Mathematical Statistics
Stockholm University

Some results in the theory of Markov chains and in renewal theory

Andreas Nordvall Lagerås

Research Report 2005:12
Licentiate thesis

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2005:12**
Licentiate thesis,
http://www.math.su.se/matstat

Some results in the theory of Markov chains and in renewal theory

Andreas Nordvall Lagerås*

October 2005

Abstract

This licentiate thesis consists of two articles

1. "Central limit theorems for contractive Markov chains"†, published in Nonlinearity **18** (2005) 1955–1965
2. "A renewal process type expression for the moments of inverse subordinators"‡, to be published in Journal of Applied Probability **42.4** (2005)

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Stockholm, Sweden. E-mail: andreas@math.su.se.

Acknowlegdements

First of all I wish to thank my supervisor Thomas Höglund. He was my first lecturer in probability theory and immediately got me interested in the subject. He always takes the time to read what I have written, and often finds the embarrassing errors. Anders Martin-Löf has also been inspiring and always helpful in explaining difficult topics in the most intuitive way imaginable.

In the first year of my graduate studies the following persons were also important: Allan Gut who during his graduate course in probability in Uppsala gave me the confidence to "do mathematics", and my co-author Örjan Stenflo, who really showed what it takes to do research and write an article, i.e. diligence.

I further want to thank all colleagues at the division of Mathematical Statistics, especially Mathias Lindholm, who is also a dear friend of mine. Other close friends, who provide the relaxation from work that is at times much needed, are (in no particular order) Emelie Olsson, David Seim, Johan Holmgren and Frank Majgren.

Last, but not least, I want to thank my family: my father Anders Nordwall, my mother Maria Lagerås, my grandparents Ann-Marie Lagerås, Irene and Eric Nordwall, my brothers and sister Max Lagerås, Daniel and Columbine Nordwall. Without you and your love and support this thesis would not have been written.

The two articles of this licentiate thesis treat two different areas of probability theory and the theory of stochastic processes. I will in this short introduction try to explain some background and the main results without going into details.

1 Central limit theorems for Markov chains

Let us start with the first article that I have written together with Örjan Stenflo. As implied by its title, it is about central limit theorems for contractive Markov chains. In order to understand the meaning of all this we start with a simple but interesting Markov chain $\{X_n\}$ on $\mathbb{R}^2$. Let its starting point be $X_0 = (0.5, 0.5)$. In every step X_n will jump half-way towards $(0,0), (0,1)$ or $(1,0)$ with equal probability. If we draw the points that the chain visits, we will get a picture that looks like figure 1.

This set is called the Sierpinski triangle. The stationary distribution of the chain will be the uniform distribution on this set.

Let us define three functions that describe the jumps of the chain:

$$
\begin{aligned}
w_1(x_1, x_2) &= (0.5x_1, 0.5x_2) && \text{jump towards } (0,0)\\
w_2(x_1, x_2) &= (0.5x_1, 0.5x_2 + 0.5) && \text{jump towards } (0,1)\\
w_3(x_1, x_2) &= (0.5x_1 + 0.5, 0.5x_2) && \text{jump towards } (1,0)
\end{aligned}
$$

We can now write $X_n = w_{I_n} \circ w_{I_{n-1}} \circ \cdots \circ w_{I_1}(0.5, 0.5)$ where $I_1, I_2, \ldots$ are independent random variables equally distributed on $\{1, 2, 3\}$ that indicates what type of jump X_n makes in every step. It is actually possible to write any Markov chain as such an "Iterated function system" (IFS), where $\{w_i\}_{i\in[0,1]}$ is a set of functions and the indices $I_1, \ldots$ are independent and uniformly distributed on $[0, 1]$. This is no news for anyone who has ever simulated a Markov chain with a given transition matrix. We call the Markov chain contractive if all the functions $\{w_i\}$ are contractions, that is $d(w_i(x), w_i(y)) < d(x, y)$ for all i, where $d(x, y)$ is the distance between x and y.

The functions $\{w_i(x)\}$ can be quite complicated as functions of x. It is therefore sometimes easier to let the distribution of I_n depend on X_{n-1}. In our example we could change the transition probabilities so that the chain has a higher probability of jumping towards the closest corner. So the corner are in some sense attracting the chain. If we let the chain jump towards the closest corner with probability 0.8 and jump towards either of the other two corners with equal probabilities 0.1 we get the picture of figure 2.

We can introduce functions $\{p_i(x)\}$ defined by $p_i(x) = P(I_n = i|X_{n-1} = x)$ describing the probabilities of the different jumps depending on what point the chain is in at the moment.

$$p_1(x_1,x_2) = \begin{cases} 0.8 & \text{if } x_1 \le 0.5 \text{ and } x_2 \le 0.5 \\ 0.1 & \text{else} \end{cases}$$
$$p_2(x_1,x_2) = \begin{cases} 0.8 & \text{if } x_1 > 0.5 \\ 0.1 & \text{else} \end{cases}$$
$$p_3(x_1,x_2) = \begin{cases} 0.8 & \text{if } x_2 > 0.5 \\ 0.1 & \text{else} \end{cases}$$

A type of law of large numbers for Markov chains $\{X_n\}$ says that if the Markov chain has a unique stationary distribution, then

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to Ef(X) \quad \text{a.s.}$$

where X has the stationary distribution and f is some real-valued function. If you have a law of large numbers then you obviously ask if you also have something similar to a central limit theorem. In our case we could want to see if

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(f(X_k) - Ef(X)) \quad \text{or} \quad \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(f(X_k) - E[f(X_k)])$$

converge to Gaussian random variables. A stronger form of central limit theorems are the so called functional central limit theorems. They describe the convergence of not only the sum but all partial sums at the same time. We consider

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}(f(X_k) - Ef(X))$$

where $0 \le t \le 1$. Subject to some conditions this converges, considered as a stochastic process in t, to a Brownian motion on $[0,1]$.

The main result of the article is conditions on f and $\{p_i\}$ that implies functional central limit theorems. Loosely speaking the conditions are about how continuous f and the p_i's have to be. If the p_i's are highly regular then we can allow more "wild" f's and vice versa. One can also state the results with the rate of convergence towards the stationary distribution for X_n. Often one consider chains that in a certain sense have "exponential" rate of convergence, but our results also work with even slower convergence.

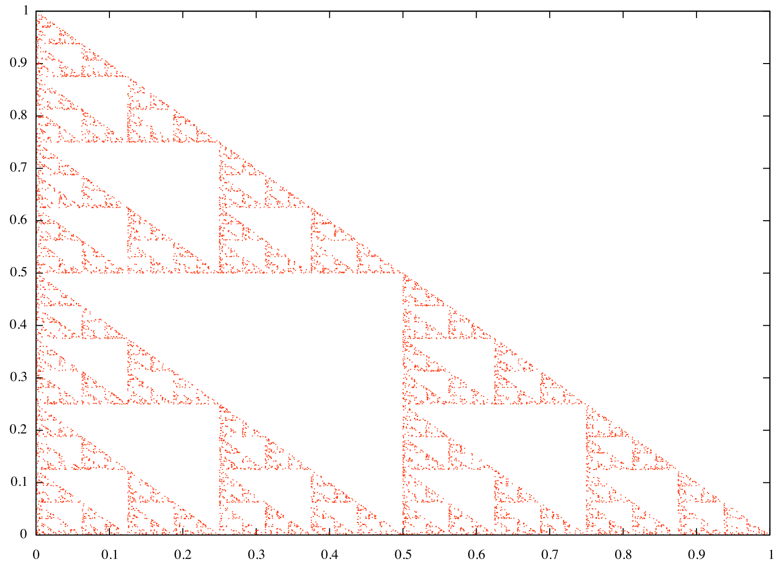

Figure 1: The Sierpinski triangle drawn with the first 10000 points visited by $\{X_n\}$ when it makes the jumps independent of location.

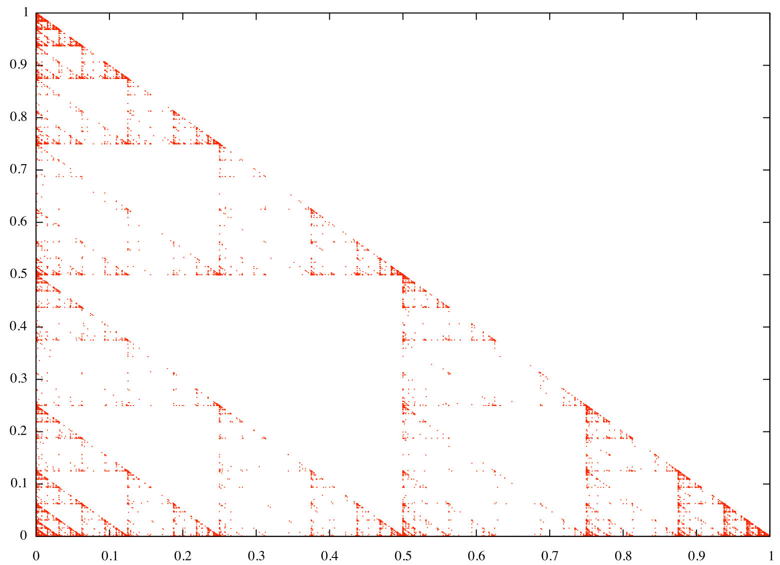

Figure 2: The first 50000 points visited by $\{X_n\}$ when the corners are attracting. Note that more points are included here compared to figure 1.

2 Inverse subordinators

The second article is about properties of a type processes that are called inverse subordinators. In order to understand these processes and why one would want to study them it is easiest to recall some facts about renewal processes. The simplest description of a renewal process is that it gives the number of light bulbs that have been changed at a given time, if one light bulb is always on, one changes the light bulb immediately when it goes out and all the life lengths of the light bulbs are independent and identically distributed. We need not model the changing of light bulbs, but could of course use the model for other components than light bulbs. The model can obviously be used whenever one has reoccurring events (such as the changing of a light bulb) after which the process always "renews" itself, i.e. the time to the next event is independent of what has happened before and the times between events are equally distributed.

If $X_1, X_2, \ldots$ are the life lengths of the light bulbs, the value of the renewal process at time t is given by $N(t) = \min(n; \sum_{k=1}^{n} X_k \geq t)$. An important observation is that this is the inverse of the random walk $S(n) = \sum_{k=1}^{n} X_k$. See figures 3 and 4. If X_1 is given a certain distribution different from the distribution of $X_2, X_3, \ldots$ the whole renewal process can become stationary.

From a modelling point of view it can be problematic that $N(t)$ is integer-valued. One could want to have a process on the whole of $\mathbb{R}_+$ but with some renewal structure similar to renewal processes. The solution is the inverse subordinators. To make the transition from integer-valued to truly real-

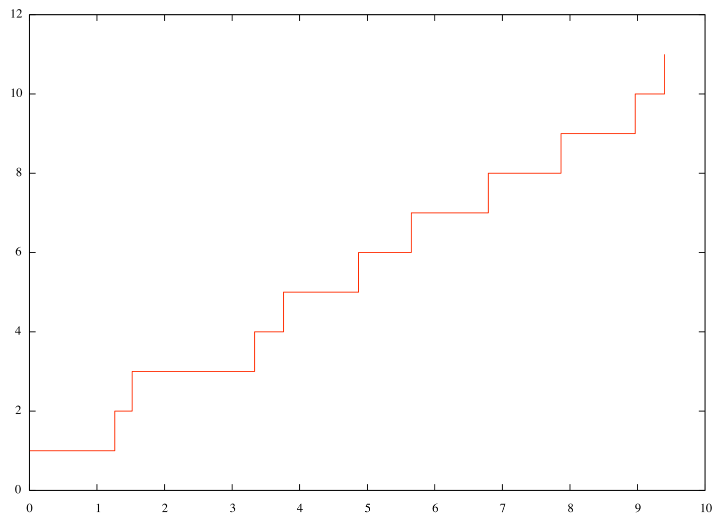

Figure 3: Renewal process.

