Budapest University of Technology and Economics
Department of Telecommunications and Media Informatics

Characterisation of Self-Similar Traffic in Data Networks

András Gefferth

Scientific supervisors
Dr. Sándor Molnár and Dr. Darryl N. Veitch

Budapest, 2005

Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

Önhasonló Hálózati Forgalom Matematikai Leírása

Gefferth András

Tudományos vezetők
Dr. Molnár Sándor és Dr. Darryl N. Veitch

Budapest, 2005

A disszertáció bírálatai, valamint a nyílvános védésről készült jegyzőkönyv a Dékáni Hivatalban tekinthetők meg.

The recension of this Ph.D. dissertation and the minutes of the Ph.D. discussion are available in the Dean's Office

Acknowledgement

I wish to thank my supervisor, Dr. Darryl Veitch, for his encouraging support and guidance. His enthusiasm for interesting mathematical details contributed to a very exhaustive development of the subject. I will never forget our long and fruitful consultations far into the night.

I wish to express particular thanks to my supervisor, Dr. Sándor Molnár, for his conscientious work and his practical approach in guiding me towards my PhD degree.

István Maricza was always happy to answer all my questions. I wish to thank him for his help and for his valuable comments on the dissertation.

I'm especially thankful to these people for their friendship, and also because I always felt that my PhD degree was at least as important for them as it was for me.

I would also like to thank those people and institutions that provided the environment for my research.

I'm very thankful to Tamás Henk and all members of the High Speed Networks Laboratory at the Department of Telecommunications and Media Informatics at the Budapest University of Technology and Economics.

I'm very thankful to the Australian Department of Training Education and Youth Affairs for making it possible for me to stay in Melbourne through its Australian-European Award Program. This funding has allowed me to study at the Royal Melbourne Institute of Technology and the University of Melbourne and work under the supervision of Dr. Darryl Veitch.

I'm very thankful to Darryl Veitch and all members of the EMUlab at the University of Melbourne for providing financial support and a very pleasant work environment, enabling me to focus on my research.

Abstract

In order to design telecommunication networks, network protocols and applications it is important to know the properties of the traffic intensity of the network. This intensity can be characterised using stochastic descriptors. Stochastic models for telephone networks have been developed in the beginning of the 20th century and by now well established methods exist for the dimensioning and planning of such networks.

It was found that the traffic intensity of computer networks differs significantly from that of telephone networks. The traffic of the former has very strong self-correlation resulting in a so-called long-range dependent and self-similar characteristics.

These processes have been used for a long time to describe different natural and man-made systems, however, it was found that the corresponding theory is still incomplete; some incorrect statements still appear in the literature, some questions are not answered, or not even raised and the definitions used by different authors are incompatible with each other.

This work contributes to the clarification of the theoretical background of these processes. It does so by stating and proving important statements, and by providing a new approach that enables simple but comprehensive treatment.

After presenting the basic terms and definitions in Chapter 1, Chapter 2 deals with discrete self-similarity. Here we compare the different definitions, clear up the common misconception that fractional Gaussian noise is the only self-similar process and give the whole set of self-similar processes.

Then we focus our attention on asymptotically self-similar processes and provide typical examples of these.

Chapter 3 deals with long-range dependent processes, which form an important subset of asymptotically self-similar processes. Again the different definitions are gathered, thoroughly compared and a new definition is provided. Among other results a fundamental theorem is proved about the behaviour of long-range dependent processes on different time scales.

It will also be shown that the class of asymptotically self-similar processes converging to a long-range dependent fractional Gaussian noise is wider than usually believed.

This research started with the investigation of practical questions from the field of computer networks. During this work it turned out that the shortcomings of the theoretical background hinder practical work. Therefore the most important contribution of this work is that it provides a solid, well-structured theoretical background that supports application-oriented research. Chapter 5 demonstrates with a few examples how the new approach simplifies the treatment of second-order scaling processes.

Although the most important benefits are presented in Chapter 5 some practical applications of the results are provided in Chapter 4, where we point out the hidden hazards of long-range dependence estimation.

Kivonat

Távközlőhálózatok, számítógép-hálózatok, távközlési protokollok, illetve alkalmazások tervezéséhez fontos a hálózati forgalom viselkedésének ismerete. Ezt a forgalmat sztochasztikus leírókkal jellemezhetjük. A klasszikus telefonhálózat forgalmát jellemző sztochasztikus modellek a huszadik század elején születtek és mára jól bevált módszerek léteznek az ilyen hálózatok méretezésére, tervezésére.

A számítógép-hálózatok forgalma azonban jelentősen különbözik a telefonhálózatokétól. Itt ugyanis a forgalomban igen erős korreláció van jelen, ami sokszor egy úgynevezett hosszútávon összefüggő és önhasonló folyamathoz vezet.

Ezek a folyamatok már régebb óta ismeretesek, különböző természetes és mesterséges rendszerekben fedeztek fel ilyeneket. Ennek ellenére azt találtam, hogy az idevonatkozó elméletben hibák, hiányosságok vannak. Az irodalomban fellelhetők tévedések, tisztázatlan, illetve meg nem fogalmazott kérdések, továbbá a különböző szerzők definíciói gyakran különböznek egymástól.

Jelen munka fontos tételek kimondásával, illetve bizonyításával, valamint egy újfajta, könnyebben kezelhető megközelítés bevezetésével az ilyen folyamatok elméletének tisztázásához járul hozzá.

Az első fejezetben a továbbiakban használt definíciókat és fogalmakat vezetjük be. Ezután a második fejezet foglalkozik a diszkrét idejű önhasonló folyamatokkal. Összehasonlítjuk a fellelhető definíciókat, és eloszlatjuk azt a többször előforduló tévhitet miszerint a frakcionális Gaussi zaj az egyetlen lehetséges önhasonló folyamat, egyben megadjuk az önhasonló folyamatok teljes halmazát. Ezután az aszimptotikusan önhasonló folyamatokra térünk át, és bemutatjuk ezek néhány tipikus példáját.

A 3. fejezetben a hosszútávon összefüggő folyamatokkal foglalkozunk, melyek az aszimptotikusan önhasonló folyamatok gyakorlati szempontból fontos részhalmazát képzik. Ezen folyamatoknak is áttekintjük, és összevetjük különböző definícióit, valamint egy újabb definíciót adunk. Több eredmény mellett megadjuk egy a hosszútávon összefüggő folyamatok viselkedését különböző időskálákon leíró fontos és ismert tétel részletes bizonyítását is. Itt mutatjuk be, hogy a frakcionális Gaussi zajhoz konvergáló aszimptotikusan önhasonló folyamatok osztálya tágabb mint azt az irodalomban általában feltételezik.

