x_{m+n},x_m) \;\le\; \frac{c\alpha^m}{1-\alpha}$$

for all $m, n \geq 0$ and therefore that $(x_n)_{n\geq 1}$ is again a Cauchy sequence in $\mathbb{X}$. All further necessary arguments to complete the proof follow in a similar manner as given for the previous results. $\square$

Appendix D
Hausdorff measures and dimension

The following short introduction of *Hausdorff measures* and *Hausdorff dimension* is based on the more detailed expositions of this subject by ELSTRODT [43], EDGAR [41] and FALCONER [48].

From measure theory, we recall that a set function $\mu : \mathfrak{P}(\mathbb{X}) \to [0,\infty]$ for an arbitrary set $\mathbb{X} \neq \emptyset$ is called *outer measure* if $\mu(\emptyset) = 0$ and μ is *subadditive*, that is

$$\mu\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} \mu(A_n)$$

for any sequence $A_1, A_2, \ldots$ of subsets of $\mathbb{X}$. Following CARATHÉODORY, a set $A \subset \mathbb{X}$ is called *μ-measurable* if

$$\mu(C) \geq \mu(C \cap A) + \mu(C \cap A^c)$$

for all $C \subset \mathbb{X}$ which, by the subadditivity of μ, is actually equivalent to

$$\mu(C) = \mu(C \cap A) + \mu(C \cap A^c)$$

for all $C \subset \mathbb{X}$. One of the fundamental results in measure theory [☞ e.g. [43, Satz II.4.4]] states that the system $\mathfrak{A}_\mu$ of all μ-measurable sets forms a σ-field and that μ is a measure on $(\mathbb{X}, \mathfrak{A}_\mu)$.

Now let $(\mathbb{X}, \rho)$ be a metric space with Borel σ-field $\mathscr{B}(\mathbb{X})$. For $x \in \mathbb{X}$ and $A, B \subset \mathbb{X}$, we put

$$\begin{aligned} \rho(A,B) &:= \inf\{\rho(x,y) : x \in A,\, y \in B\}, \\ \rho(x,B) &:= \inf\{\rho(x,y) : y \in B\} = \rho(\{x\},B), \\ \mathit{diam}(A) &:= \sup\{\rho(x,y) : x,y \in A\}, \end{aligned}$$

and call $\rho(A,B)$ the *distance between A and B*, $\rho(x,B)$ the *distance between x and B*, and $diam(A)$ the *diameter of A*, where $diam(\emptyset) := 0$. The notion defined next goes again back to CARATHÉODORY.

Definition D.1. An outer measure $\mu : \mathfrak{P}(\mathbb{X}) \to [0,\infty]$ is called *metric outer measure* if

$$\mu(A \cup B) \;=\; \mu(A) + \mu(B)$$

for all $A, B \subset \mathbb{X}$ with positive distance $\rho(A,B)$.

Bound for the definition of Hausdorff measures, the next theorem and the subsequent construction scheme are of importance.

Theorem D.2. *Given an outer measure $\mu : \mathfrak{P}(\mathbb{X}) \to [0,\infty]$ on a metric space $(\mathbb{X},\rho)$, the Borel σ-field $\mathscr{B}(\mathbb{X})$ is contained in $\mathfrak{A}_\mu$ iff μ is a metric outer measure.*

Proof. ☞ [43, Satz II.9.3]. □

A general construction scheme. Here is a general procedure for obtaining metric outer measures: Let $\mathfrak{C} \subset \mathfrak{P}(\mathbb{X})$ an arbitrary family of subsets of $\mathbb{X}$ containing the empty set $\varnothing$, and let $\xi : \mathfrak{C} \to [0,\infty]$ be a set function with $\xi(\emptyset) = 0$. For $A \subset \mathbb{X}$ and $\varepsilon > 0$, we put