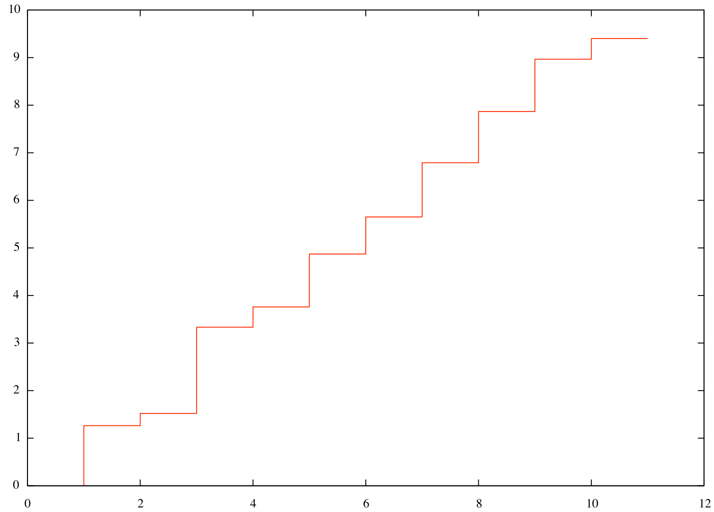

Figure 4: Random walk—the inverse of the renewal process in figure 3.

valued we use the connection between renewal processes and random walks. The inverse of a random walk in continuous time gives us precisely a real-valued process. What is then a random walk in continuous time? It is a process with independent increments over disjoint time intervals, and its increments are homogeneous in the sense that increments over intervals of equal length are equally distributed. Such processes are called Lévy processes. We only consider random walks with positive increments so that there exists an inverse. Increasing Lévy processes are called subordinators, and thus we have at least explained the name inverse subordinators. Se figures 5 and 6 for examples of a subordinator and its inverse. It should be mentioned that inverse subordinators not only arise due to modelling needs, but they also appear in completely theoretical settings. The local time of a large class of Markov processes are inverse subordinators for example.

What are then the properties of inverse subordinators? One nice property is that if we round down the value of the inverse subordinator, we get a renewal process. We actually get (scaled) renewal processes if we round down to the closest n:th part or n-multiple as well. A large class of renewal processes can thus be seen as approximations of an inverse subordinator. A property of renewal processes is that the number of events in disjoint time intervals are dependent (except when the times between events are exponentially distributed). This means that the increments are dependent. This is also the case for inverse subordinators. It is very difficult to obtain explicit expressions for the joint distribution of the increments of renewal processes and subordinators. Nevertheless an expression for joint moments

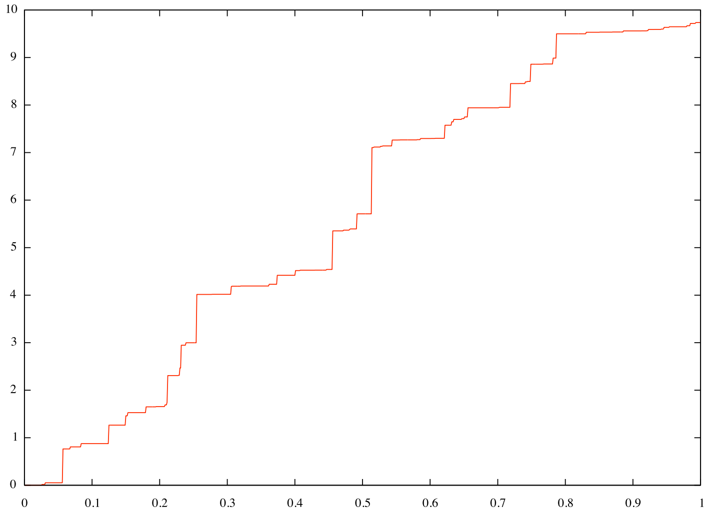

Figure 5: A subordinator with gamma increments. The distribution at 1 is gamma(10,1)

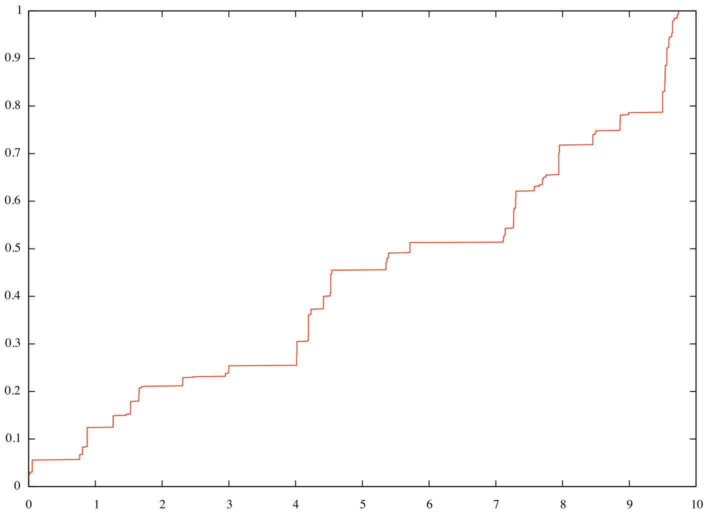

Figure 6: Inverse subordinator—the inverse of the subordinator of figure 5.

of any order is known for renewal processes. We call $E[X^{[n]}] = E[X(X-1)\cdots(X-n+1)]$ the factorial moment of order n. It just takes some algebra to relate the factorial moments and the ordinary moments. If $V(t) = EN(t)$, we can calculate the factorial moments of the increments over the disjoint intervals $(s_i, t_i], i = 1, \ldots, n$ with

$$E\Big[\prod_{i=1}^{n}(N(t_i) - N(s_i))^{[k_i]}\Big] = \prod_{i=1}^{n} k_i! \cdot \int_C \prod_{j=1}^{k} V(dx_j - x_{j-1})$$

where $k = k_1 + \cdots + k_n$ and C is a particular subset of $\mathbb{R}^k$. This result is interesting in itself, but it becomes even more interesting when we compare it to the main result of the second article, the corresponding expression for inverse subordinators $\tau(t)$

$$E\Big[\prod_{i=1}^{n}(\tau(t_i) - \tau(s_i))^{k_i}\Big] = \prod_{i=1}^{n} k_i! \cdot \int_C \prod_{j=1}^{k} U(dx_j - x_{j-1})$$

where $U(t) = E\tau(t)$. Note that the only difference is that we have ordinary moments here and not factorial.

How does one show such a result? One could possibly use the aforementioned property that an inverse subordinator can be approximated with scaled renewal processes. My proof instead uses previously known results about a type processes called Cox processes. A Cox process is loosely speaking an inhomogeneous Poisson process with a randomized intensity measure, i.e. the expected number of points in any interval is given by a *random* measure. It is easily shown that the factorial moments of a Cox process equal the ordinary moments of the random measure. The connection with inverse subordinators is the following: If the (cumulative) intensity measure is given by an inverse subordinator then the Cox process is also a renewal process. In this case we can use the first expression to calculate the factorial moments of the process since it is a renewal process. But since it is a Cox process these factorial moments equal the ordinary moments of the intensity measure. Noting that $U = V$, we are done.

With this connection one can also show how to obtain a stationary version of the inverse subordinator, and prove so called renewal theorems.

Beside the main result, the article also provides new proofs of known expressions for the so called double Laplace transform of the marginal distribution of the inverse subordinators, and explicit expressions for $U(t)$ for some types of inverse subordinators.

Nonlinearity **18** (2005) 1955–1965 doi:10.1088/0951-7715/18/5/005

Central limit theorems for contractive Markov chains

Andreas Nordvall Lagerås and Örjan Stenflo

Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden

E-mail: andreas@math.su.se and stenflo@math.su.se

Received 9 February 2005, in final form 3 May 2005
Published 6 June 2005
Online at stacks.iop.org/Non/18/1955

Recommended by C P Dettmann

Abstract
We prove limit theorems for Markov chains under (local) contraction conditions. As a corollary we obtain a central limit theorem for Markov chains associated with iterated function systems with contractive maps and place-dependent Dini-continuous probabilities.

Mathematics Subject Classification: 60F05, 60J05, 60B10, 37H99

1. Introduction

Let (X, d) be a compact metric space, typically a closed and bounded subset of $\mathbb{R}$ or $\mathbb{R}^2$ with the Euclidean metric and let $\{w_i\}_{i=1}^N$ be a family of (strict) contraction maps on X, i.e. there exists a constant $c < 1$ such that $d(w_i(x), w_i(y)) \leqslant cd(x, y)$, for any $x, y \in X$ and integer $1 \leqslant i \leqslant N$. Such a system is called an iterated function system (IFS) (see [1]). Hutchinson [12] and Barnsley and Demko [1] introduced these objects in order to describe fractals. It is easy to see that there exists a unique compact set K that is invariant for the IFS in the sense that $K = \bigcup_{i=1}^N w_i(K)$. The set K is called the fractal set, or attractor, associated with the IFS. If the maps w_i are non-degenerate and affine and the sets $w_i(K)$, $1 \leqslant i \leqslant N$, are 'essentially' disjoint, then K will have the characteristic 'self-similar' property of a fractal. The huge class of examples of fractals that can be described in this way includes the Sierpinski gasket, Barnsley's fern, the Cantor set and many, many others. Despite fractals being totally deterministic objects, the simplest way of drawing pictures of fractals is often via Barnsley's 'random iteration algorithm': attach probabilities, p_i, to each map w_i ($\sum_i p_i = 1$). Choose a starting point $Z_0(x) := x \in X$. Choose a function, w_{I_1}, at random from the IFS, with $P(w_{I_1} = w_k) = p_k$. Let $Z_1(x) = w_{I_1}(x)$. Next, independently, choose a function, w_{I_2}, in the same manner and let $Z_2(x) = w_{I_2}(Z_1(x)) = w_{I_2} \circ w_{I_1}(x)$. Repeat this 'random iteration' procedure inductively and define $Z_n(x) = w_{I_n} \circ w_{I_{n-1}} \circ \cdots \circ w_{I_1}(x)$. The random sequence $\{Z_n(x)\}$ forms a Markov chain with a unique stationary probability distribution, μ, supported on K. Since

$$\frac{\sum_{k=0}^{n-1} f(Z_k(x))}{n} \to \int f \, \mathrm{d}\mu \quad \text{a.s.},$$

as $n \to \infty$, for any real-valued continuous function f on X, by Birkhoff's ergodic theorem (note that x can be chosen to be *any* fixed point by the contraction assumption), we will 'draw a picture of the attractor K' by 'plotting' the orbit $\{Z_n(x)\}$, possibly ignoring some of the first points in order to reach the stationary regime. This algorithm will be an efficient way of 'drawing a picture of K' provided the probabilities are chosen in such a way as to make the stationary distribution as uniform as possible on K and the stationary state is reached sufficiently fast. The choice of p_i can sometimes be made by inspection, by searching for a stationary distribution with the same dimension as K itself. The convergence rates towards the stationary state are 'heuristically justified' by central limit theorems (CLTs), where

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(f(Z_k(x)) - \int f\,\mathrm{d}\mu\right)$$

converges in distribution to the normal distribution for f belonging to some suitably rich class of real-valued functions on X, or by stronger forms of CLTs, the so-called invariance principles or functional CLTs, where the stochastic process

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{[nt]-1}\left(f(Z_k(x)) - \int f\,\mathrm{d}\mu\right), \qquad 0 \leqslant t \leqslant 1 \tag{1}$$

converges in distribution to a Brownian motion. (Here $[x]$ denotes the integer part of x.) Note that expression (1) above is a function-valued random element. See [5] for details about the concept of convergence in distribution for function-valued random elements.