A kutatás eredeti céljaként számítógép-hálózatokban felmerülő gyakorlati problémákra kerestem megoldást. Eközben derült ki, hogy az elméletben meglevő hiányosságok a gyakorlati munkát hátráltatják. Így tehát jelen munkának, mely ezeket a hiányosságokat pótolja, legfőbb jelentősége abban áll, hogy gyakorlati, alkalmazott kutatásokhoz egy jól áttekinthető elméleti alapot nyújt. Az 5. fejezetben példát is mutatunk, olyan esetre ahol a jelen munkában kidolgozott megközelítés lényegesen leegyszerűsíti a folyamatok több időskálán való viselkedésének vizsgálatát.

Habár – mint említettük – a munka legfontosabb kimenetele az elmélet tisztázása, a 4. fejezetben az eredmények egyik közvetlen hasznosítását ismertetjük, a hosszútávon összefüggőségi tulajdonság becslésével kapcsolatban.

Contents

List of abbreviations and notations

Acronym	Symbol	Name	Page No.
Æ	$f \sim g$	asymptotic equivalence, $\lim f/g = 1$	37
ACF	ρ	autocorrelation function	16
ACVF	γ	autocovariance function	15
AFN		alternating fractional noise	47
ASS		asymptotically self-similar	24
CRV		continuous regular variation	40
CSF	s	correlation sum function	16
CSV		continuous slow variation	40
CTF	ϕ	correlation-time function	16
CVSF	S	covariance sum function	16
DoA		domain of attraction	27
DRV		discrete regular variation	41
DSV		discrete slow variation	41
	$\mathbf{E}[.]$	expected value	
FN		fractional noise	33
LRD		long-range dependent/dependence	43
PSD		positive semi-definite	17
SRD		short-range dependent/dependence	33
SS		self-similar	21
	$\mathcal{V}$	variance	15
VTF	ω	variance-time function	16
	RV_α	set of regular varying functions w. index α	40, 41
	$\mathbf{R}$ / $\mathbf{R}^+$	set of real/positive real numbers	
	$\mathbf{Z}$ / $\mathbf{Z}^+$	set of integer/positive integer numbers	

Introduction

Describing the traffic of telecommunication networks is important for different engineering tasks, such as dimensioning, admission control, designing routing protocols, etc., depending on the networking technology. Because of its continuously changing nature, the traffic can be best described using stochastic models. Such models for telephone networks have been first developed by Agner Krarup Erlang not long after the first few commercial networks have been installed. By his work he established the discipline of traffic theory [4].

He developed the famous "Erlang B" formula, which is still being used for the dimensioning of telephone networks. Suppose that the average call intensity, which is defined by the product of the average number of calls during an hour and the average length of calls, is known. Assume now that we would like to keep the call blocking probability below a given level. The Erlang B formula yields the minimum number of telephone lines that is needed to carry this amount of traffic with the given blocking probability. The requirements of this formula are the Poissonian process of the calls and the exponential call holding times.

The network traffic was found to be static in the sense that the behaviour of the "typical costumer" could be given. This static nature was the reason that led to the Poissonian model of the new calls. The main assumptions of this model are the independence of the calls, and the exponential distribution of call holding times. The formulas developed for this kind of traffic proved to be robust enough to be applicable in cases when the above mentioned assumption of exponential holding times were not totally satisfied. The problem of dimensioning telecommunication networks was basically solved.

The advent of packet switched computer networks, however, raised new challenges for the traffic engineering community.

The different types of applications generate different types of traffic, the "typical user" no longer exists. Most of the applications generate variable bitrate data. The total load on the network is not simply proportional to the number of users or connections, but it is a superposition of many variable bitrate sources. The traffic of different sources interact, for example during data transfer they compete for the available bandwidth. This competition violates the assumption of the independence of traffic sources.

In the case of PSTN we have seen that the distribution of call holding times has a fast exponential decay. The term "call holding time" is no longer applicable for data networks, but one can speak of user session time, or for example the size of a transferred data file. These are the quantities that correspond to call holding time, since they describe for how long the network has to allocate resources for the connection. It was found that the distribution of these is significantly different from exponential, because they have a much slower power-law

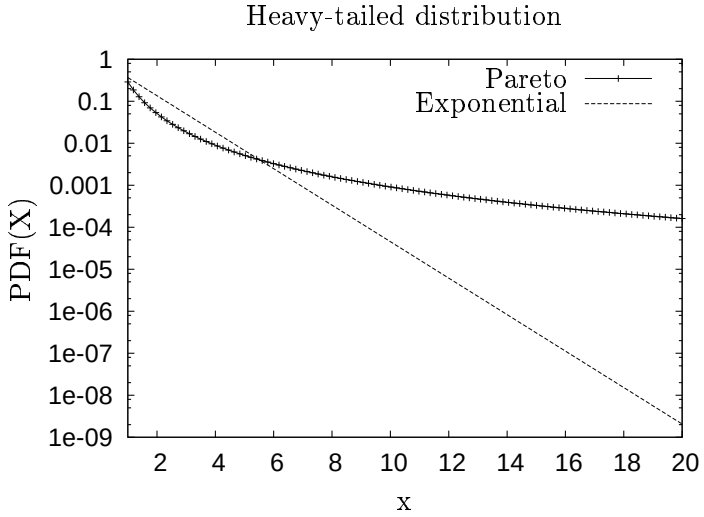

Figure 1: Probability density functions of exponential and heavy-tailed Pareto random variables with the same mean. For the Pareto variable the mass of the probability is concentrated in the tail, the variance of this variable is infinite.

decay. These distributions are called heavy-tailed, because a significant amount of mass is concentrated in the tail of the probability distribution, as illustrated in Figure 1. The main consequence of the heavy-tailedness is that large values appear much more often than in the case of exponential distribution.

We have seen that the assumptions of the Poissonian model are violated, therefore it is no surprise that the aggregated traffic intensity of computer networks was found to differ significantly from that of PSTN. Traffic has a more complex structure.

Periods with low traffic intensity are followed by periods of high intensity. These high intensity periods are called bursts. This bursty nature is typical for data networks.

Such bursts can also be observed in telephone networks when traffic intensity is calculated for very small intervals, but these bursts disappear if the intensity is investigated with a sufficiently coarse granularity. What is different in data networks is that these bursts are not only present at small timescales but also at very large timescales. This is nicely shown in Figure 2. Here we see two different processes at different timescales. The one on the left hand side is bursty on all the depicted timescales. This multi-scale burstiness is also called burst-within-burst.

For data networks one can not identify a given timescale where traffic should be investigated. This type of traffic is referred to as traffic with scaling nature, or scale-free traffic , or traffic with a fractal structure. But it is not only the bursty nature of the traffic that does not change with timescale. In Figure 2 we see that there is also a visual similarity between traffic intensities at different scales. Investigations show that the self-correlation structure of the traffic also remains similar across timescales.