$$\eta_\varepsilon(A) \;:=\; \inf\left\{\sum_{n\ge 1} \xi(A_n) : A_n \in \mathfrak{C},\, diam(A_n) \le \varepsilon \text{ for all } n \ge 1,\, A \subset \bigcup_{n\ge 1} A_n\right\}$$

with the usual convention $\inf\emptyset := \infty$. As one can easily verify [☞ Problem D.13], η_ε defines an outer measure, and the mapping $\varepsilon \mapsto \eta_\varepsilon(A)$ is nonincreasing for each $A \subset \mathbb{X}$. Further defining

$$\eta(A) \;:=\; \sup_{\varepsilon>0} \eta_\varepsilon(A) \;=\; \lim_{\varepsilon\downarrow 0} \eta_\varepsilon(A), \quad A \subset \mathbb{X},$$

we find, for arbitrary $A_1, A_2, \ldots \subset \mathbb{X}$ and $\varepsilon > 0$, that

$$\eta_\varepsilon\left(\bigcup_{n\ge 1} A_n\right) \;\le\; \sum_{n\ge 1} \eta_\varepsilon(A_n) \;\le\; \sum_{n\ge 1} \eta(A_n).$$

and therefore $\eta(\bigcup_{n\ge 1} A_n) \le \sum_{n\ge 1} \eta(A_n)$. Hence η is also an outer measure, and the subsequent argument shows that it is even a metric outer measure. Pick any $A, B \subset \mathbb{X}$ with $\rho(A,B) > 0$. Clearly, it remains to argue that $\eta(A \cup B) \ge \eta(A) + \eta(B)$ for which we may assume $\eta(A \cup B) < \infty$. For any $0 < \varepsilon < \rho(A,B)$, the latter implies the existence of $C_n \in \mathfrak{C}$ with $\xi(C_n) \le \varepsilon$ for $n \ge 1$ and $A \cup B \subset \bigcup_{n\ge1} C_n$. But then no C_n can have common elements with both, A and B, so that $(C_n)_{n\ge1}$ can be decomposed into a covering of A and a covering of B, giving

$$\sum_{n\ge1} \xi(C_n) \;\ge\; \eta_\varepsilon(A) + \eta_\varepsilon(B).$$

This being actually true for any ε-covering of $A \cup B$, $\eta_\varepsilon(A \cup B) \ge \eta_\varepsilon(A) + \eta_\varepsilon(B)$ follows and thereupon $\eta(A \cup B) \ge \eta(A) + \eta(B)$ as desired when letting ε tend to 0. Note that η_ε needs not be a metric outer measure.

Example D.3. Consider the real line $\mathbb{X}$ with the usual Euclidean distance $\rho(x,y) := |x-y|$, let $\mathfrak{C}$ be the system of left open intervals $(a,b]$ and $\xi(A) := \mathit{diam}(A)$. Since each $(a,b]$ can be decomposed into pairwise disjoint intervals of the same type (thus elements of $\mathfrak{C}$) of length at most ε for arbitrary $\varepsilon > 0$, we see that $\eta = \eta_\varepsilon$ does not depend on ε and equals Lebesgue outer measure $\lambda\!\!\!\lambda^*$ on $\mathbb{R}$. Theorem D.2 above now implies the known result that all Borel subsets of $\mathbb{R}$ are Lebesgue measurable, i.e., measurable with respect to $\lambda\!\!\!\lambda^*$.

The previous example is about a special and particularly important Hausdorff measure, and we now proceed with its general definition. Again, for an arbitrary metric space $(\mathbb{X},\rho)$, let $\mathfrak{C}$ be the system of all bounded subsets A of $\mathbb{X}$, i.e. $\mathit{diam}(A) < \infty$, and $\xi(A) := \mathit{diam}(A)^\alpha$ for nonempty $A \in \mathfrak{C}$ and a fixed $\alpha \ge 0$, where $0^0 := 1$. The above construction scheme then provides us with the outer measures