The purpose of this paper is to study Markov chains generated by IFSs with place-dependent probabilities. (Such Markov chains have also been studied under the name 'random systems with complete connections', see [13].) We are given a set of contraction maps $\{w_i\}$, with associated continuous functions $p_i = p_i(x)$, where $p_i : X \to (0, 1)$, with $\sum_i p_i(x) = 1$, for any $x \in X$. The Markov chains are characterized by the transfer operator T defined for real-valued measurable functions f on X by $Tf(x) = \sum_i p_i(x) f(w_i(x))$. Intuitively, the Markov chains considered are generated by fixing a starting point x and letting $Z_0(x) := x$, and inductively letting $Z_{n+1}(x) := w_i(Z_n(x))$ with probability $p_i(Z_n(x))$ for $n \geqslant 0$.

One motivation for studying such chains is that it gives more freedom when trying to generate a 'uniform' stationary probability distribution on K. Such Markov chains also arise naturally in the thermodynamic formalism of statistical mechanics. It is well known that they do not necessarily possess a unique stationary distribution (see [4, 6, 20, 26, 27]), but with some additional regularity conditions, uniqueness holds (see [11, 14, 27, 28]).

The operator T (without the normalizing condition $\sum_i p_i(x) = 1$) is known as the Ruelle–Perron–Frobenius operator. Fan and Lau [10] proved a limit theorem for iterates of the Ruelle–Perron–Frobenius operator under the Dini-continuity assumptions on the p_i, by lifting a similar result from Walters [29] on symbolic spaces. (Recall that p_i is Dini-continuous if $\int_0^1 (\Delta_{p_i}(t)/t)\,\mathrm{d}t < \infty$, or equivalently $\sum_{n=0}^{\infty} \Delta_{p_i}(c^n) < \infty$, for some (and thus all) $0 < c < 1$, where $\Delta_{p_i}(t) := \sup_{d(x,y)\leqslant t} |p_i(x) - p_i(y)|$ is the modulus of uniform continuity of p_i). Uniqueness in stationary distributions still holds (in the normalized cases) if the contraction assumptions of the w_i are relaxed to 'average contraction' under the Dini-continuity assumption (see [2, 17]) but information about rates of convergence in these 'average contractive' cases seems to be unknown.

The Dini-condition is somewhat stronger than the weakest known conditions for uniqueness in stationary probability distributions (in the normalized cases with strict contraction maps), but weaker than, e.g., Hölder-continuity.

In the Dini-continuous cases it follows that the unique equilibrium measure will have the Gibbs (approximation) property (see [10]). This property is of importance when analysing the multidimensional spectra of measures.

In this paper we will prove the perhaps initially surprising fact (corollary 2) that Markov chains generated by IFSs with Dini-continuous probabilities obey a CLT, despite the well-known fact that such Markov chains do not typically converge with an exponential rate. Our main result, theorem 1, expresses this in a natural generality.

CLTs/functional CLTs for iterated random functions under conditions that imply exponential (or other rapid) rates of convergence have previously been proved in, e.g., [3, 15, 16, 22, 30, 31]. We discuss the connection between some of these results and our results in remarks 4 and 6.

2. Preliminaries

Let $\mathcal{B}$ denote the Borel σ-field generated by the metric d, and let $\boldsymbol{P} : X \times \mathcal{B} \to [0, 1]$ be a transition probability. That is, for each $x \in X$, $\boldsymbol{P}(x, \cdot)$ is a probability measure on $(X, \mathcal{B})$ and for each $A \in \mathcal{B}$, $\boldsymbol{P}(\cdot, A)$ is $\mathcal{B}$-measurable. The transition probability generates a Markov chain with transfer operator defined by $Tf(x) = \int_X f(y)\boldsymbol{P}(x, \mathrm{d}y)$ for real-valued measurable functions f on X. A probability measure μ is stationary for $\boldsymbol{P}$ if $\mu(\cdot) = \int_X \boldsymbol{P}(x, \cdot)\,\mathrm{d}\mu(x)$.

There are several ways of representing a Markov chain with a given transfer operator. One common way is to find a measurable function $w : X \times [0, 1] \to X$, let $\{I_j\}_{j=1}^{\infty}$ be a sequence of independent random variables uniformly distributed in [0, 1], and consider the random dynamical system defined by

$$Z_n(x) := w_{I_n} \circ w_{I_{n-1}} \circ \cdots \circ w_{I_1}(x), \quad n \geqslant 1, \qquad Z_0(x) := x,$$

for any $x \in X$, where

$$w_s(x) = w(x, s).$$

It is always possible to find such a representation, w, such that the transition probability generated by $\{Z_n\}$ is $\boldsymbol{P}$, i.e. $Ef(Z_n(x)) = T^n f(x)$, for any x, n and f (see [19]).

For two fixed points $x, y \in X$ and $\boldsymbol{x} = (x, y)$ we can consider the Markov chain $\{\boldsymbol{Z}_n(\boldsymbol{x})\}$, on X^2, where $\boldsymbol{Z}_n(\boldsymbol{x}) := (Z_n(x), Z_n(y))$. When proving theorems based on contraction conditions we are typically interested in representations that minimize $d(Z_n(x), Z_n(y))$ (in some average sense).

More generally, if $W : X^2 \times [0, 1] \to X^2$, is a measurable map and $\{I_j\}_{j=1}^{\infty}$ is a sequence of independent random variables uniformly distributed in [0, 1], we will consider the random dynamical system defined by

$$\boldsymbol{Z}_n(\boldsymbol{x}) := W_{I_n} \circ W_{I_{n-1}} \circ \cdots \circ W_{I_1}(\boldsymbol{x}), \quad n \geqslant 1, \qquad \boldsymbol{Z}_0(\boldsymbol{x}) := \boldsymbol{x}, \tag{2}$$

where $W_s(\boldsymbol{x}) = W(\boldsymbol{x}, s)$, such that, for any $\boldsymbol{x} = (x, y) \in X^2$, the Markov chain $\boldsymbol{Z}_n(\boldsymbol{x}) =: (Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$ on X^2 has marginals $\boldsymbol{P}^n(x, \cdot) = P(Z_n^{(x,y)}(x) \in \cdot)$, and $\boldsymbol{P}^n(y, \cdot) = P(Z_n^{(x,y)}(y) \in \cdot)$, for any n.

Thus $\{Z_n^{(x,y)}(x)\}$ and $\{Z_n^{(x,y)}(y)\}$ denote two Markov chains on X, defined on the same probability space, with the former starting at $x \in X$ and the latter starting at $y \in X$, both with transition probability $\boldsymbol{P}$.

Let d_w be the Monge–Kantorovich metric defined by $d_w(\pi, \nu) = \sup(\int f\,\mathrm{d}(\pi - \nu);\ f : X \to \mathbb{R}, |f(x) - f(y)| \leqslant d(x, y)\,\forall x, y)$, for probability measures π and ν on X. The Monge–Kantorovich metric metrizes the topology of weak convergence on the set of probability measures on X (see [9]). It follows from the definitions that for any stationary probability measure μ, we have

$$d_w(\boldsymbol{P}^n(x,\cdot),\mu(\cdot)) \leqslant \sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)). \tag{3}$$

Therefore if $\sup_{x,y} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \to 0$ as $n \to \infty$, then there is a unique stationary distribution for $\boldsymbol{P}$.

We will sometimes drop the upper index, i.e. write $Z_n(x)$ instead of $Z_n^{(x,y)}(x)$ etc, when we are not interested in the joint distribution of the pair $(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$.

The following proposition gives sufficient conditions for the existence of a CLT.

Proposition 1. *Suppose there exists a unique stationary distribution μ for $\boldsymbol{P}$, and let f be a real-valued measurable function on X with $\|f\|_{L^2}^2 = \int f^2\,\mathrm{d}\mu < \infty$. Suppose that for some $\delta > 0$,*

$$\lim_{n\to\infty} n^{-1/2}(\log n)^{1+\delta} \sup_{x,y\in X} E\sum_{k=0}^{n-1} | f(Z_k^{(x,y)}(x)) - f(Z_k^{(x,y)}(y)) | = 0. \tag{4}$$

Let

$$S_n^x = \sum_{k=0}^{n-1}(f(Z_k(x)) - Ef(Z_k(x))),$$

$$S_n^{x,\mu} = \sum_{k=0}^{n-1}\left(f(Z_k(x)) - \int f\,\mathrm{d}\mu\right)$$

and

$$B_n^x(t) = \frac{S_{[nt]}^x}{\sqrt{n}}, \qquad 0 \leqslant t \leqslant 1,$$

$$B_n^{x,\mu}(t) = \frac{S_{[nt]}^{x,\mu}}{\sqrt{n}}, \qquad 0 \leqslant t \leqslant 1.$$

Then the limit

$$\sigma^2 = \sigma^2(f) := \lim_{n\to\infty}\frac{1}{n}E[(S_n^Z)^2] \tag{5}$$

exists and is finite, where Z is a μ-distributed random variable, independent of $\{I_j\}_{j=1}^{\infty}$. Furthermore, if $B = \{B(t) : 0 \leqslant t \leqslant 1\}$ denotes the standard Brownian motion, then

$$B_n^x \xrightarrow{\mathrm{d}} \sigma B \tag{6}$$

and

$$B_n^{x,\mu} \xrightarrow{\mathrm{d}} \sigma B, \tag{7}$$

as $n \to \infty$, for any $x \in X$, where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution for random elements taking values in the space of right-continuous functions on $[0, 1]$ *with left-hand limits equipped with the Skorokhod topology.*

Remark 1. Proposition 1 above is valid when $(X, \mathcal{B})$ is a general measurable space.

Remark 2. General CLTs for Markov chains started at a point have been proved by Derriennic and Lin [7]. Proposition 1 complements their result in cases of 'uniform' ergodicity. The proof of proposition 1, given later, relies on a slightly stronger result by Peligrad and Utev [23] for

Markov chains starting according to the unique stationary probability distribution. Theorems about convergence, allowing Markov chains to start at a point, are important in the theory for Markov chain–Monte Carlo methods.

Remark 3. In an earlier draft of this paper we proved a weaker (non-functional) form of the CLT in proposition 1, where our result was based on a CLT by Maxwell and Woodroofe [21]. The recent paper by Peligrad and Utev [23], which was helpfully pointed out to us by a referee, enabled us to state our CLT in the current functional CLT form.

Remark 4. Wu and Woodroofe considered general state spaces in [31]. The conditions in their CLT (theorem 2) imply (4) in the case of a compact X. This can be seen as follows: their proof of this theorem amounts to showing that $\sum_{n=0}^{\infty} \|T^n f\|_{L^2} < \infty$, for centred functions f. Restricting X to be compact allows a strengthening of their lemma 3, so that its result holds even when starting $\{Z_k(x)\}$ from a point. With some minor modifications to the proof, it is possible to show that $\sum_{n=0}^{\infty} \sup_{x,y} E|f(Z_n^{(x,y)}(x)) - f(Z_n^{(x,y)}(y))| < \infty$. Thus the conditions of our proposition 1 hold.

Checking the L^2 boundedness condition could be difficult if we have no *a priori* information about the (possibly non-unique) stationary measures. The following corollary circumvents these problems and might therefore be more directly applicable in our case when (X, d) is compact.

Corollary 1. *If*

$$\lim_{n\to\infty} \sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) = 0, \tag{8}$$

then there exists a unique stationary distribution μ for $\boldsymbol{P}$.

Let f be a real-valued continuous function on X. Suppose $\Delta_f : \mathbb{R}^+ \to \mathbb{R}^+$ is an increasing concave function with $\Delta_f(t) \geqslant \sup_{d(x,y)\leqslant t} |f(x) - f(y)|$, for any $t \geqslant 0$ and suppose, in addition to (8), that for some $\delta > 0$,

$$\lim_{n\to\infty} \sqrt{n}(\log n)^{1+\delta} \Delta_f \left(\sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \right) = 0 \tag{9}$$

also holds, then the conclusions of proposition 1 hold for f, i.e. the limit (5) exists for f and is finite and (6) and (7) hold.

Remark 5. The function Δ_f may thus be chosen to be the modulus of uniform continuity of f in cases when this function is concave.