Such type of traffic is referred to as self-similar. There are different version of self-

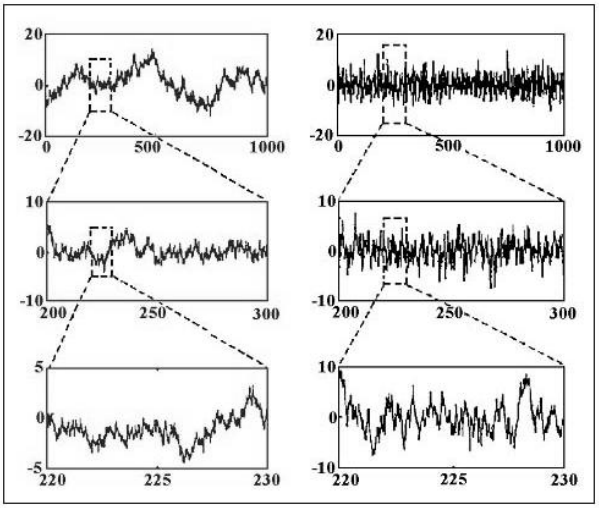

Figure 2: Depicting the same process on different timescales. If the time scale is chosen to be small (bottom row) both processes show high variability. As the time scale increases the on on the right-hand side shows a smooth curve. In contrast the left-hand side process has the burst-within-burst structure and so the bursts do not disappear even if the time granularity is coarser (middle and top rows).

similarity: exact, asymptotic, second-order, discrete, continuous, etc. and also combinations of these. They all refer to the self-correlation structure of the given process and describe that it is qualitatively independent of the chosen timescale. This behaviour of the correlation structure appears in the intensity plot (which is a sample path of the stochastic process) as seen in Figure 2

It is important to note that this bursty structure is not incompatible with stationarity and indeed our models for this bursty self-similar network traffic are stationary. Because of stationarity the most commonly used descriptors for the correlation structure, namely the autocovariance and the autocorrelation functions are in this case one-dimensional, ie. the correlation between two values depends only on the distance between these two values. Whether a process is self-similar or not is closely related to the shape of the autocovariance function. For a large subclass of the asymptotical type of self-similar processes the autocovariance function shows a very slow, power-law decay. We say that these processes are long-range dependent, and they appear as the intensity function of data networks. Therefore they are in the focus of this research.

This strong correlation has significant implications. One is the already discussed burst-within-burst structure. This makes network dimensioning difficult. The covariance function is not summable. This is in strong contrast to traditional models. We will return to long-range dependent processes in the second part of the introduction, where an extensive overview will be given about their most important properties and currently available results.

We have now seen that dimensioning requires different approaches in the case of data and voice networks. This is however not the only difference between these two types of networks.

Consider for example admission control. In the case of PSTN where all sources generate the same constant bitrate stream it is straightforward to determine if there is a slot for a new connection between two endpoints or not, at least from a theoretical point of view. But in the case of packet switched networks the admission of any new flow occupies resources from the network that were previously shared by all the other flows. The investigation of how a new connection degrades the quality of the existing ones is a crucial issue in providing guaranteed quality of service.

Packet buffers, needed to store packet bursts may cause delay but cannot totally eliminate packet loss. Different issues such as buffer size, maximum delay, delay variance and acceptable level of packet loss should be taken into account.

The diversity of the different applications present on the network complicates the analysis even further. From time to time new applications appear, which might have different traffic characteristics and also different quality of service requirements. Some applications are more sensitive to delays, some are sensitive to packet loss, etc.

These and similar issues have been in the centre of interest in the last few decades. Many research papers appeared in this field answering many open questions.

Despite the big effort to clarify these issues many questions still remained open. The problem that motivated this work was the reliable estimation of the mean of long-range dependent processes and a related issue about testing the constancy of the mean. An example for the application of mean estimation is the above mentioned admission control. A simple algorithm could be based on the mean of the current traffic intensity and the estimated mean of the new traffic flow. Another example which requires the detection of mean change, is routing. If an increase in traffic intensity could be discovered, then alternate routes might be allocated to some traffic streams to balance the load on the different links.

The reason why the simple question about the mean of a stochastic process does not have a simple answer, and the reason why the investigation of this issue led to a thorough analysis of long-range dependent and the closely related self-similar (SS) processes will be explained below.

Introduction to long-range dependence and self-similarity

Time series that can be efficiently modelled by so called *long-range dependent* (LRD) stochastic processes have been discovered in various natural or man-made systems. Their first discovery relates to hydrology. In the 1950s Hurst analysed the historical data of the water level of the river Nile which was available for several hundred years. Hurst observed the presence of the very strong correlation. Since then long-range dependence has been encountered in various other fields such as agriculture, physics, and soil science ([2]).

By now long-range dependence has a well-developed literature, and many results already exist that describe one or another of its aspects. Although authors define long-range dependence in slightly different ways all definitions agree in the following main properties:

P1 LRD is usually defined for discrete time, stationary stochastic processes.

P2 The autocovariance function $\gamma(k)$ has a slow power-law decay and so its infinite sum $\sum_{i=0}^{\infty} \gamma(k)$ is infinite.

P3 The variance of the sample mean calculated as $\bar{X} := \frac{\sum_{i=1}^{m} X(i)}{m}$ decreases slowly, i.e. slower than a constant times m^{-1}, where m is the sample size, and X is the long-range dependent process.

P4 The autocorrelation function, which describes the qualitative behaviour of the correlation structure, converges pointwise to a constant function when viewing the same process on successively larger time scales. Viewing the process on larger (or coarser) time scales is achieved by dividing the process into non-overlapping, contiguous blocks of the same size, and replacing the values in each block by their average. This operation, which forms a new process from the original, is also called aggregation, or aggregation of level m, where m is the size of each block. In Figure 2 the same time series is depicted on different timescales.

Property 3 has a direct impact on the estimation of mean. Bigger variance of the estimate means bigger confidence interval, and so smaller accuracy.

Property 4 influences the detection of mean changes. Consider for example the following simple algorithm: the process is subdivided into non-overlapping contiguous intervals and the sample mean is calculated for each interval. A change of mean is concluded if the difference between empirical means is bigger than a given threshold. However this threshold does not only depend on the confidence interval of the estimate but also on the correlation between the consecutive estimates. Property 4 states that even asymptotically when the size of the intervals converge to infinity a correlation of the sequence of estimates will be present that has to be taken into account. Processes that share this property are called asymptotically self-similar. Long-range dependent processes are thus an important subset of asymptotically self-similar processes.

These issues lead to the investigation of the literature of long-range dependence and (asymptotic) self-similarity.

The paper of Cox [6] is an attempt to summarise the most important results of second-order self-similarity and long-range dependence without the need of exact mathematical precision.

Perhaps the most extensive work on different aspects of long-range dependence including estimation and forecasting among others is the monograph of Beran ([2]).

Another set of papers focuses on the application of LRD and self-similarity. From the field of computer networks these papers include [7, 14, 8],[17]. Although these papers use the mathematical disciplines of LRD and SS, their primary focus is on the application of results and so are not a relevant source for the study of the mathematical aspects.