$$\mathscr{H}_{\alpha,\varepsilon}(A) \;:=\; \inf\left\{\sum_{n\ge1} \mathit{diam}(A)^\alpha : A \subset \bigcup_{n\ge1} A_n,\ \mathit{diam}(A_n) \le \varepsilon \text{ f.a. } n \ge 1\right\} \tag{D.1}$$

for $A \subset \mathbb{X}$, and the metric outer measure

$$\mathscr{H}_\alpha(A) \;:=\; \sup_{\varepsilon>0} \mathscr{H}_{\alpha,\varepsilon}(A), \quad A \subset \mathbb{X}, \tag{D.2}$$

which, by Theorem D.2, is a measure on $\mathscr{B}(\mathbb{X})$. It does not take much to see that $\mathscr{H}_\alpha(A)$ remains unchanged when allowing only coverings of open (closed) A_n in the definition of the $\mathscr{H}_{\alpha,\varepsilon}$. In particular, it is enough to consider measurable coverings in (D.1).

Definition D.4. The (outer) measure $\mathscr{H}_\alpha$ given by (D.2) is called *α-dimensional (outer) Hausdorff measure*.

The following result identifies $\mathscr{H}_\alpha$ in the case $\alpha = 0$.

Proposition D.5. *The zero-dimensional Hausdorff measure $\mathscr{H}_0$ is the counting measure on $\mathbb{X}$.*

Proof. If $A = \{a_1, ..., a_n\}$ consists of n elements, then $A \subset \bigcup_{i=1}^n \mathbb{B}(a_i, \varepsilon)$ for all $\varepsilon > 0$, where $\mathbb{B}(x, \varepsilon) := \{y \in \mathbb{X} : \rho(x,y) \leq \varepsilon\}$ denotes the closed ε-ball with center x. It follows that

$$\mathscr{H}_{0,\varepsilon}(A) \ \leq \ \sum_{i=1}^n diam(\mathbb{B}(a_i, \varepsilon))^0 \ = \ n.$$

But the inequality obviously turns into an identity as soon as 2ε becomes smaller than $\min_{1 \leq i \neq j \leq n} \rho(a_i, a_j)$. Hence $\mathscr{H}_0(A) = \lim_{\varepsilon \downarrow 0} \mathscr{H}_{0,\varepsilon}(A) = n$. Finally, if A is an infinite set, then it contains a subset A_n with n elements for each $n \geq 1$. As a consequence, $\mathscr{H}_0(A) \geq \sup_{n \geq 1} \mathscr{H}_0(A_n) = \infty$. □

As a function of α, $\mathscr{H}_\alpha$ is easily seen to be nonincreasing, but the next result provides a much stronger assertion which, in particular, leads to the notion of Hausdorff dimension.

Proposition D.6. *For any $A \subset \mathbb{X}$ and $0 < \alpha < \beta$,*

$$\mathscr{H}_\alpha(A) < \infty \quad \Rightarrow \quad \mathscr{H}_\beta(A) = 0,$$
$$\mathscr{H}_\beta(A) > 0 \quad \Rightarrow \quad \mathscr{H}_\alpha(A) = \infty.$$

Consequently, there exists a unique number, denoted as $\dim_H A$ and called ***Hausdorff dimension of*** *A, such that*

$$\mathscr{H}_\alpha(A) \ = \ \begin{cases} \infty, & \text{if } \alpha < \dim_H A, \\ 0, & \text{if } \alpha > \dim_H A. \end{cases}$$

Proof. For each $B \subset \mathbb{X}$ with diameter $diam(B) \leq \varepsilon$, we have

$$diam(B)^\beta \ \leq \ \varepsilon^{\beta-\alpha}\, diam(B)^\alpha.$$

It follows that $\mathscr{H}_{\beta,\varepsilon}(A) \leq \varepsilon^{\beta-\alpha}\,\mathscr{H}_{\alpha,\varepsilon}(A)$ and therefore in the case $\mathscr{H}_\alpha(A) < \infty$ that

$$\mathscr{H}_\beta(A) \ = \ \lim_{\varepsilon \downarrow 0} \mathscr{H}_{\beta,\varepsilon}(A) \ \leq \ \lim_{\varepsilon \downarrow 0} \varepsilon^{\beta-\alpha}\,\mathscr{H}_{\alpha,\varepsilon}(A) \ = \ 0.$$