Remark 6. If $\sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \sim O(c^n)$, for some constant $c < 1$, satisfied for instance the average-contractive IFSs with place-independent probabilities, then it follows from corollary 1 that the CLT holds with respect to any f of modulus of uniform continuity Δ_f, of order $\Delta_f(c^n) \sim o(1/\sqrt{n}(\log n)^{1+\delta})$. This condition is satisfied by, e.g., Dini-continuous functions f. Corollary 1 thus strengthens theorem 2.4. of [3] (who considered Lipschitz-continuous f). Wu and Shao [30] considered functions f that are stochastically Dini-continuous with respect to the stationary distribution. (It should be noted that [3] and [30] treated average contractive IFSs on more general metric spaces.)

Proof (proposition 1). Let $f \in L^2(\mu)$ be a real-valued measurable function on X satisfying assumption (4).

Since

$$\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\left\|\sum_{k=0}^{n-1}T^k\left(f-\int f\,\mathrm{d}\mu\right)\right\|_{L^2}=\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\left\|\sum_{k=0}^{n-1}\left(T^kf-\int f\,\mathrm{d}\mu\right)\right\|_{L^2}$$
$$\leqslant\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\sup_{x\in X}\left|\sum_{k=0}^{n-1}\left(T^kf(x)-\int f\,\mathrm{d}\mu\right)\right|$$
$$=\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\sup_{x\in X}\left|\sum_{k=0}^{n-1}\left(T^kf(x)-\int T^kf\,\mathrm{d}\mu\right)\right|$$
$$\leqslant\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\sup_{x,y\in X}\left|\sum_{k=0}^{n-1}\left(T^kf(x)-T^kf(y)\right)\right|$$
$$=\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\sup_{x,y\in X}\left|E\sum_{k=0}^{n-1}(f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y)))\right|$$
$$\leqslant\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\sup_{x,y\in X}E\sum_{k=0}^{n-1}\left|f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y))\right|<\infty,$$

it follows from theorem 1.1 of [23] that $\sigma^2=\lim_{n\to\infty}(1/n)E[(S_n^Z)^2]$ exists and is finite, and $B_n^{Z,\mu}\stackrel{\mathrm{d}}{\to}\sigma B$, where Z is a μ-distributed random variable, independent of $\{I_j\}_{j=1}^{\infty}$. By Chebyshev's inequality,

$$P\left(\sup_{0\leqslant t\leqslant 1}|B_n^{x,\mu}(t)-B_n^{Z,\mu}(t)|\geqslant\epsilon\right)$$
$$=P\left(\max_{0\leqslant m\leqslant n}\frac{1}{\sqrt{n}}\left|\sum_{k=0}^{m-1}(f(Z_k^{(x,Z)}(x))-f(Z_k^{(x,Z)}(Z)))\right|\geqslant\epsilon\right)$$
$$\leqslant P\left(\frac{1}{\sqrt{n}}\max_{0\leqslant m\leqslant n}\sum_{k=0}^{m-1}\left|f(Z_k^{(x,Z)}(x))-f(Z_k^{(x,Z)}(Z))\right|\geqslant\epsilon\right)$$
$$\leqslant P\left(\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left|f(Z_k^{(x,Z)}(x))-f(Z_k^{(x,Z)}(Z))\right|\geqslant\epsilon\right)$$
$$\leqslant\frac{1}{\epsilon\sqrt{n}}E\sum_{k=0}^{n-1}\left|f(Z_k^{(x,Z)}(x))-f(Z_k^{(x,Z)}(Z))\right|$$
$$\leqslant\frac{1}{\epsilon\sqrt{n}}\sup_{x,y\in X}E\sum_{k=0}^{n-1}\left|f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y))\right|\to 0,$$

as $n\to\infty$. By theorem 4.1 in [5], $B_n^{x,\mu}\stackrel{\mathrm{d}}{\to}\sigma B$.

The difference between S_n^x and $S_n^{x,\mu}$ lies in how the summands are centred. The difference is negligible in the limit:

$$\begin{aligned}\sup_{0\leqslant t\leqslant 1}|B_n^{x,\mu}(t)-B_n^x(t)| &= \max_{0\leqslant m\leqslant n}\frac{1}{\sqrt{n}}\left|\sum_{k=0}^{m-1}\left(Ef(Z_k(x))-\int f\,\mathrm{d}\mu\right)\right| \\ &\leqslant \frac{1}{\sqrt{n}}\max_{0\leqslant m\leqslant n}E\left|\sum_{k=0}^{m-1}\left(f(Z_k(x))-\int f(Z_k(y))\,\mathrm{d}\mu(y)\right)\right| \\ &\leqslant \frac{1}{\sqrt{n}}E\sum_{k=0}^{n-1}\left|f(Z_k(x))-\int f(Z_k(y))\,\mathrm{d}\mu(y)\right| \\ &\leqslant \frac{1}{\sqrt{n}}\sup_{x,y\in X}E\sum_{k=0}^{n-1}\left|f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y))\right|\to 0,\end{aligned}$$

as $n\to\infty$. Thus also $B_n^x \xrightarrow{\mathrm{d}} \sigma B$. □

Proof (corollary 1). The first part of the corollary follows from (3) above.

For the proof of the second part of corollary 1, first note that by assumption (9),

$$\Delta_f\left(\sup_{x,y\in X}Ed(Z_n^{(x,y)}(x),Z_n^{(x,y)}(y))\right)\sim o\left(\frac{1}{\sqrt{n}(\log n)^{1+\delta}}\right),$$

implying that

$$\sum_{k=0}^{n-1}\Delta_f\left(\sup_{x,y\in X}Ed(Z_k^{(x,y)}(x),Z_k^{(x,y)}(y))\right)\sim o\left(\frac{\sqrt{n}}{(\log n)^{1+\delta}}\right).$$

(To see this, note that the derivative $F'(t)$ of $F(t)=\sqrt{t}/(\log t)^{1+\delta}$ satisfies $F'(t)\geqslant 1/(3\sqrt{t}(\log t)^{1+\delta})$, for large t.)

Thus,

$$\lim_{n\to\infty}n^{-1/2}(\log n)^{1+\delta}\sum_{k=0}^{n-1}\Delta_f\left(\sup_{x,y\in X}Ed(Z_k^{(x,y)}(x),Z_k^{(x,y)}(y))\right)=0.$$

Since by the definition of Δ_f and Jensen's inequality,

$$\begin{aligned}\Delta_f\left(\sup_{x,y\in X}Ed(Z_n^{(x,y)}(x),Z_n^{(x,y)}(y))\right) &\geqslant \sup_{x,y\in X}\Delta_f(Ed(Z_n^{(x,y)}(x),Z_n^{(x,y)}(y))) \\ &\geqslant \sup_{x,y\in X}E\Delta_f(d(Z_n^{(x,y)}(x),Z_n^{(x,y)}(y))) \\ &\geqslant \sup_{x,y\in X}E|f(Z_n^{(x,y)}(x))-f(Z_n^{(x,y)}(y))|\end{aligned}$$

and

$$\sum_{k=0}^{n-1}\sup_{x,y\in X}E|f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y))|\geqslant\sup_{x,y\in X}E\sum_{k=0}^{n-1}|f(Z_k^{(x,y)}(x))-f(Z_k^{(x,y)}(y))|,$$

we see that an application of proposition 1 completes the proof of the second part of corollary 1. □

3. Main results

Theorem 1. *Let* $W : X^2 \times [0,1] \to X^2$ *be a measurable map such that for any fixed* $(x, y) \in X^2$ *the map* $W(x, y, \cdot) := (W^{(x,y)}(x), W^{(x,y)}(y))(\cdot)$ *defines random variables with* $P(W^{(x,y)}(x) \in \cdot) = \boldsymbol{P}(x, \cdot)$ *and* $P(W^{(x,y)}(y) \in \cdot) = \boldsymbol{P}(y, \cdot)$*, where* P *denotes the uniform probability measure on the Borel subsets of* $[0, 1]$.

Let $\Delta : [0, \infty) \to [0, 1)$*, be an increasing function with* $\Delta(0) = 0$*. Suppose there exists a constant* $c < 1$*, such that*

$$P(d(W^{(x,y)}(x), W^{(x,y)}(y)) \leqslant cd(x, y)) \geqslant 1 - \Delta(d(x, y)), \tag{10}$$

for any two points $x, y \in X$.

Then

(i) (Distributional stability theorem)

$$d_w(\boldsymbol{P}^n(x, \cdot), \mu(\cdot)) \leqslant \sup_{x,y\in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leqslant ED_n, \tag{11}$$

for any stationary probability distribution μ*, where* D_n *is a homogeneous Markov chain with* $D_0 = \mathrm{diam}(X) := \sup_{x,y} d(x, y)$*,*

$$P(D_{n+1} = ct \mid D_n = t) = 1 - \Delta(t)$$

and

$$P(D_{n+1} = \mathrm{diam}(X) \mid D_n = t) = \Delta(t),$$

for any $0 \leqslant t \leqslant \mathrm{diam}(X)$.

If

$$\sum_{n=1}^{\infty} \prod_{k=1}^{n} (1 - \Delta(c^k)) = \infty, \tag{12}$$

then $ED_n \to 0$ *and thus by corollary 1 there is a unique stationary distribution,* μ.

(ii) (Central limit theorem)

If $\sum_{k=0}^{\infty} \Delta(c^k) < \infty$*, then the conclusions of proposition 1 hold for any Hölder-continuous function* f *with exponent* $\alpha > \frac{1}{2}$.

Proof (theorem 1(i)). Fix two points x and y in X. Define $Z_0^{(x,y)}(x) = x$, $Z_0^{(x,y)}(y) = y$ and inductively

$$Z_n^{(x,y)}(x) = W^{(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y))}(Z_{n-1}^{(x,y)}(x))$$

and

$$Z_n^{(x,y)}(y) = W^{(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y))}(Z_{n-1}^{(x,y)}(y)),$$

as in (2). Then $Z_n^{(x,y)}(x)$ and $Z_n^{(x,y)}(y)$ are random variables such that $Ef(Z_n^{(x,y)}(x)) = T^n f(x)$ and $Ef(Z_n^{(x,y)}(y)) = T^n f(y)$, for any n.

We have from assumption (10) that

$$\begin{aligned} &P(d(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leqslant ct \mid d(Z_{n-1}^{(x,y)}(x), Z_{n-1}^{(x,y)}(y)) \leqslant t) \\ &\qquad \geqslant 1 - \Delta(t) = P(D_n = ct \mid D_{n-1} = t), \end{aligned}$$

for any $t \in \{c^k \mathrm{diam}(X)\}_{k=0}^{\infty}$. (Note that D_n takes values in the discrete state space $\{c^k \mathrm{diam}(X)\}_{k=0}^{\infty}$.)

D_n is therefore stochastically larger than $d(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$, and consequently $ED_n \geqslant Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y))$, for any $x, y \in X$. The other inequality of (11) follows from (3).

Since $\{D_n\}$ is a non-ergodic Markov chain under condition (12) (see [24], p 80), it follows that $ED_n \to 0$ as $n \to \infty$, if (12) holds, and we have thus proved theorem 1(i).

In order to prove theorem 1(ii), we first observe that it is well known that $\sum_{k=0}^{\infty} \Delta(c^k) < \infty$ implies that D_n is transient (see [24], p 80). Therefore (see [25], p 575), $\sum_{k=0}^{\infty} P(D_k = \mathrm{diam}(X)) < \infty$ and it follows that

$$\begin{aligned}
\sum_{k=0}^{\infty} ED_k &= \sum_{k=0}^{\infty}\sum_{j=0}^{k} c^j \mathrm{diam}(X) P(D_k = c^j \mathrm{diam}(X)) \\
&\leqslant \mathrm{diam}(X) \sum_{k=0}^{\infty}\sum_{j=0}^{k} c^j P(D_{k-j} = \mathrm{diam}(X)) \\
&= \frac{\mathrm{diam}(X)}{1-c} \sum_{k=0}^{\infty} P(D_k = \mathrm{diam}(X)) < \infty.
\end{aligned}$$

By stochastic monotonicity ED_k is decreasing, and thus $\sum_{k=1}^{n} ED_k \geqslant nED_n$, for any n. This implies that $ED_n \leqslant c_0/n$, for $c_0 := \sum_{k=0}^{\infty} ED_k$.