After studying the corresponding literature I returned to the problems of mean estimation and mean-change detection. As mentioned earlier the variance of the sample mean decreases slower for long-range dependent samples than for the traditional short-range dependent ones. Beran [2] gives an asymptotical result for the size of the sample-mean variance. My goal was to extend this result for finite dataset or find a lower cut-off scale where the asymptotical result can be applied with a certain accuracy. Although the asymptotical result was found in many places in the corresponding literature a rigorous proof could not be found. Most papers usually cite other papers which also mention this result, but don't prove them either ([2]). Other papers which gave a proof for this result either placed some additional restrictions

([13]), or used some unproved lemma for the proof ([23]).

Although the study of the corresponding literature was motivated by the above specific problem, during the study it also turned out that this is not the only case where misconceptions are present. Similar, but different definitions for long-range dependence have not been compared. Long-range dependence was sometimes expanded to include processes with regularly varying autocovariance function, but discrete time regular variation wasn't properly defined. Implicitly the inverse of the theorem stating the slow decay of the aggregated variance was also assumed. (See for example the variance time plot estimation method in Chapter 4 on Page 50.)

Therefore it was decided that a clarification of these issues should be the the first step.

As a result the above asymptotic theorem was proved (Property **LRD**–(i), page 45) and also other important properties of long-range dependence and discrete regular variation (Chapter 3) have been collected proved and summarised.

After clarifying the theoretical issues attention was focused on practice. Sample average of long-range dependent processes with known correlation structure was calculated for finite sample size and compared to the asymptotical result. These processes were taken from a family called fractionally integrated autoregressive moving average [5]. By numerical calculations it was found that for all the processes the convergence of the sample mean variance to its asymptotical value was significantly faster than expected. (Section 5.2.1)

During the study of this behaviour, which could not be explained by the existing results. attention was focused on the literature on exact and asymptotic self-similarity.

The most relevant papers investigating discrete self-similarity from a purely mathematical point of view include the papers of Major [16], Sinai [21] and the monograph of Samorodnitsky and Taqqu [20].

These papers give a precise mathematical treatment of discrete self-similarity. In contrast to this work they explicitly prescribe the sequence of normalising factors, so a whole family of processes, which deserve to be called self-similar remain undiscovered. The different definitions that appear in these papers will be reviewed and compared in Section 2.1.

The study of this issue led to a thorough treatment of all exactly and asymptotically self-similar processes, not just long-range dependent ones. In contrast to many previous works which derived the properties of discrete self-similarity from its continuous counterpart this work investigates the discrete self-similarity in its own right. The other important novelty of this work is that instead of the autocovariance function the variance time function is used to describe processes on different time scales. This new approach turned out to be fruitful from many aspects. It largely simplified the analysis of asymptotically self-similar processes, revealed a previously undiscovered family of self-similar processes and could also be used to show why fractionally integrated autoregressive moving average processes show such a special behaviour.

Although the chronological order of research was as presented above the dissertation was structured in a different way, hopefully more suitable for the reader to follow.

As a summary this dissertation concentrates on one dimensional discrete second-order self-similar and asymptotically self-similar processes, paying special attention to an important subset of the latter, the processes with long-range dependence.

After discussing the basic properties of second-order processes in Chapter 1 Chapter 2

deals with self-similar and Chapter 3 deals with long-range dependent processes. Chapter 4 is dedicated to some direct practical impacts of the new results. The final chapter (Chapter 5) demonstrates the advantages of the new approach presented in this work for analysing processes on different time scales while showing the above mentioned fast convergence of fractionally integrated autoregressive moving average processes.

Chapter 1

Definitions and notations

The aim of this chapter is to introduce the notations and measures used for second-order description of processes and their basic relationships.

1.1 Second-order stationary processes

Consider a discrete time stochastic process $\{X(t), t \in \mathbf{Z}\}$. The symbol $\mu(t) := \mathbf{E}[X(t)]$ stands for the mean of $X(t)$ and $\mathcal{V}(t) := \mathbf{E}[(X(t) - \mu(t))^2]$ is its variance. The function $\gamma(k,t) := \mathbf{E}[(X(t+k) - \mu(t+k))(X(t) - \mu(t))]$ is the *autocovariance function* (ACVF) of X. If for the process X: $\mu(t) = \mu(0)$ and $\gamma(k,t) = \gamma(k,0)$ for each $t, k \in \mathbf{Z}$, then X is called *second-order stationary*, and the autocovariance function is defined as

$$\gamma(k) := \mathbf{E}[(X(k) - \mu)(X(0) - \mu)], \tag{1.1}$$

where μ is defined as $\mu := \mu(0)$.

In this work only second-order stationary processes will be investigated, therefore Definition (1.1) is used for the autocovariance function. Also for these processes $\mathcal{V}(t)$ does not depend on t, so $\mathcal{V} = \gamma(0)$ will denote the common variance of all $X(t)$ values, that is the variance of the process. It will be assumed that X is not trivial, that is its variance $\mathcal{V}$ is greater than zero. Apart from the existence of a finite variance no assumption is made on the distributions of the process. Technically this paper deals with functional relationships satisfied by the covariance function. Random variables play no essential role. For our purposes a process is characterised by its autocovariance function. For the calculation of the autocovariance function μ is subtracted from $X(t)$. This means that the ACVF does not depend on μ, so to simplify calculations without loss of generality $\mu = 0$ will be assumed, as far as the ACVF and its derivatives are concerned, which is the case in this work.

Table 1.1 introduces the functions and notations that will be extensively used. The symbols $\gamma_\infty = \lim_{k\to\infty} \gamma(k)$ etc. denote the corresponding limits, when they exist. Infinity is a valid limit for S_∞ and ω_∞, and their normalised forms s_∞ and ϕ_∞.

In this study we also include processes, where non of ρ_∞, s_∞ and ϕ_∞ exist. As an example for a process where ρ_∞ does not exist we define $Y = \{\ldots, X, aX, X, aX, X, \ldots\}$, $a \in [-1, 1]$. Using a fair coin independent of X, we assign the origin of time to X or

aX to ensure stationarity. It is easy to see that $\rho = \{1, b, 1, b, \ldots\}$ is periodic, where $b = 2a/(1+a^2) \in [-1, 1]$. Now if $a = -1$, then $b = -1$ and so, the correlation sum function is $s = 1, -1, 1, -1, 1, \ldots$. The correlation time function for the same process is $\phi = 1, 0, 1, 0, 1, \ldots$.