If $\mathscr{H}_\beta(A) > 0$, then

$$\mathscr{H}_\alpha(A) \ = \ \lim_{\varepsilon \downarrow 0} \mathscr{H}_{\alpha,\varepsilon}(A) \ \geq \ \lim_{\varepsilon \downarrow 0} \varepsilon^{\alpha-\beta}\,\mathscr{H}_{\beta,\varepsilon}(A) \ = \ \infty,$$

which completes the proof. □

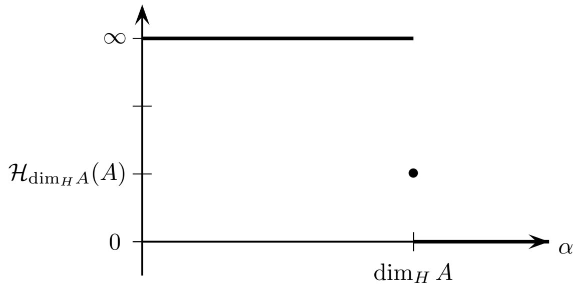


Fig. D.1 The function $\alpha \mapsto \mathscr{H}_\alpha(A)$. Its value at $\alpha = \dim_H A$ (here positive and finite) can be anything in $[0,\infty]$.

In view of the previous result the reader might expect the value of $\mathscr{H}_\alpha(A)$ to be positive and finite if $\alpha = \dim_H A$. Unfortunately, the true answer is that this value can be anything between and including 0 and ∞, although for many prominent examples like the Cantor ternary set the value is indeed in $\mathbb{R}_>$.

Hausdorff measures and Hausdorff dimension play an important role in the study of so-called *fractal sets* or *fractals* which are thus called because their dimension is not an integer. Again, the Cantor ternary set provides a well-known example (of dimension $\log 2/\log 3$ as will be shown in Example D.12) of a fractal, but there are many others including those random Cantor sets defined and studied in Section 5.2. In the case of a one-dimensional curve, a two-dimensional surface or a three-dimensional body one can measure arc length, surface area or volume, respectively. *Fractal measures* like the $\mathscr{H}_\alpha$ allow the very same in a space of fractal dimension. A more detailed exposition of this subject can be found in [41, 42] or [48].

That Hausdorff measures generalize the notion of ordinary volume measure in spaces of integral dimension is further sustained by the next result. A bijective map $\varphi : \mathbb{X} \to \mathbb{X}$ satisfying $\rho(\varphi(x), \varphi(y)) = \rho(x,y)$ for all $x, y \in \mathbb{X}$ is called a *motion (in* $\mathbb{X}$*)*, and the set of all motions in $\mathbb{X}$ forms a group which contains the translations as a subgroup if $\mathbb{X}$ is itself a group. Since the diameter of a set is obviously invariant under motions, we have:

Proposition D.7. *Any Hausdorff measure $\mathscr{H}_\alpha$ on a metric space $(\mathbb{X}, \rho)$ is motion invariant, that is $\mathscr{H}_\alpha(\varphi(A)) = \mathscr{H}_\alpha(A)$ for all $A \subset \mathbb{X}$ and motions φ.*