Thus $\sup_{x,y} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \leqslant c_0/n$, for any $n \geqslant 1$. If f is a Hölder-continuous function on X, with modulus of uniform continuity Δ_f satisfying $\Delta_f(t) \leqslant c_1 t^{\alpha}$, for some constants c_1 and $\alpha > \frac{1}{2}$, and any $t \geqslant 0$, it follows that for any $\delta > 0$,

$$\begin{aligned}
&\lim_{n\to\infty} \sqrt{n}(\log n)^{1+\delta} \Delta_f \left(\sup_{x,y \in X} Ed(Z_n^{(x,y)}(x), Z_n^{(x,y)}(y)) \right) \\
&\quad \leqslant \lim_{n\to\infty} \sqrt{n}(\log n)^{1+\delta} c_1 \left(\frac{c_0}{n}\right)^{\alpha} = 0.
\end{aligned}$$

An application of corollary 1 now completes the proof of theorem 1(ii). □

4. IFSs with place-dependent probabilities

Let $\{w_i\}_{i=1}^{\infty}$ be a set of strictly contracting maps, i.e. there exist a constant $c < 1$ such that $d(w_i(x), w_i(y) \leqslant cd(x, y)$, for any $x, y \in X$ and any integer i. Let $\{p_i(x)\}_{i=1}^{\infty}$ be associated place-dependent probabilities, i.e. non-negative continuous functions, with $\sum_i p_i(x) = 1$, for any $x \in X$. This system defines a Markov chain with transfer operator defined by $Tf(x) = \sum_{i=1}^{\infty} p_i(x) f(w_i(x))$, for real-valued measurable functions f on X.

Let

$$\Delta(t) = \frac{1}{2} \sup_{d(x,y)\leqslant t} \sum_{i=1}^{\infty} |p_i(x) - p_i(y)| = 1 - \inf_{d(x,y)\leqslant t} \sum_{i=1}^{\infty} \min(p_i(x), p_i(y)) \tag{13}$$

and let for any two points $x, y \in X$, $W^{(x,y)}(x)$ and $W^{(x,y)}(y)$ be random variables defined by

$$P(W^{(x,y)}(x) = w_i(x), W^{(x,y)}(y) = w_i(y)) = \min(p_i(x), p_i(y)) \tag{14}$$

and

$$P(W^{(x,y)}(x) = w_i(x), W^{(x,y)}(y) = w_j(y)) = \frac{(p_i(x) - \min(p_i(x), p_i(y)))(p_j(y) - \min(p_j(x), p_j(y)))}{1 - \sum_{k=1}^{\infty} \min(p_k(x), p_k(y))}, \tag{15}$$

when $i \neq j$. (If $p_i(x) = p_i(y), \forall\, i$, then we understand the expression in (15) as zero.)

It is straightforward to check that by construction $P(W^{(x,y)}(x) = w_i(x)) = p_i(x)$, and $P\big(W^{(x,y)}(y) = w_j(y)\big) = p_j(y)$ for any i and j.

It follows from (14) that

$$P(d(W^{(x,y)}(x), W^{(x,y)}(y)) \leqslant c d(x, y)) \geqslant \sum_{i=1}^{\infty} \min(p_i(x), p_i(y)) \geqslant 1 - \Delta(d(x, y)),$$

and we may therefore apply theorem 1 to obtain the following.

Corollary 2. *Let $\{w_i\}_{i=1}^{\infty}$ be an IFS with strictly contractive maps, and let $\{p_i(x)\}$ be associated place-dependent probabilities. Then the conclusions of theorem 1 hold with Δ defined as in (13) above.*

Let us illustrate the above corollary with an example.

Example 1. Let w_1 and w_2 be two maps from [0, 1] into itself defined by

$$w_1(x) = \beta x \qquad \text{and} \qquad w_2(x) = \beta x + (1 - \beta),$$

where $0 < \beta < 1$ is a constant parameter. Consider the Markov chain with transfer operator $T : C([0, 1]) \to C([0, 1])$ defined by

$$T f(x) = p(x) f(w_1(x)) + (1 - p(x)) f(w_2(x)), \qquad f \in C([0, 1]),$$

where $p : [0, 1] \to (0, 1)$, is a continuous function with modulus of uniform continuity $\Delta = \Delta_p$.

The case when $p(x) \equiv \frac{1}{2}$ and $\beta = \frac{1}{2}$, where the uniform distribution on [0, 1] is the unique stationary distribution, and the case when $p(x) \equiv \frac{1}{2}$ and $\beta = \frac{1}{3}$, where the uniform distribution on the (middle third) Cantor set is the unique stationary distribution, are two important particular cases of this model.

For general p, Markov chains of this form always possess a stationary probability distribution, but they may possess more than one stationary probability distribution (see [26]).

From theorem 1 it follows that the distribution will be unique (for any fixed value of the parameter β) provided (12) holds, and this theorem also enables us to quantify the rate of convergence as a function of the modulus of uniform continuity of p. It also follows that this Markov chain will obey the functional CLT (6) and (7) for Hölder-continuous functions f with exponent $\alpha > \frac{1}{2}$ provided p is Dini-continuous. Observe that our conditions are only sufficient. It is an interesting open problem to try to find critical smoothness properties of p to ensure a unique stationary measure and a CLT.

Remark 7. If $X = \{1, \ldots, N\}^{\mathbb{N}}$ and for two elements $x = x_0 x_1 \ldots$ and $y = y_0 y_1 \ldots$ in X, we define $d(x, y) := 2^{-\min(k \geqslant 0;\, x_k \neq y_k)}$ if $x \neq y$, and $d(x, y) := 0$ if $x = y$, then (X, d) is a compact metric space. Let g be a continuous function from X to (0, 1], such that $\sum_{x_0=1}^{N} g(x_0 x_1 \ldots) = 1$ for all $x_1 x_2 \ldots \in X$. g describes an IFS with place-dependent probabilities: $\{(X, d), w_i(x), p_i(x), i \in \{1, \ldots, N\}\}$, where $w_i(x) = ix$ and $p_i(x) = g(ix)$, and corollary 2 applies. This generalizes theorem 1 in [28] and also implies a CLT for the associated Markov chains under the 'summable variations' condition used in [8] or [29]. Stationary probability measures for such Markov chains are sometimes called g-measures, a concept coined by Keane [18].

Acknowledgments

We are grateful to the referees for pointing out useful references and for providing us with comments for improving this paper.

References

[1] Barnsley M F and Demko S 1985 Iterated function systems and the global construction of fractals *Proc. R. Soc. Lond.* A **399** 243–75

[2] Barnsley M F, Demko S G, Elton J H and Geronimo J S 1988 Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities *Ann. Inst. H Poincaré Probab. Stat.* **24** 367–94

Barnsley M F, Demko S G, Elton J H and Geronimo J S 1989 *Ann. Inst. H Poincaré Probab. Stat.* **25** 589–90 (erratum)

[3] Benda M 1998 A central limit theorem for contractive stochastic dynamical systems *J. Appl. Probab.* **35** 200–5

[4] Berger N, Hoffman C and Sidoravicius V 2003 Nonuniqueness for specifications in $\ell^{2+\epsilon}$ *Preprint* math PR/0312344

[5] Billingsley P 1968 *Convergence of Probability Measures* (New York: Wiley)

[6] Bramson M and Kalikow S 1993 Non-uniqueness in g-functions *Israel J. Math.* **84** 153–60

[7] Derriennic Y and Lin M 2003 The central limit theorem for Markov chains started at a point *Probab. Theory Rel. Fields* **125** 73–6

[8] Doeblin W and Fortet R 1937 Sur les chaines à liaisons complètes *Bull. Soc. Math. France* **65** 132–48

[9] Dudley R M 1989 *Real Analysis and Probability* (Pacific Grove, CA: Wadsworth and Brooks/Cole)

[10] Fan A H and Lau K-S 1999 Iterated function system and Ruelle operator *J. Math. Anal. Appl.* **231** 319–44

[11] Harris T E 1955 On chains of infinite order *Pac. J. Math.* **5** 707–24

[12] Hutchinson J E 1981 Fractals and self-similarity *Indiana Univ. Math. J.* **30** 713–47

[13] Iosifescu M and Grigorescu S 1990 *Dependence with Complete Connections and its Applications* (Cambridge: Cambridge University Press)

[14] Johansson A and Öberg A 2003 Square summability of variations of g-functions and uniqueness of g-measures *Math. Res. Lett.* **10** 587–601

[15] Kaijser T 1973 Some limit theorems for Markov chains with applications to learning models and products of random matrices *PhD Thesis* Institut Mittag-Leffler, Djursholm, Sweden

[16] Kaijser T 1979 Another central limit theorem for random systems with complete connections *Rev. Roum. Math. Pures Appl.* **24** 383–412

[17] Kaijser T 1981 On a new contraction condition for random systems with complete connections *Rev. Roum. Math. Pures Appl.* **26** 1075–117

[18] Keane M 1972 Strongly mixing g-measures *Invent. Math.* **16** 309–24

[19] Kifer Y 1986 *Ergodic Theory of Random Transformations* (Boston, MA: Birkhäuser)

[20] Lacroix Y 2000 A note on weak-$\star$ perturbations of g-measures *Sankhyā* A **62** 331–8

[21] Maxwell M and Woodroofe M 2000 Central limit theorems for additive functionals of Markov chains *Ann. Probab.* **28** 713–24

[22] Norman M F 1972 *Markov Processes and Learning Models* (New York: Academic)

[23] Peligrad M and Utev S 2005 A new maximal inequality and invariance principle for stationary sequences *Ann. Probab.* **33** 798–815

[24] Prabhu N U 1965 *Stochastic Processes. Basic Theory and its Applications* (New York: Macmillan)

[25] Shiryaev A N 1996 *Probability* 2nd edn (New York: Springer)

[26] Stenflo Ö 2001 A note on a theorem of Karlin *Stat. Probab. Lett.* **54** 183–7

[27] Stenflo Ö 2002 Uniqueness of invariant measures for place-dependent random iterations of functions *IMA Vol. Math. Appl.* **132** 13–32

[28] Stenflo Ö 2003 Uniqueness in g-measures *Nonlinearity* **16** 403–10

[29] Walters P 1975 Ruelle's operator theorem and g-measures *Trans. Am. Math. Soc.* **214** 375–87

[30] Wu W B and Shao X 2004 Limit theorems for iterated random functions *J. Appl. Probab.* **41** 425–36

[31] Wu W B and Woodroofe M 2000 A central limit theorem for iterated random functions *J. Appl. Probab.* **37** 748–55

Applied Probability Trust (15 September 2005)

A RENEWAL PROCESS TYPE EXPRESSION FOR THE MOMENTS OF INVERSE SUBORDINATORS

ANDREAS NORDVALL LAGERÅS,* *Stockholm University*

Abstract

We define an inverse subordinator as the passage times of a subordinator to increasing levels. It has previously been noted that such processes have many similarities with renewal processes. Here we present an expression for the joint moments of the increments of an inverse subordinator. This is an analogue of a result for renewal processes. The main tool is a theorem on the processes which are both renewal processes and Cox processes.

Keywords: Subordinator; Passage time; Renewal theory; Cox process; Local time

AMS 2000 Subject Classification: Primary 60K05

Secondary 60G51; 60G55; 60E07

1. Introduction

Subordinators are non-decreasing processes with independent and stationary increments. The corresponding processes in discrete time are the partial-sum processes with positive, independent and identically distributed summands. Renewal processes can be considered to be passage times of partial-sum processes to increasing levels. Analogously we can define a process by the passage times of a subordinator. We call such a process an inverse subordinator.