Notation	Name	Abbr.	Definition	Domain	Limit
$\mathcal{V}$	Variance		$\mathbf{E}[X^2(t)]$		
$\gamma(k)$	AutoCoVariance Function	ACVF	$\mathbf{E}[(X(t+k))(X(t))]$	$k \in \mathbf{Z}$	γ_∞
$\rho(k)$	AutoCorrelation Function	ACF	$\gamma(k)/\gamma(0)$	$k \in \mathbf{Z}$	ρ_∞
$S(n)$	CoVariance Sum Function	CVSF	$\sum_{k=-n}^{n} \gamma(k)$	$n \geq 0$	S_∞
$s(n)$	Correlation Sum Function	CSF	$S(n)/S(0)$	$n \geq 0$	s_∞
$\omega(m)$	Variance Time Function	VTF	$\sum_{n=0}^{m-1} S(n)$	$m \geq 1$	ω_∞
$\phi(m)$	Correlation Time Function	CTF	$\omega(m)/\omega(1)$	$m \geq 1$	ϕ_∞

Table 1.1: Definitions: the covariance and related functions, and normalised forms.

The importance of the covariance sum S and variance time ω functions will be justified progressively below. The first point is that there is a one-to-one mapping between the ACVF, CVSF and the VTF, so they are equivalent ways to describe the covariance structure of the process. The VTF can be expressed in terms of the ACVF using a double-summation:

$$\omega(m) = \sum_{n=0}^{m-1} S(n) = \sum_{n=0}^{m-1} \sum_{i=-n}^{n} \gamma(i), \tag{1.2}$$

so expressing the ACVF in terms of the VTF involves a double-differencing. To simplify notations, the double-differencing operator $\mathbf{D}$ will be introduced, defined as:

$$\mathbf{D}\{f\}(n) = \frac{1}{2} \begin{cases} 2f(1) & : n = 0 \\ f(2) - 2f(1) & : n = 1 \\ f(n+1) - 2f(n) + f(n-1) & : n > 1. \end{cases}$$

With the help of $\mathbf{D}$ the equation that relates γ to ω can simply be written as:

$$\gamma(n) = \mathbf{D}\{\omega\}(n). \tag{1.3}$$

Because of its physical meaning, the ACVF description is the most commonly used, however, it turns out that in many cases the VTF leads to more direct and more elegant solutions, especially in the context of self-similarity. As it will be made clear the variance time function is more than just another equivalent form of describing the covariance structure. It provides a novel approach that simplifies the analysis of second-order processes over different time-scales. The CVSF is not generally considered, except through its limiting value S_∞, the *covariance sum*, which is of considerable importance in various contexts. Note that $\omega(1) = S(0) = \gamma(0) = \mathcal{V}$, so $\phi(1) = s(1) = \rho(1) = 1$, and $\gamma(k) = \gamma(-k)$.

The *autocovariance matrix* of a random vector $\{A_1, A_2, \ldots, A_n\}$ is a $n \times n$ matrix, defined as $[\Gamma_n]_{i,j} = \mathrm{Cov}[A_i, A_j]$. For a stationary stochastic process, for any vector consisting of n contiguous values the autocovariance matrix is of Toeplitz type, satisfying $[\Gamma_n]_{i,j} = \gamma(i-j)$. This matrix, depicted in Figure 1.1, can be used to visualise the connection between the ACVF, CVSF, VTF and more.

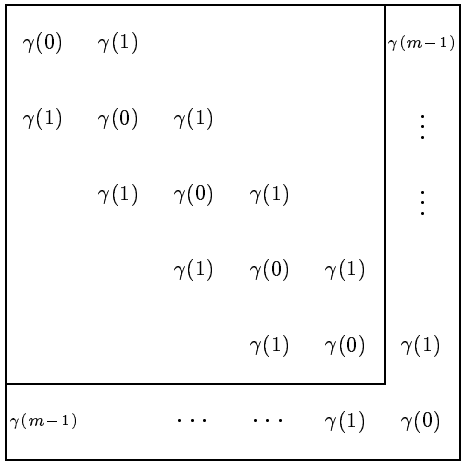

Figure 1.1: The Autocovariance Matrix $[\Gamma_m]_{i,j}$. The sum of terms in the L-shaped form is $S(m-1)$, and the sum over all terms in the matrix is $\omega(m)$. Clearly $\omega(m) = \omega(m-1) + S(m-1)$.

1.2 Positive semi-definiteness

It is natural to ask what possible forms a covariance function may take. As it is well known, the answer is closely related to the property of *positive semi-definiteness* (PSD, [5]).

Let $\mathbf{Y} = \{X_1, X_2, \ldots, X_n\}^T$ be a random vector of n contiguous elements from a process X, and $\mathbf{a}$ any constant real vector of length n. Since variance cannot be negative, one clearly has $\mathrm{Var}[\mathbf{a}^T\mathbf{Y}] \geq 0$, and therefore

$$\mathrm{Var}[\mathbf{a}^T\mathbf{Y}] = \mathbf{E}[\mathbf{a}^T\mathbf{Y}\mathbf{Y}^T\mathbf{a}] = \mathbf{a}^T\Gamma_n\mathbf{a} \geq 0. \tag{1.4}$$

For stationary X this is equivalent to

$$\sum_{1\leq i,j\leq n} a_i\gamma(i-j)a_j \geq 0. \tag{1.5}$$

The matrix Γ_n is *positive semi-definite* if (1.4) holds for all possible vectors $\mathbf{a}$ of length n, whereas the function γ is *positive semi-definite* if (1.5) holds for all possible vectors $\mathbf{a}$, and for any n. Positive semi-definiteness is not only necessary for the existence of a stationary process with the given ACVF. It can be shown that if a function f is positive semi-definite then there exists a Gaussian process with ACVF equal to f [5]. A diagonalisable matrix is PSD if and only if all its eigenvalues are non-negative.

It is notoriously difficult to determine if a given function is positive semi-definite or not. Two of the more obvious and well known necessary conditions are that $\gamma(k) \in [-\mathcal{V}, \mathcal{V}]$, and $\omega(m) \geq 0$, which together with (1.2) imply $\omega(m) \in [0, \mathcal{V}m^2]$. Furthermore it is easier in general to show that a function is not PSD than the opposite. This is because it is sufficient to find a single evidence, like an appropriately chosen vector, or a negative eigenvalue that immediately shows if the function is not PSD. To prove the positive semi-definiteness of a

function is much more difficult. Therefore the existence of processes with given ACFs is usually not proved via the PSDness of the ACF, but the process is constructed from other processes. This way the existence of the process is guaranteed, and also the PSDness of its ACF.