Specializing to $\mathbb{X} = \mathbb{R}^d$ with the Euclidean metric, any motion is the composition of a translation and an orthogonal map and Lebesgue measure $\lambda\!\!\lambda^d$ the unique motion invariant measure with $\lambda\!\!\lambda^d([0,1]^d) = 1$ [☞ [12, Thms. 8.1 and 8.3]]. One can easily verify that $0 < \mathscr{H}_d([0,1]^d) < \infty$ and hence $\dim_H [0,1]^d = d$ by Proposition D.6 holds true. Consequently, $\mathscr{H}_d = \kappa_d \lambda\!\!\lambda^d$ for some positive κ_d. It can further be shown that $\mathscr{H}_k$ for $k \in \{1, ..., d-1\}$ and up to a positive scalar equals k-dimensional volume measure (length if $k = 1$, and surface area if $k = 2$) on k-dimensional manifolds. The behavior of $\mathscr{H}_\alpha$ under dilations $\psi_t : (x_1,...,x_d) \mapsto (tx_1,...,tx_d)$, $t \in \mathbb{R}\backslash\{0\}$, is stated in the next result and again immediate when observing that $diam(tA) = |t|\, diam(A)$ for any real t.

Proposition D.8. *If $\mathbb{X} = \mathbb{R}^d$ with Euclidean metric $\rho(x,y) = |x-y|$, then*

$$\mathscr{H}_\alpha(\psi_t(A)) \;=\; \mathscr{H}_\alpha(tA) \;=\; |t|^\alpha\, \mathscr{H}_\alpha(A) \tag{D.3}$$

for all $\alpha \ge 0$, $t \in \mathbb{R}\backslash\{0\}$ and $A \subset \mathbb{R}^d$.

Given a second metric space $(\mathbb{Y},\chi)$, a function $f : (\mathbb{X},\rho) \to (\mathbb{Y},\chi)$ is called *Hölder continuous of order* $\beta > 0$ if

$$\chi(f(x), f(y)) \;\le\; \kappa\, \rho(x,y)^\beta$$

for all $x,y \in \mathbb{X}$ and a constant $\kappa \in \mathbb{R}_>$. If $\beta = 1$, f is also called *Lipschitz continuous*. In the following, we will use the same symbol $\mathscr{H}_\alpha$ for the α-dimensional Hausdorff measure on $(\mathbb{X},\rho)$ and on $(\mathbb{Y},\chi)$.

Proposition D.9. *Let $f : (\mathbb{X},\rho) \to (\mathbb{Y},\chi)$ be a Hölder continuous function of order β and $A \subset \mathbb{X}$. Then $\mathscr{H}_{\alpha/\beta}(f(A)) \le \kappa^{\alpha/\beta} \mathscr{H}_\alpha(A)$ and $\dim_H f(A) \le \beta^{-1} \dim_H A$. If f is injective on A with both f and $f^{-1} : (f(A),\chi) \to (A,\rho)$ being Lipschitz continuous, then $\dim_H f(A) = \dim_H A$.*

Proof. Let $\varepsilon > 0$ be arbitrary and $(A_n)_{n\ge1}$ an ε-covering of A. Since

$$diam(f(A \cap A_n)) \le \kappa\, diam(A_n)^\beta \le \delta(\varepsilon) := \kappa\varepsilon^\beta,$$

we see that $(f(A \cap A_n))_{n\ge1}$ forms a $\delta(\varepsilon)$-covering of $f(A)$, and

$$\sum_{n\ge1} diam(f(A \cap A_n))^{\alpha/\beta} \;\le\; \kappa^{\alpha/\beta} \sum_{n\ge1} diam(A_n)^\alpha.$$

Hence, $\mathscr{H}_{\alpha/\beta,\delta(\varepsilon)}(f(A)) \le \kappa^{\alpha/\beta} \mathscr{H}_{\alpha,\varepsilon}(A)$. By letting ε tend to 0, the first two assertions follow. But the last assertion is now easily obtained, because $\dim_H A = \dim_H f^{-1}(f(A)) \le \dim_H f(A) \le \dim_H A$. □

As a direct consequence of this proposition, we note that the Hausdorff dimensions of a set under *equivalent metrics* are equal, where two metrics ρ_1, ρ_2 on $\mathbb{X}$ are called equivalent if

$$\kappa_1 \rho_1(x,y) \;\le\; \rho_2(x,y) \;\le\; \kappa_2 \rho_1(x,y)$$

for all $x, y \in \mathbb{X}$ and suitable $\kappa_1, \kappa_2 \in \mathbb{R}_>$. For example, on $\mathbb{R}^d$ the ℓ^p-metrics

$$\rho_p(x,y) \ := \ \begin{cases} \left(\sum_{i=1}^d |x_i - y_i|^p\right)^{1/p}, & \text{if } 1 \le p < \infty, \\ \max_{1\le i\le d} |x_i - y_i|, & \text{if } p = \infty \end{cases}$$

are all equivalent.