The inverse subordinators appear in diverse areas of probability theory: As Bertoin [2] notes, the local times of a large class of well-behaved Markov processes are really inverse subordinators, and any inverse subordinator is the local time of some Markov process. It is well known, see Karatzas and Shreve [9], that the local time of the Brownian motion is the inverse of a 1/2-stable subordinator. Inverses of α-stable

* Postal address: Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden
Email address: andreas@math.su.se

subordinators with $0 < \alpha < 1$ arise as limiting processes of occupation times of Markov processes, see Bingham [4]. Some recent applications of inverse subordinators in stochastic models can be found in [8], [11] and [14]. Kaj and Martin-Löf [8] consider superposition and scaling of inverse subordinators with applications in queueing theory, Kozlova and Salminen [11] uses diffusion local time as input in a so-called storage process and Winkel [14] uses inverse subordinators in financial modeling.

In this paper we study some general distributional properties of inverse subordinators, using renewal theory and some theory about Cox processes. In particular we find an expression for the joint moments of their increments. Other results for inverse subordinators analogous to those in renewal theory has been proved by Bertoin, van Harn and Steutel, see [3] and [7].

Some well-known results on subordinators and infinitely divisible distributions on the positive real line are given in section 2 of this paper. Section 3 introduces the inverse subordinators and hints that they may have properties similar to the renewal processes. In section 4 the main result is given: An expression for the joint moments of the increments of an inverse subordinator. This is proved using a representation of the class of point processes that are both Cox processes and renewal processes. With this representation one can also give an alternative proof of the fact that inverse subordinators can be delayed to be given stationary increments, see [7]. We also provide a bound of the upper tail of the marginal distribution of an inverse subordinator. Finally, section 5 exemplifies the results with three types of inverse subordinators.

2. Some basic facts about subordinators

The following results on infinitely divisible distributions and Lévy processes can be found in [13]. Let $\{Y_t\}$ be a Lévy process, i.e. a stochastic process in continuous time with $Y_0 = 0$ and stationary and independent increments. The distribution F of Y_1 is necessarily infinitely divisible, i.e. for all $n \in \mathbb{N}$ there is a distribution F_n such that $F_n^{\star n} = F$. Here $F_n^{\star n}$ is the n-fold convolution of F_n. The converse is also true: Given an infinitely divisible distribution F there is a Lévy process $\{Y_t\}$ such that the distribution of Y_1 is F. Define $F^{\star t}$ for positive, non-integer t by $F^{\star t}(x) = P(Y_t \leq x)$.

One recognizes that $F_n = F^{\star 1/n}$.

If one restricts F to be a distribution on $\mathbb{R}_+$ then the increments of $\{Y_t\}$ are all non-negative. Lévy processes with non-negative increments are called subordinators. It is well known that the Laplace-Stieltjes transform of $F^{\star t}$, where F is an infinitely divisible distribution on $\mathbb{R}_+$, can be written

$$\widehat{F^{\star t}}(u) = \int_0^\infty e^{-ux} F^{\star t}(dx) = e^{-t\psi(u)} = \widehat{F}(u)^t,$$

where $\psi(u)$ is called the Lévy exponent. It can be written in the following form

$$\psi(u) = \delta u + \int_0^\infty (1 - e^{-ux})\nu(dx),$$

where $\delta \geq 0$ is called the drift and $\nu(dx)$ is called the Lévy measure. If Y_1 has drift δ then $Y_1 - \delta$ has drift 0. If $\int_0^\infty \nu(dx) < \infty$ then $\{Y_t\}$ is a compound Poisson process, with drift if $\delta > 0$, and thus only makes a finite number of jumps in any finite interval. We call a function π the Lévy density if $\nu(A) = \int_A \pi(x)dx$. If we define $\mu = E[Y_1]$, then $\mu = \delta + \int_0^\infty x\nu(dx)$. Since $\psi'(u) = \delta + \int_0^\infty e^{-ux} x\nu(dx)$, we have

$$\psi'(0) = \mu \text{ and } \psi'(u) \searrow \delta \text{ as } u \nearrow \infty. \tag{1}$$

Some parts of the reasoning in the following sections do not apply to compound Poisson processes without drift. Therefore we will henceforth, albeit somewhat artificially, exclude the compound Poisson processes without drift when referring to subordinators.

3. Inverse subordinators and renewal processes

It is advantageous to recall some results on renewal processes before a more thorough study of subordinators and their inverses. Let $X_2, X_3, \ldots$ be a sequence of independent and identically distributed (strictly) positive random variables with distribution F, and X_1 a positive random variable with distribution H, independent of $X_2, X_3, \ldots$. Let $S_0 = 0$ and $S_n = \sum_{k=1}^n X_k$, and we call $\{S_n\}$ a partial-sum process. Given a partial-sum process we define the renewal process with interarrival distribution F by $N_t = \min(n \in \mathbb{N} : S_n > t) - 1$. The -1 in the definition comes from the fact that we do not want to count the renewal at the origin, as is sometimes done. If $F = H$ then $\{N_t\}$ is called an ordinary renewal process.

It is well known that $\{N_t\}$ has stationary increments if and only if $H(x) = \frac{1}{\mu}\int_0^x (1 - F(y))dy$, where $\mu = E[X_2] = \int_0^\infty (1 - F(x))dx$, and μ necessarily is finite, see [5]. Then one also has

$$E[X_1] = \frac{E[X_2^2]}{2\mu}, \tag{2}$$

and the Laplace-Stieltjes transform of H is

$$\widehat{H}(s) = \frac{1}{\mu s}(1 - \widehat{F}(s)). \tag{3}$$

We note, as in [3], that subordinators are continuous time analogues of partial-sum processes. A Lévy process sampled at equidistant time points does produce a partial-sum process with infinitely divisible F, e.g. $Y_n = \sum_{k=1}^n (Y_k - Y_{k-1})$, when the time points are the integers. As the renewal processes are integer valued inverses to partial-sum processes, an inverse of a subordinator could be expected to have some properties similar to renewal processes. Given a subordinator $\{Y_t\}$, we define $\tau_t = \inf(\tau > 0 : Y_\tau > t)$, and call the process $\{\tau_t\}_{t\geq 0}$ the inverse subordinator.

The properties of the paths of $\{\tau_t\}$ differ depending on $\{Y_t\}$. Let us first consider a compound Poisson process $\{Y_t\}$ with drift $\delta > 0$. Since a jump in $\{Y_t\}$ corresponds to a flat period in its inverse, $\{\tau_t\}$ alternates between linear increasing with slope $\frac{1}{\delta}$ for exponential periods of time and being constant for periods of time with lengths drawn from the compounding distribution, with all these periods having independent lengths. It is more tricky when $\{Y_t\}$ is not compound Poisson and the drift is zero. Due to the fact that $\{Y_t\}$ in this case makes an infinite number of jumps in any finite interval, the trajectories of $\{\tau_t\}$ are continuous singular almost surely.

Now we will show that $\{\tau_t\}$ can be arbitrarily closely approximated by a scaled renewal process. For any $c > 0$, let $\{Y_t^c\}$ be defined by $Y_t^c = Y_{t/c}$. Note that $\{Y_t^c\}$ is a subordinator with $Y_1^c \sim F^{\star 1/c}$. Also define the renewal process $N_t^c = \min(n \in \mathbb{N} :$

$Y_n^c > t) - 1$. Since

$$\begin{aligned} c\tau_t &= c\inf(\tau > 0 : Y_\tau > t) \\ &= \inf(c\tau > 0 : Y_\tau > t) \\ &= \inf(\tau > 0 : Y_{\tau/c} > t) \\ &= \inf(\tau > 0 : Y_\tau^c > t) \\ &\geq \min(n \in \mathbb{N} : Y_n^c > t) - 1 \\ &\geq \inf(\tau > 0 : Y_\tau^c > t) - 1 \\ &= c\tau_t - 1, \end{aligned}$$

$$\tau_t = \tfrac{1}{c} N_t^c + r_t, \text{ where } 0 \leq r_t \leq \tfrac{1}{c},$$

and the approximation becomes arbitrarily good as $c \to \infty$. This result suggests that the inverse subordinators may have some properties similar to renewal processes. That this is in fact true will be shown in the following section.

An important function in the theory of renewal processes is the so called renewal function $V(t) = E[N_t]$. We note that for an ordinary renewal process $V(t) = \sum_{k=1}^\infty F^{\star k}(t)$, and for a stationary renewal process $V(t) = \frac{t}{\mu}$. If there is a function v such that $V(t) = \int_0^t v(s)ds$, then v is called the renewal density. If the renewal process would have been defined to also count the renewal at the origin, then the renewal function would be $V(t)+1$. One can also define a renewal function for the inverse subordinator. Given an inverse subordinator $\{\tau_t\}$, we define its renewal function U by $U(t) = E[\tau_t]$. The renewal function can be expressed as follows:

$$U(t) = E[\tau_t] = \int_0^\infty P(\tau_t > x)dx = \int_0^\infty P(Y_x \leq t)dx = \int_0^\infty F^{\star x}(t)dx$$

The expression on the right hand side might be hard to evaluate, but its Laplace-Stieltjes transform is easily calculated:

$$\begin{aligned} \widehat{U}(s) &= \int_0^\infty e^{-st} \int_0^\infty F^{\star x}(dt)dx = \int_0^\infty \widehat{F}(s)^x dx \\ &= \int_0^\infty e^{-x\psi(s)}dx = \frac{1}{\psi(s)} \end{aligned} \tag{4}$$

Thus there is a one-to-one correspondence between the renewal function and the distribution of $\{\tau_t\}$. This also correlates with the similar result for ordinary renewal processes and their renewal functions. Define the factorial power $n^{[k]}$ for $n, k \in \mathbb{N}$ by:

$$n^{[k]} = \begin{cases} n(n-1)\cdots(n-k+1) & \text{for } n \geq k \geq 1 \\ 1 & \text{for } k = 0 \\ 0 & \text{for } n < k, k \geq 1. \end{cases}$$

Given a renewal process and its renewal function, moments of all orders can be calculated as stated in the following proposition, see [5].

Proposition 1. *Let $\{N_t\}$ be a renewal process with interarrival distribution F and let $V(t) = \sum_{k=1}^{\infty} F^{\star k}(t)$. If $\{N_t\}$ is an ordinary renewal process then, for $0 \leq s_1 < t_1 \leq s_2 < \cdots < t_n$ and $k_1, \ldots, k_n \in \mathbb{N} \setminus \{0\}$ such that $k_1 + \cdots + k_n = k$,*

$$E\Big[\prod_{i=1}^{n} (N_{t_i} - N_{s_i})^{[k_i]}\Big] = \prod_{i=1}^{n} k_i! \cdot \int\limits_C \prod_{j=1}^{k} V(dx_j - x_{j-1}), \tag{5}$$

where $C = \{x_0, \ldots, x_k; x_0 = 0, s_i < x_{k_0+\cdots+k_{i-1}+1} < \cdots < x_{k_0+\cdots+k_i} \leq t_i, i = 1, \ldots, n, k_0 = 0\}$. If $\{N_t\}$ is stationary, then the proposition also holds with the first factor of the rightmost product in equation (5) replaced by $\frac{dx_1}{\mu}$.