Positive semi-definiteness has a much simpler form in frequency domain. If the discrete time Fourier transform of the ACVF exist, then it is called the spectral density function of the process. This is a real-values, symmetric and periodic function. The ACVF is PSD if and only if its spectral density function is non-negative at all values. the spectral

1.3 Aggregation

From the original process X, for each fixed $m \geq 1$ a new process $X^{(m)}$ can be defined as

$$X^{(m)}(t) := \frac{1}{m} \sum_{j=m(t-1)+1}^{mt} X(j).$$

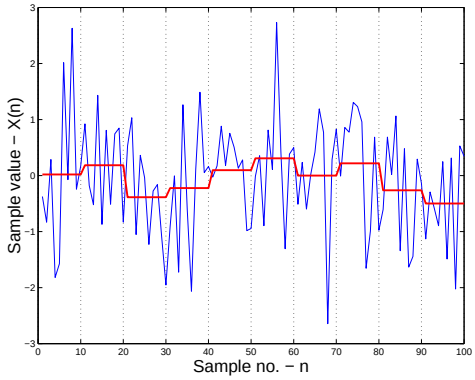

Figure 1.2: The process X and its 10-aggregate: $X^{(10)}$

This operation, which is just an averaging of X over non-overlapping blocks of width m, will be called *aggregation of level m*. Studying the properties of $X^{(m)}$ is useful for several purposes. It is a common approach when studying limiting properties of processes, and it provides a natural way to investigate the original process over many time scales, and thereby self-similarity. Another motivation is that the aggregated process is related to the estimation of the mean of the original process. The variance of $X^{(m)}$ is the variance of the sample mean over m data points: $\bar{\mu}_m := \frac{1}{m}\sum_{j=1}^{m} X(j)$. The ACVF of $X^{(m)}$ therefore corresponds to the covariances of mean estimates over neighbouring blocks, which is required for calculations such as accurate confidence intervals for mean estimates, and the detection of level shifts.

The variance and the functions γ, ρ, S, s, ω and ϕ of $X^{(m)}$ will be denoted by $\mathcal{V}^{(m)}$, $\gamma^{(m)}$, $\rho^{(m)}$, $S^{(m)}$, $s^{(m)}$, $\omega^{(m)}$ and $\phi^{(m)}$ respectively.

We now relate these quantities to those of the original process $X = X^{(1)}$, beginning with the simplest, the variance. Proceeding from the definitions it is straightforward to show that

$$\mathcal{V}^{(n)} = E\left[\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)^2\right] = \frac{1}{n^2}\sum_{j=0}^{n-1}\sum_{i=-j}^{j}\gamma(i) = \frac{1}{n^2}\sum_{j=0}^{n-1}S(j) = \frac{\omega(n)}{n^2}. \quad (1.6)$$

As Equation (1.6) holds for any process, it holds for the m-th level aggregation of the original, thus

$$\mathcal{V}^{(m)(n)} = \frac{\omega^{(m)}(n)}{n^2},$$

where $\mathcal{V}^{(m)(n)}$ is the variance of the n-th aggregation of the m-th aggregation, which is simply the nm-th aggregation of the original process. It follows that $\mathcal{V}^{(m)(n)} = \mathcal{V}^{(mn)}$, which according to (1.6) is equal to $\frac{\omega(mn)}{(mn)^2}$. Putting this together yields

$$\omega^{(m)}(n) = \frac{\omega(mn)}{m^2}, \quad (1.7)$$

from which we obtain the important normalised version:

$$\phi^{(m)}(n) = \frac{\phi(mn)}{\phi(m)}. \quad (1.8)$$

Because $\omega(m)$ is so closely related to the variance of the m-aggregated process, the mapping that relates $\omega^{(m)}$ to ω is very simple. The mapping becomes increasingly complex as we move from the variance time function through to the covariance sum function

$$S^{(m)}(n) = \sum_{k=-n}^{n}\gamma^{(m)}(k) = \frac{1}{m^2}\sum_{i=mn}^{mn+m-1}S(i) \quad (1.9)$$

and to the autocovariance

$$\begin{aligned}\gamma^{(m)}(k) &= E\left[\left(\frac{X_1 + X_2 + \cdots + X_m}{m}\right)\left(\frac{X_{m(k-1)+1} + X_{m(k-1)+2} + \cdots + X_{mk}}{m}\right)\right] \\ &= \frac{1}{m^2}\left[m\gamma(km) + \sum_{i=1}^{m-1} i(\gamma((k-1)m+i) + \gamma((k+1)m-i))\right]. \quad (1.10)\end{aligned}$$

We think of aggregation as a family of operators, indexed by m, mapping time series to new time series. Corresponding to this is another family of operators mapping the space of covariance functions into itself, and another acting on the correlation functions. We explicitly adopt this operator viewpoint in Section 2.2.

1.4 Summary

In this chapter second-order stationary stochastic processes and their basic descriptors were introduced. The descriptors include among others the autocovariance, and covariance time functions and their normalised forms the autocorrelation and correlation time functions.

Positive semi-definiteness as a sufficient and necessary condition for a function to serve as the autocovariance function of a stochastic process was also introduced.

It was also shown how the aggregation operator forms a new process from the original and how aggregation enables the study of the original process on different time-scales. The above mentioned basic descriptors of the aggregated process were expressed in terms of the corresponding quantities of the original process.

Chapter 2

Self-similarity and asymptotic self-similarity

In this chapter a review of the different definitions of self-similarity (SS) will be given. A novel definition will be provided which extends a previously used one in the sense that it includes all the processes covered by the previous definition plus some more. A definition for asymptotic self-similarity will also be given, which is based on the above mentioned new form of self-similarity. Then in Section 2.2 an operator formalism will be developed, which extends the equivalence of the autocovariance and the covariance time functions to the asymptotic regions. Then using this equivalence the set of processes satisfying the above definitions will be explored in Sections 2.3 and 2.4.

2.1 Definitions

The discrete version of self-similarity, which is the scope of this work, was originally deduced from the continuous version. Therefore continuous time self-similarity will be introduced first.

Consider a stochastic process $\{Y(t), t \in \mathbf{R}\}$. Assume that Y is such that

$$Y(at) \stackrel{f.d.}{=} f(a)Y(t), \tag{2.1}$$

for all $a > 0$, where $f(a)$ is some real valued function and $\stackrel{f.d.}{=}$ means equality in all finite dimensional distribution, i.e.

$$A(t) \stackrel{f.d.}{=} B(t) \Leftrightarrow (A_{t_1}, \ldots, A_{t_k}) \stackrel{d}{=} (B_{t_1}, \ldots, B_{t_k}),$$

for all $k \in \mathbf{Z}^+$ and all $t_1, \ldots, it_k \geq 0$.

In this case $Y(at)$ is "similar" to $Y(t)$, hence the word self-similar. Now if we consider $Y(abt)$ then on one hand it is equal (in distribution) to $f(a)f(b)Y(t)$ on the other hand it is equal to $f(ab)Y(t)$, which yields that

$$f(a)f(b) = \pm f(ab), \tag{2.2}$$

for all $a, b > 0$, since if Y is a symmetric random variable then $Y(t)$ and $-1Y(t)$ have the same distribution. But even because of this symmetry we do not reduce the class of processes satisfying (2.1) if we require the function f to be non-negative. Therefore we can rewrite Equation (2.3) in the following form:

$$f(a)f(b) = f(ab), \tag{2.3}$$

It is well-known ([3, Theorem 1.1.9, page 6]) that the only possible solution (under very general conditions on continuity) of (2.3) is $f(a) = a^H$, with $H \in \mathbf{R}$.