Generally, the Hausdorff dimension provides only very limited information about the topological structure of a set A. Only when $\dim_H A < 1$, the distribution of the elements of A is so coarse that no two of them belong to the same connected component of A, in which case the set is called *totally disconnected*.

Proposition D.10. *Any subset A of $\mathbb{R}^d$ with $\dim_H A < 1$ is totally disconnected.*

Proof. Pick two distinct elements x, y of A and define $f : (\mathbb{R}^d, |\cdot|) \to (\mathbb{R}, |\cdot|)$ by $f(z) := |x - z|$. Then f is Lipschitz continuous, for

$$|f(z) - f(w)| \ = \ \Big||x-z| - |x-w|\Big| \ \le \ |z-w|$$

for all $z, w \in \mathbb{R}^d$. Consequently, $\dim_H f(A) \le \dim_H A < 1$ by Proposition D.9 which in turn implies $\lambda\!\!\lambda(f(A)) = \mathscr{H}_1(f(A)) = 0$ and thus that $f(A)^c$ forms a dense subset of $\mathbb{R}$. By now picking some $r \in f(A)^c \cap (0, f(y))$, it follows that

$$A \ = \ \{z \in A : |x - z| < r\} \cup \{z \in A : |x - z| > r\},$$

that is, A is contained in two disjoint open sets with x being an element of the first and y being an element of the second set. In particular, x and y belong to distinct connected components of A. ☐

Calculating Hausdorff dimensions. We close this excursion into geometric measure theory with a quick look at the question of how to calculate Hausdorff dimensions in applications. Working directly with the definition is typically difficult if not impossible. In this regard, FALCONER [48, p. 54] writes

> Rigorous dimension calculations often involve pages of complicated manipulations and estimates that provide little intuitive enlightenment.

While an upper bound for $\dim_H A$ is often obtained in a relatively easy manner by calculation of $\sum_{n\ge1} diam(A_n)^\alpha$ for *special* coverings $(A_n)_{n\ge1}$ of A, a lower bound usually requires the estimation of $\sum_{n\ge1} diam(A_n)^\alpha$ for *all* ε-coverings of A which, in view of the generally enormous number of such coverings, raises the question for an alternative approach. Such an alternative does indeed exist in situations where the elements of A are sufficiently dispersed so that no small set, in terms of its diameter, can cover too much of A. The following *mass distribution principle* makes this idea precise: find a mass distribution on A, namely a measure Λ on A with $0<\Lambda(A)<\infty$, such that the mass $\Lambda(B)$ covered by B is bounded with respect to $\mathit{diam}(B)^\alpha$ for any sufficiently small set B. The following result is again taken from [48, p. 55].

Proposition D.11. *Let $A\in\mathscr{B}(\mathbb{X})$ and suppose there exists a measure Λ on $(\mathbb{X},\mathscr{B}(\mathbb{X}))$ and $c,\alpha,\varepsilon>0$ such that*

$$\Lambda(B)\ \le\ c\,\mathit{diam}(B)^\alpha$$

for all $B\in\mathscr{B}(\mathbb{X})$ with $\mathit{diam}(B)\le\varepsilon$. Then $\mathscr{H}_\alpha(A)\ge\Lambda(A)/c$.