A sketch of a proof: We can write $N_t - N_s = \int_{(s,t]} N(dx)$ and

$$(N_t - N_s)^k = \int_{(s,t]^k} \prod_{j=1}^{k} N(dx_j).$$

Note that n^k is the number of k-tuples of integers from 1 to n, and $n^{[k]}$ is the number of k-tuples of integers such that no integers in the k-tuple are the same. Thus we can write

$$(N_t - N_s)^{[k]} = \int_A \prod_{j=1}^{k} N(dx_j),$$

where $A = \{(x_1 \ldots x_k) \in (s,t]^k; x_p \neq x_q \text{ for } p \neq q\}$. The renewal property is used in the following:

$$\begin{aligned} E\Big[\prod_{j=1}^{k} N(dx_j)\Big] &= P(N(dx_1) = 1, \ldots, N(dx_k) = 1) \\ &= P(N(dx_{(1)}) = 1) \prod_{j=2}^{k} P(N(dx_{(j)}) = 1 | N(dx_{(j-1)}) = 1) \\ &= P(N(dx_{(1)}) = 1) \prod_{j=2}^{k} V(dx_{(j)} - x_{(j-1)}), \end{aligned}$$

and the first factor equals $V(dx_{(1)})$ and $\frac{dx_{(1)}}{\mu}$ in the ordinary and stationary case, respectively. Let $A_i = \{(y_{i1}, \ldots, y_{ik_i}) \in (s_i, t_i]^{k_i}; y_{ip} \neq y_{iq} \text{ for } p \neq q\}$ and $B_i = \{(y_{i1}, \ldots, y_{ik_i}); s_i < y_{i1} < \cdots < y_{ik_i} \leq t_i\}$. Thus, in the ordinary case,

$$
\begin{aligned}
E\Big[\prod_{i=1}^{n}(N_{t_i} - N_{s_i})^{[k_i]}\Big] &= E\Big[\prod_{i=1}^{n}\int_{A_i}\prod_{j=1}^{k_i} N(dy_{ij})\Big] \\
&= E\Big[\prod_{i=1}^{n} k_i! \int_{B_i}\prod_{j=1}^{k_i} N(dy_{ij})\Big] \\
&= \prod_{i=1}^{n} k_i! \cdot E\Big[\int_{C}\prod_{l=1}^{k} N(dx_l)\Big] \\
&= \prod_{i=1}^{n} k_i! \cdot \int_{C}\prod_{l=1}^{k} V(dx_l - x_{l-1}).
\end{aligned}
$$

4. Inverse subordinators and Cox processes

An expression similar to (5) for the moments of $\{\tau_t\}$ can be obtained. First recall the definition of a Cox process. Let $\{N_t^\lambda\}$ be an inhomogeneous Poisson process on $\mathbb{R}_+$ with intensity measure λ. Let Λ be a random measure on $\mathbb{R}_+$. If the point process $\{M_t\}$ has the distribution of $\{N_t^\lambda\}$ conditional on $\Lambda = \lambda$, then $\{M_t\}$ is called a Cox process directed by Λ. Note that if $\{\widetilde{N}_t\}$ is a Poisson process with constant intensity equal to one and independent of Λ, then $M_t \stackrel{d}{=} \widetilde{N}(\Lambda((0,t]))$, and $\{M_t\}$ can be considered to be a homogeneous Poisson process subjected to a random time change by the random function $\Lambda((0,t])$. The interpretation of the Cox process as a time changed Poisson process also describes how the points of the Cox process can be obtained from the points of the Poisson process: If we let $K(t)$ be the inverse of $\Lambda((0,t])$ and $t_1, t_2, \ldots$ are the points of $\{\widetilde{N}_t\}$, then $K(t_1), K(t_2), \ldots$ are the points of $\{M_t\}$.

Also define a slight generalization of the inverse subordinators: Let $\widetilde{Y}_0$ have the distribution G on $\mathbb{R}_+$ and be independent of the subordinator $\{Y_t\}$ with $Y_1 \sim F$. Define the process $\{\widetilde{Y}_t\}$ by $\widetilde{Y}_t = Y_t + \widetilde{Y}_0$. Let $\tau_t = \inf(\tau > 0 : \widetilde{Y}_\tau > t)$, and call the process $\{\tau_t\}_{t\geq 0}$ a general inverse subordinator. If $\widetilde{Y}_0 \equiv 0$ then we call $\{\tau_t\}$ an ordinary inverse subordinator.

We will see in Proposition 4 that $\widetilde{Y}_0$ can be chosen so that the general inverse subordinator $\{\tau_t\}$ has stationary increments, if $\mu = E[Y_1] < \infty$. The following proposition is by Kingman [10] and Grandell [6].

Proposition 2. *The Cox process $\{M_t\}$ directed by Λ is a renewal process if and only if $\Lambda((s,t]) = \tau_t - \tau_s$ for all $t > s$, where $\{\tau_t\}$ is a general inverse subordinator.*

We will only prove the easier if-part of the proposition. Only that part will be used in theorem 1.

Proof. We note that $\widetilde{Y}_t$ is the inverse of $\Lambda((0,t]) = \tau_t$. If we use the representation of $\{M_t\}$ as a time changed Poisson process $\{\widetilde{N}_t\}$ with intensity one, then the points of $\{M_t\}$ are $\widetilde{Y}(t_1), \widetilde{Y}(t_2), \dots$, where $t_1, t_2, \dots$ are the points of $\{\widetilde{N}_t\}$. Since $\{Y_t\}$ is a subordinator, $\widetilde{Y}(t_1), \widetilde{Y}(t_2) - \widetilde{Y}(t_1), \widetilde{Y}(t_3) - \widetilde{Y}(t_2), \dots$ are independent and $\widetilde{Y}(t_2) - \widetilde{Y}(t_1), \widetilde{Y}(t_3) - \widetilde{Y}(t_2), \dots$ are furthermore equally distributed. Thus $\{M_t\}$ is a renewal process.

We can say more about the interarrival distribution of $\{M_t\}$. Let $Z = \widetilde{Y}(t_2) - \widetilde{Y}(t_1) = Y(t_2) - Y(t_1)$, and let $\varepsilon \sim \mathrm{Exp}(1)$, independent of $\{Y_t\}$. $Z = Y(t_2) - Y(t_1) \stackrel{d}{=} Y(t_2 - t_1) \stackrel{d}{=} Y(\varepsilon)$, so the interarrival distribution of $\{M_t\}$ is thus compound exponential. The Laplace-Stieltjes transform of the distribution of Z is given by

$$\widehat{F}_Z(s) = E[e^{-sZ}] = E[E[e^{-sY(\varepsilon)}|\varepsilon]] = E[e^{-\psi(s)\varepsilon}] = \frac{1}{1+\psi(s)}, \tag{6}$$

where $\psi(s)$ is the Lévy exponent of Y_1. We now have the tools to prove the main result:

Theorem 1. *Let $\{\tau_t\}$ be an ordinary inverse subordinator with renewal function $U(t)$. Then, for $0 \le s_1 < t_1 \le s_2 < \cdots < t_n$ and $k_1, \dots, k_n \in \mathbb{N}\setminus\{0\}$ such that $k_1 + \cdots + k_n = k$,*

$$E\Big[\prod_{i=1}^{n}(\tau_{t_i} - \tau_{s_i})^{k_i}\Big] = \prod_{i=1}^{n} k_i! \cdot \int_C \prod_{j=1}^{k} U(dx_j - x_{j-1}) \tag{7}$$

where C is as in Proposition 1. If $\{\tau_t\}$ is stationary, then the theorem also holds with the change that the first factor of the rightmost product in equation (7) is replaced by $\frac{dx_1}{\mu}$, but with the same U in the remaining factors as the ordinary inverse subordinator.

Proof. Define the random measure Λ on $\mathbb{R}_+$ by $\Lambda((s,t]) = \tau_t - \tau_s$ for all $t > s \in \mathbb{R}_+$, and let $\{M_t\}$ be the Cox process directed by Λ. By Proposition 2, $\{M_t\}$ is also a renewal process. Write $V(t)$ for its renewal function. Then

$$V(t) = E[M_t] = E[E[M_t|\tau_t]] = E[\tau_t] = U(t). \tag{8}$$

Thus one can replace $V(t)$ by $U(t)$ in (5) when calculating the factorial moments of $\{M_t\}$. As noted in [5], the factorial moments of the Cox process coincide with the ordinary moments of its directing measure, and by the construction of the directing measure the stated result follows.

A renewal theorem for the inverse subordinators can also be given following Bertoin [2], Theorem I.21.

Proposition 3. *If $\mu < \infty$, then $U(t) \sim \frac{t}{\mu}$ as $t \to \infty$.*

Proof. Let $\{M_t\}$ be a Cox process directed by $\{\tau_t\}$ as in Proposition 2, and $V(t)$ its renewal function. By (8), $V(t) = U(t)$. An application of the renewal theorem for renewal processes, see [5], provides the desired result.

Similar to renewal processes, the inverse subordinators can be delayed to become stationary. This has been proved by different methods in [7] and [8]. We state the result and provide a proof based on the connection with Cox processes.

Proposition 4. *Let $\{\tau_t\}$ be a general inverse subordinator with $\widetilde{Y}_0 \sim G$ and $Y_1 = \widetilde{Y}_1 - \widetilde{Y}_0 \sim F$ and $\mu = E[Y_1] < \infty$, where*

$$\psi(s) = -\log \widehat{F}(s) = \delta s + \int_0^\infty (1 - e^{-sx})\nu(dx) \text{ and}$$
$$G(x) = \begin{cases} \frac{1}{\mu}\left(\delta + \int_0^x \int_y^\infty \nu(dz)dy\right) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases} \tag{9}$$

Then $\{\tau_t\}$ has stationary increments.

Proof. By Theorem 1.4 in [6], a Cox process is stationary if and only if its directing measure Λ has stationary increments. Therefore it suffices to check that the Cox process $\{M_t\}$ directed by $\{\tau_t\}$ is stationary. Its interarrival distribution is F_Z given by (6). The X_1 of $\{M_t\}$ can be decomposed into $X_1 \stackrel{d}{=} \widetilde{Y}_0 + Z$, with $\widetilde{Y}_0$ and Z independent, since the inverse subordinator is delayed a time $\widetilde{Y}_0$ during which it is constant equal to 0. The Laplace-Stieltjes transform of the distribution H of X_1 is $\widehat{H}(s) = \widehat{G}(s)\widehat{F}_Z(s)$, where

$$\begin{aligned}\widehat{G}(s) &= \frac{1}{\mu}\int_0^\infty e^{-sx}\left(\delta + \int_x^\infty \nu(dy)\right)dx = \frac{1}{\mu}\left(\frac{\delta}{s} + \int_0^\infty\int_0^y e^{-sx}dx\nu(dy)\right)\\ &= \frac{1}{\mu s}\left(\delta + \int_0^\infty (1-e^{-sy})\nu(dy)\right) = \frac{\psi(s)}{\mu s}. \end{aligned} \tag{10}$$

Combining (6) and (10), we get

$$\widehat{H}(s) = \widehat{G}(s)\widehat{F}_Z(s) = \frac{\psi(s)}{\mu s}\frac{1}{1+\psi(s)} = \frac{1}{\mu s}(1-\widehat{F}_Z(s)).$$

By (3), X_1 thus has the right distribution to make $\{M_t\}$ stationary.

Let W_t be the excess of the renewal process and Cox process $\{M_t\}$, i.e. the time from t to the next point of the process. When $\{M_t\}$ is stationary, $W_t \stackrel{d}{=} X_1 \stackrel{d}{=} \widetilde{Y}_0 + Z$. The decomposition of the excess can be given the following interpretation: From any given time t the inverse subordinator will remain constant a period which has the distribution G. During this time no points in the Cox process will occur. After that time the inverse subordinator starts anew and the distribution to the next point in the point process is given by F_Z. In the stationary case, we do not have to know G explicitly to calculate $E[\widetilde{Y}_0]$, if we use (2): $E[X_1] = \frac{E[Z^2]}{2E[Z]}$. $E[X_1] = E[\widetilde{Y}_0] + EZ$, and by straightforward calculation, using e.g. (6), $E[Z] = E[Y_1]$ and $E[Z^2] = \mathrm{Var}(Y_1) + 2E[Y_1]^2$. Collecting and rearranging yields $E[\widetilde{Y}_0] = \frac{\mathrm{Var}(Y_1)}{2EY_1}$.