The definition of continuous time self-similarity is therefore given as

Definition 2.1.1 (Continuous time self-similarity ([2]))
The stochastic process $\{Y(t), t \in \mathbf{R}\}$ is called self-similar if

$$Y(at) \stackrel{f.d.}{=} a^H Y(t),$$

for all $a > 0$, with $H \in \mathbf{R}$.

As can be seen from the definition, unless $H = 0$, the process is not stationary, since e.g. $Y(1)$ and $Y(2)$ do not have the same marginal distribution. The process, however, can have stationary increments.

Definition 2.1.2 (Stationary increments)
If for any $k \geq 1$ and any k time points $t_1, t_2, \ldots, t_k$, the distribution of $(Y_{t_1+c} - Y_{t_1+c-1}, \ldots, Y_{t_k+c} - Y_{t_k+c-1})$ does not depend on $c \in \mathbf{R}$, then we say that Y_t has *stationary increments.*

We can also define the *increment process* of $Y(t)$ as a discrete time process $X(n) = Y(n) - Y(n-1)$, with $n \in Z^+$.

The discrete definition of self-similarity is not a simple copy of the continuous version, but it is more closely related to its increment process.

Assume that $X(n)$ is the increment process of a self-similar process. In this case we have that

$$(X(1), X(2), \ldots) f(m) \stackrel{d}{=} ((X(1) + X(2) + \cdots + X(m)), (X(m+1) + \cdots + X(2m)), \ldots), \tag{2.4}$$

for all $m = 1, 2, \ldots$, where $f(m)$, as we know, is equal to m^H.

Equation (2.4) can be verified by replacing $X(k)$ by $Y(k) - Y(k-1)$ and $\sum_{i=(k-1)m+1}^{km} X(i)$ by $Y(km) - Y((k-1)m)$.

One of the most general definitions of discrete self-similarity is based on Equation (2.4).

Definition 2.1.3 (Discrete Self-Similarity)
Let $X(t)$ be a process and define $X^{'(m)}$ as

$$X^{'(m)}(t) := A_m \sum_{j=m(t-1)+1}^{mt} X(j) = m A_m X^{(m)}(t). \tag{2.5}$$

The process X is said to exhibit self-similarity if X and $X^{'(m)}$ have the same distribution for all $m \in \mathbf{Z}^+$, where A_m is a sequence of predefined normalising constants. Similarly to the continuous case we do not restrict the class of eligible processes by requiring the positivity of A_m.

The sequence A_m in this definition corresponds to the inverse of $f(m)$ of (2.4)

Note that Definition 2.1.3 does not require the stationarity of $X(t)$, nor the existence of its finite first and second moments. In many places however (like in this work), where this definition is used attention is later on restricted to processes satisfying one or both of these criteria. Because, as we have seen $f(m)$ is restricted to m^H, in many places where this definition is found A_m is given as $A_m := m^{-H}$. Thus the first form of self-similarity which applies to discrete time processes requires the equivalence of X and $m^{1-H}X^{(m)}(t)$.

This definition is found for example in Samorodnitsky and Taqqu ([20]), where the stationarity of $X(t)$ is also required.

Sinai ([21]) and Major ([16]) use a very similar definition but do not restrict their attention to the one dimensional stochastic process, $\{X(t),\ t \in \mathbf{Z}\}$. These works also consider random fields in higher dimensions, $X(t_1, t_2, \ldots, t_d)$. Sinai also investigates non-stationary processes.

Since this work focuses on second-order stationary processes the second-order stationarity of $X(t)$ is required, which implies that both the first and the second moments of X are finite. The corresponding definition is the following:

Definition 2.1.4 (Discrete Self-Similarity #1 (SS1))
Let $X(t)$ be a second-order stationary process and define $X^{'(m)}$ as

$$X^{'(m)}(t) := m^{-H} \sum_{j=m(t-1)+1}^{mt} X(j) = m^{1-H}X^{(m)}(t). \tag{2.6}$$

The process X is said to exhibit discrete second-order self-similarity if X and $X^{'(m)}$ have the same autocovariance function for all $m \in \mathbf{Z}^+$.

It can (and will later on) be shown that in this case H must to be in the range of $[0, 1]$.

In the continuous time case we have seen that the distribution of $Y(t)$ and some constant times $Y(at)$ is equal, therefore they were called self similar. In the second-order stationary discrete case it is the autocovariance function of X and of some constant times $X^{(m)}$ that are equal. This means that the autocovariance functions of X and $X^{(m)}$ differ only in a multiplicative constant, that is $\gamma^{(m)} \equiv C_m\gamma$, where the constant $C_m = (mA_m)^{-2} = m^{2H-2}$ is explicitly defined by the sequence A_m. This immediately yields that $\rho^{(m)} \equiv \rho$, for all $m = 1, 2, \ldots$. The converse is also true, namely if for a process X and all $m = 1, 2, \ldots$: $\rho^{(m)} \equiv \rho$ then there is a sequence A_m, such that (2.5) is satisfied.

A new definition of second-order self-similarity, which will be used in this work, can therefore be given as

Definition 2.1.5 (Discrete Self-Similarity #2 (SS2))
A second-order stationary process X is self-similar if X and $X^{(m)}$ have the same autocorrelation functions ($\rho \equiv \rho^{(m)}$) for all $m \in \mathbf{Z}^+$.

As we have seen Definition 2.1.5 also requires the existence of a sequence A_m such that (2.5) is satisfied. The difference between Definitions 2.1.4 and 2.1.5 is that in the latter case A_m is not restricted to m^{-H}.

To decide whether these two definitions are equivalent or not is not trivial. So far we have only considered the $A_m = m^{-H}$ case. This was because in continuous time Equation (2.3) does not have any other solution. However here it is not required that (2.3) holds for all real values of a, it only has to be satisfied on integer values. In the discrete world there is a much bigger set of solutions for Equation (2.3), as will be discussed in Section 2.3. This means that $A_m = m^{-H}$ is not necessarily the only possibility in the discrete case.

On the other hand even if the sequence A_m is such that the corresponding $f(a)$ satisfies (2.3) it is not necessarily possible to find a stochastic process satisfying Equation (2.5) with the given A_m sequence.

The answer to this question will be given in Section 2.3 where it will turn out that these two definitions are not equivalent.

Consistent with Definition 2.1.5 of self-similarity a definition for asymptotic self-similarity is given as follows:

Definition 2.1.6 (Asymptotic self-similarity (ASS))
A process X with ACF ρ is asymptotically self-similar if $\lim_{m\to\infty} \rho^{(m)}(k) = \rho^*(k)$ exists for all $k \in \mathbf{Z}$.

This means that the process itself is not (necessarily) self-similar, so it might change its statistical behaviour, but under aggregation the autocorrelation function converges to a constant function, asymptotically it behaves like self-similar.