Proof. Given an arbitrary, w.l.o.g. measurable [☞ remark after (D.2)] ε-covering $(A_n)_{n\ge1}$ of A, we infer

$$\Lambda(A)\ \le\ \Lambda\left(\bigcup_{n\ge1}A_n\right)\ \le\ \sum_{n\ge1}\Lambda(A_n)\ \le\ c\sum_{n\ge1}\mathit{diam}(A_n)^\alpha$$

and hence $\mathscr{H}_\alpha(A)\ge\mathscr{H}_{\alpha,\varepsilon}(A)\ge\Lambda(A)/c$ □

Example D.12. We illustrate the procedure by another look at the Cantor ternary set $\mathfrak{C}$ which may be defined as follows: Let $\mathbb{T}_2=\bigcup_{n\ge0}\{1,2\}^n$ be the homogeneous tree of order 2 in Ulam-Harris labeling, define [☞ also Section 5.2]

$$\begin{aligned}&J_\varnothing:=[0,1],\\ &J_1:=[0,1/3],\quad J_2:=[2/3,1],\\ &J_{11}:=[0,1/9],\quad J_{12}:=[2/9,1/3],\quad J_{21}:=[2/3,7/9],\quad J_{22}:=[8/9,1],\\ &\quad\vdots\end{aligned}$$

and then

$$\mathfrak{C}\ :=\ \bigcap_{n\ge1}\ \bigcup_{\mathsf{v}\in\{1,2\}^n}J_\mathsf{v}.$$

For each $n\in\mathbb{N}$, the family $(J_\mathsf{v})_{\mathsf{v}\in\{1,2\}^n}$ consists of 2^n intervals of length (= diameter) 3^{-n} and thus forms an ε-covering for $\varepsilon=3^{-n}$. Since

$$\mathscr{H}_{\alpha,3^{-n}}(\mathfrak{C})\ \le\ \sum_{\mathsf{v}\in\{1,2\}^n}\mathit{diam}(J_\mathsf{v})^\alpha\ =\ \frac{2^n}{3^{\alpha n}}\ =\ e^{n(\log2-\alpha\log3)}$$

and the last term tends to 0 for $\alpha>\log2/\log3$, we infer $\mathscr{H}_\alpha(\mathfrak{C})=0$ for these α and therefore $\dim_H\mathfrak{C}\le\log2/\log3$.

Now we turn to a proof of $\dim_H\mathfrak{C}\ge\log2/\log3$ by making use of the mass distribution principle. Let Λ_n be the uniform distribution on $\bigcup_{\mathsf{v}\in\{1,2\}^n}J_\mathsf{v}$. Then it is a well-known fact that Λ_n converges weakly to the "uniform distribution" Λ on $\mathfrak{C}$, called *Cantor distribution*. This distribution is continuous but singular with respect to $\lambda\!\!\lambda$, and it satisfies $\Lambda(J_\mathsf{v})=\Lambda_n(J_\mathsf{v})=(3/2)^{|\mathsf{v}|}\lambda\!\!\lambda(J_\mathsf{v})=2^{-|\mathsf{v}|}$ for each $n\ge|\mathsf{v}|$. Now let $n \in \mathbb{N}$ be arbitrary and $B \in \mathscr{B}([0,1])$ be any set with $3^{-(n+1)} \leq diam(B) < 3^{-n}$. Then B intersects at most one interval I_n, say, of the J_ν for $\nu \in \{1,2\}^n$. Consequently,

$$\Lambda(B) \leq \Lambda(I_n) = 2^{-n} = \left(3^{-n}\right)^{\log 2/\log 3} \leq \left(3\, diam(B)\right)^{\log 2/\log 3}$$

and the asserted inequality follows by Proposition D.11. □

Problems

Problem D.13. Prove that η_ε defined by the general construction scheme after Theorem D.2 is indeed an outer measure for any $\varepsilon > 0$.

Problem D.14. For any $0 < a < 1/2$, consider the Cantor set $\mathfrak{C}_a$ obtained by first removing the interval $((1-a)/2, (1+a)/2)$ from the unit interval $[0,1]$ and then proceeding indefinitely in the obvious manner. Find the Hausdorff dimension of $\mathfrak{C}_a$ by making use of the mass distribution principle.

Index