The expression (7) may be hard to use in practice to calculate higher joint moments. Nonetheless the results above show that the covariance of two increments of a stationary inverse subordinator is a simple expression in the renewal function. Let $\{\tau_t\}$ be stationary and let $U(t)$ denote the renewal function of the corresponding ordinary inverse subordinator. Also let $0 < r \le s < t$.

$$\begin{aligned}\mathrm{Cov}(\tau_r, \tau_t - \tau_s) &= E[\tau_r(\tau_t-\tau_s)] - E[\tau_r]E[\tau_t-\tau_s]\\ &= \int_0^r\int_s^t U(dx-y)\frac{dy}{\mu} - \frac{r}{\mu}\frac{t-s}{\mu}\\ &= \frac{1}{\mu}\int_0^r (U(t-y)-U(s-y))dy - \frac{r(t-s)}{\mu^2}.\end{aligned}$$

Now consider the particular case where $r = 1, s = n \geq 1$ and $t = n + 1$ and U has a density u, such that $U(t) = \int_0^t u(s)ds$. Also assume, for simplicity, that $\mu = 1$. Then the following approximation can be done:

$$\mathrm{Cov}(\tau_1, \tau_{n+1} - \tau_n) = \int_0^1 (U(n+1-y) - U(n-y))dy - 1 \approx u(n) - 1.$$

Given the distribution of the subordinator $\{Y_t\}$, the distribution of its inverse is given by $P(\tau_t \leq x) = P(Y_x > t)$. It may still be hard to find a closed form expression of this distribution function. The tail probabilities for the ordinary inverse subordinator can nonetheless be estimated. Only the case $\delta = 0$ is interesting since if the drift δ is positive then $\{Y_x - \delta x\}$ is non-negative and thus $P(Y_x \leq t) = P(Y_x - \delta x \leq t - \delta x) = 0$ for $x > \frac{t}{\delta}$. Let $s \geq 0$. Then we have that

$$P(\tau_t > x) = P(Y_x \leq t) = P(e^{-sY_x} \geq e^{-st}) \leq \frac{E\left[e^{-sY_x}\right]}{e^{-st}} = e^{st - x\psi(s)}.$$

By (1) the last expression has unique minimum as a function of s. If x is large enough ($x > \frac{t}{\mu}$), the s that minimizes the expression is non-zero and given by $s = \psi'^{-1}(\frac{t}{x})$, where ψ'^{-1} is the inverse of ψ'. Thus, for large enough x,

$$P(\tau_t > x) \leq \exp\left(t\psi'^{-1}(\tfrac{t}{x}) - x\psi(\psi'^{-1}(\tfrac{t}{x}))\right). \tag{11}$$

There is another result on the marginal distribution of $\{\tau_t\}$ that deserves mentioning. This result can be found in [8] and [12], but we give a short proof based on identifying Laplace transforms as probabilities.

Proposition 5. *Let ε_s be exponentially distributed with mean $\frac{1}{s}$ and independent of $\{\tau_t\}$. Then the Laplace-Stieltjes transform of the distribution of $\tau(\varepsilon_s)$ is given by:*

$$E[e^{-u\tau(\varepsilon_s)}] = 1 - \frac{u\widehat{G}(s)}{u + \psi(s)},$$

where $\widehat{G}(s)$ is the Laplace-Stieltjes transform of the distribution of $\widetilde{Y}_0$.

Proof. Let $\widetilde{\varepsilon}_u$ be exponentially distributed with mean $\frac{1}{u}$ and independent of ε_s and $\{\tau_t\}$. We note that for a non-negative random variable X independent of $\widetilde{\varepsilon}_u$,

$P(\widetilde{\varepsilon}_u \geq X) = E[P(\widetilde{\varepsilon}_u \geq X|X)] = E[e^{-uX}]$, the Laplace-Stieltjes transform of the distribution of X. Thus we have, in the ordinary case,

$$\begin{aligned} E[e^{-u\tau(\varepsilon_s)}] &= P(\widetilde{\varepsilon}_u \geq \tau(\varepsilon_s)) = P(Y(\widetilde{\varepsilon}_u) > \varepsilon_s) = 1 - P(\varepsilon_s \geq Y(\widetilde{\varepsilon}_u)) \\ &= 1 - E[e^{-sY(\widetilde{\varepsilon}_u)}] = 1 - E[E[e^{-sY(\widetilde{\varepsilon}_u)}|\widetilde{\varepsilon}_u]] = 1 - E[e^{-\psi(s)\widetilde{\varepsilon}_u}] \\ &= 1 - \frac{u}{u+\psi(s)}. \end{aligned}$$

Likewise, in the general case,

$$\begin{aligned} E[e^{-u\tau(\varepsilon_s)}] &= 1 - P(\varepsilon_s \geq \widetilde{Y}(\widetilde{\varepsilon}_u)) \\ &= 1 - P(\varepsilon_s \geq \widetilde{Y}(\widetilde{\varepsilon}_u)|\varepsilon_s \geq \widetilde{Y}_0)P(\varepsilon_s \geq \widetilde{Y}_0) \\ &= 1 - P(\varepsilon_s \geq Y(\widetilde{\varepsilon}_u))P(\varepsilon_s \geq \widetilde{Y}_0) \\ &= 1 - \frac{u\widehat{G}(s)}{u+\psi(s)}, \end{aligned}$$

where we have used the memorylessness of the exponential distribution.

5. Examples

The α-stable distribution on $\mathbb{R}_+$ has Lévy exponent $\psi(s) = s^\alpha$ with $0 < \alpha < 1$. This gives a renewal density $u(t) = 1/(\Gamma(\alpha)t^{1-\alpha})$ for the corresponding inverse stable subordinator by inverting (4). Theorem 1 thus confirms the moment expressions in [4], e.g. equation (18).

The main obstacle to use Theorem 1 is the possible difficulties in finding an expression for the renewal function. It is possible to find the renewal density not only for the inverse stable subordinator, but also for the inverses of subordinators with inverse gaussian and gamma distributed increments. In these two cases it is also possible to delay the processes to obtain stationary versions, which is not possible in the stable case.

For the inverse gaussian distribution, with probability density

$$f(x) = \frac{\delta}{\sqrt{2\pi x^3}} \exp\left(\delta\gamma - \frac{1}{2}\left(\frac{\delta^2}{x} + \gamma^2 x\right)\right), \delta > 0, \gamma > 0,$$

and Lévy exponent and Lévy density, respectively,

$$\psi(s) = \delta\sqrt{\gamma^2 + 2s} - \delta\gamma$$
$$\pi(x) = \frac{\delta}{\sqrt{2\pi x^3}} \exp\left(-\frac{\gamma^2 x}{2}\right),$$

we get, by (9), a probability density of the delay $\widetilde{Y}_0$ by integrating π ($\mu = \psi'(0) = \frac{\delta}{\gamma}$)

$$g(t) = \frac{1}{\mu}\int_t^\infty \pi(x)dx = \gamma\sqrt{\frac{2}{\pi t}} \exp\left(-\frac{\gamma^2 t}{2}\right) - \gamma^2 \operatorname{erfc}\left(\gamma\sqrt{\frac{t}{2}}\right)$$

Here erfc is the complementary error function defined by $\operatorname{erfc}(t) = \frac{2}{\sqrt{\pi}}\int_t^\infty \exp(-s^2)ds$. We note that the density does not depend on the parameter δ. One obtains the renewal density $u(t)$ from its Laplace transform, which is equivalent to the Laplace-Stieltjes transform of $U(t)$, by rewriting (4):

$$\begin{aligned}
\widehat{U}(s) &= \frac{1}{\psi(s)} = \frac{1}{\delta\sqrt{\gamma^2+2s} - \delta\gamma} \\
&= \frac{\gamma}{2\delta s} + \frac{1}{\delta\sqrt{\gamma^2+2s}} + \frac{\gamma^2}{2\delta s\sqrt{\gamma^2+2s}} \\
&\Rightarrow \ \{\text{by [1] (29.3.1), (29.3.11) and (29.3.44)}\} \\
u(t) &= \frac{\gamma}{\delta} + \frac{1}{\delta\sqrt{2\pi t}} \exp\left(-\frac{\gamma^2 t}{2}\right) - \frac{\gamma}{2\delta}\operatorname{erfc}\left(\gamma\sqrt{\frac{t}{2}}\right)
\end{aligned}$$

The estimate (11) gives

$$P(\tau_t > x) \leq \exp\left(-\frac{\delta^2 x^2}{2t} + \delta\gamma x - \frac{\gamma^2 t}{2}\right).$$

For the gamma distribution we have probability density, Lévy exponent and Lévy density:

$$f(x) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}, \nu > 0, \alpha > 0$$
$$\psi(s) = \nu \log\left(1 + \frac{s}{\alpha}\right)$$
$$\pi(x) = \frac{\nu}{x} e^{-\alpha x}$$

so, by (9), the density of the delay is

$$g(t) = \alpha \mathrm{E}_1(\alpha t),$$

where E_1 the exponential integral defined by $\mathrm{E}_1(t) = \int_t^\infty \exp(-s)\frac{ds}{s}$. As in the inverse gaussian case the density only depends on one parameter. The renewal density is also in the gamma case most easily obtained by first rewriting (4):

$$\begin{aligned}
\widehat{U}(s) &= \frac{1}{\nu \log(1+\frac{s}{\alpha})} \\
&= \frac{\alpha}{\nu s}\int_0^1 \left(1+\frac{s}{\alpha}\right)^u du \\
&= \frac{\alpha}{\nu}\int_0^1 \left(\frac{1}{s}\frac{1}{(1+\frac{s}{\alpha})^{1-u}} + \frac{1}{\alpha}\frac{1}{(1+\frac{s}{\alpha})^{1-u}}\right) du \\
&\Rightarrow \{\text{by [1], (29.3.11), (29.2.6) and (6.5.2)}\} \\
u(t) &= \frac{\alpha}{\nu}\int_0^1 \frac{du}{\Gamma(u)}\left(\gamma(u,\alpha t) + (\alpha t)^{u-1}e^{-\alpha t}\right),
\end{aligned}$$

where $\gamma(u,t) = \int_0^t s^{u-1}e^{-s}ds$ is the incomplete gamma function. We also have a tail estimate by (11):

$$P(\tau_t > x) \le \exp\left(\nu x - \alpha t - x\nu \log\frac{\nu x}{\alpha t}\right) = \left(\frac{\alpha t}{\nu x}\right)^{\nu x} e^{\nu x - \alpha t}.$$

Acknowledgements

I am grateful to Jan Grandell for discussing Cox processes and renewal processes with me, and also to Thomas Höglund for careful reading of an earlier draft of this paper and pointing out some errors. I also thank the anonymous referee, whose suggestions helped to improve the article.

References

[1] Abramowitz, M. and Stegun, I. A., Eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications Inc., New York.

[2] Bertoin, J. (1996). *Lévy Processes.* Cambridge University Press, Cambridge.

[3] Bertoin, J. and van Harn, K. and Steutel, F. W. (1999). Renewal theory and level passage by subordinators. *Statist. Probab. Lett.* **45,** 65–69.

[4] Bingham, N. H. (1971). Limit theorems for occupation times of Markov processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **17,** 1–22.

[5] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*, vol. 1, 2:nd ed. Springer-Verlag, New York.

[6] Grandell, J. (1976). *Doubly Stochastic Poisson Processes.* Lecture Notes in Mathematics **529**. Springer-Verlag, Berlin.

[7] van Harn, K. and Steutel, F. W. (2001). Stationarity of delayed subordinators. *Stoch. Models* **17,** 369–374.

[8] Kaj, I. and Martin-Löf, A. (2004). Scaling limit results for the sum of many inverse Lévy subordinators. Preprint 2004:13. Department of Mathematics, Uppsala University.

[9] Karatzas, I. and Shreve, S. E. (1988). *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

[10] Kingman, J. F. C. (1964). On the doubly stochastic Poisson processes. *Proc. Camb. Phil. Soc.* **60,** 923–930.

[11] Kozlova, M. and Salminen, P. (2004). Diffusion local time storage. *Stoch. Proc. Appl.* **114,** 211–229.

[12] Sato, K.-i. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge.

[13] Steutel, F. W. and van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line.* Marcel Dekker Inc., New York.

[14] Winkel, M. (2005). Electronic foreign-exchange markets and passage events of independent subordinators. *J. Appl. Prob.* **42,** 138–152