There exist γ, for which $\mathcal{V}^{(m)} = 0$ for some m , so that the corresponding $\rho^{(m)}$ are not defined [1]. This complicates a discussion based on the ACF sequence, however there are only two cases when this occurs, which can be dealt with separately: (i) there exist an $m' \geq 1$ such that $\mathcal{V}^{(m)} = 0$ for all $m \geq m'$, in which case we treat the process as asymptotically self-similar converging to the degenerate autocorrelation function of $\rho^* \equiv 0$, or (ii) zero and non-zero values of $\mathcal{V}^{(m)}$ occur infinitely often as $m \to \infty$, in which case the process does not have a limit in any reasonable sense. Thus neither of these cases creates difficulties for the definitions above, and therefore they are excluded.

This definition will be compared to an alternative definition frequently encountered in the literature in Section 2.4.1.

2.2 Aggregation as operators, equivalence of ρ and ϕ

The simplicity of Equation (1.8) that expresses $\phi^{(m)}$ in terms of ϕ suggests that the functions ω and ϕ should be used to study the effects of aggregation and in particular to explore the set of SS and ASS processes. In this and the forthcoming sections it will be shown that as expected it is really possible and fruitful to work with ω and ϕ instead of γ and ρ.

[1] Consider for example the non-ergodic process $Y = \ldots, X, -X, X, -X, X, \ldots$ with X being any symmetric random variable with non-zero variance. For every even m: $\omega^{(m)} = 0$.

First the equivalence of these two sets of descriptors will be investigated and extended to asymptotic regions.

Table 2.1 defines the operators linking together the functions γ and ω, their normalised forms ρ and ϕ, and their counterparts after aggregation.

The normalisation operator $\mathbf{N}$ acts by simply dividing by the first element of its function argument. This amounts to dividing by $\mathcal{V} \neq 0$ when operating on either of γ or ω, (we allow the difference in index, 0 for γ and 1 for ω, in a harmless abuse of notation). The operator $\mathbf{R_m}$ is the ACF based renormalisation operator of Definition 2.1.5, and $\mathbf{P_m}$ is its CTF counterpart. The operators $\mathbf{I}$ and $\mathbf{D}$ which relate γ to ω perform double summation (Integration) and double Differencing respectively.

Operator	Explicit Mapping	Eqn.
$\mathbf{N} : \gamma \mapsto \rho,\ \omega \mapsto \phi$	$\rho(k) = \gamma(k)/\gamma(0), \quad \phi(k) = \omega(k)/\omega(1)$	
$\mathbf{I} : \gamma \mapsto \omega$	$\omega(n) = \sum_{j=0}^{n-1} \sum_{i=-j}^{j} \gamma(i)$	(1.2)
$\mathbf{D} : \omega \mapsto \gamma$	$\gamma(n) = \mathbf{D}\{\omega\}(n)$	(1.3)
$\mathbf{G_m} : \gamma \mapsto \gamma^{(m)}$	$\gamma^{(m)}(n) = \frac{1}{m^2}\left[m\gamma(nm) + \sum_{i=1}^{m-1} i(\gamma((n-1)m+i) + \gamma((n+1)m-i))\right]$	(1.10)
$\mathbf{W_m} : \omega \mapsto \omega^{(m)}$	$\omega^{(m)}(n) = \omega(mn)/m^2$	(1.7)
$\mathbf{R_m} : \rho \mapsto \rho^{(m)}$	$\rho^{(m)}(n) = \frac{\mathcal{V}}{\mathcal{V}^{(m)}m^2}\left[m\rho(nm) + \sum_{i=1}^{m-1} i(\rho((n-1)m+i) + \rho((n+1)m-i))\right]$	
$\mathbf{P_m} : \phi \mapsto \phi^{(m)}$	$\phi^{(m)}(n) = \phi(mn)/\phi(m)$	(1.8)

Table 2.1: Operators linking Correlation Time Function (CTF) and Autocorrelation Function (ACF) descriptions, and original and aggregated processes.

The first observation is that the functional relationships between the functions in the table remain valid even if they do not correspond to real stochastic processes, i.e. $\gamma \notin$ PSD. These functional relationships are defined in Lemma 2.2.1 and visualised in Figure 2.1.

Lemma 2.2.1 (Operator Properties)
(a) $\mathbf{N}\{c\gamma\} = \mathbf{N}\{\gamma\}$*, where* γ *is a normalisable function and* $c \neq 0$ *a real factor.*
(b) $\mathbf{I}$ *and* $\mathbf{D}$ *are inverses of each other.*
(c) $\mathbf{I}$ *and* $\mathbf{D}$ *preserve normalisation, that is* $\phi \equiv \mathbf{I}\{\rho\}$ *and* $\rho \equiv \mathbf{D}\{\phi\}$.
(d) $\mathbf{I}$ *and* $\mathbf{D}$ *commute with* $\mathbf{N}$ *and with multiplication by a constant.*
(e) $\mathbf{I}\mathbf{G_m} = \mathbf{W_m}\mathbf{I}$.
(f) Let $\{N_j\}$ *be a sequence of real functions that converge pointwise to a function* N*, i.e.* $\lim_{j\to\infty} N_j(k) = N(k)$ *for each* k*. Then, letting* $\mathbf{T}$ *be any of* $\mathbf{I}$, $\mathbf{D}$, $\mathbf{G_m}$ *or* $\mathbf{W_m}$,

$$\mathbf{T}\{\lim_{j\to\infty} N_j\} \equiv \lim_{j\to\infty} \mathbf{T}\{N_j\}.$$

Proof
(a) Obvious from the definition.
(b) This follows from Equations (1.2) and (1.3) after a little algebra.
(c) Follows from the definitions: if $\gamma(0) = 1$ then $\mathbf{I}\{\gamma\}(1) = 1$, and if $\omega(1) = 1$ then $\mathbf{D}\{\omega\}(0) = 1$.

(d) The commutation with multiplication is obvious from the definitions. Using (c), $\mathbf{IN}\{\gamma\} \equiv \mathbf{I}\{\rho\} \equiv \mathbf{NI}\{\rho\}$, and using (a) and (d), $\mathbf{NI}\{\rho\} \equiv \mathbf{N}\gamma(0)\mathbf{I}\{\rho\} \equiv \mathbf{NI}\{\gamma(0)\rho\} \equiv \mathbf{NI}\{\gamma\}$.
(e) Follows from the one to one mapping between γ and ω, and the definitions.
(f) Consider an operator $\mathbf{T}$ with the following property: $\mathbf{T}\{N\}(k) = \sum_{h \in H_k} r_{k,h} N(h)$ where H_k, for k fixed, is a finite, fixed set of indices at which the function N is sampled, and the $r_{k,h}$ are real weights. Operators with this property are called linear operators. Each of $\mathbf{I}$, $\mathbf{D}$, $\mathbf{G_m}$ and $\mathbf{W_m}$ share this *linearity* property. The result follows from the fact that finite linear combinations and limits commute. □

Note that property (f) does not hold for $\mathbf{N}$, $\mathbf{R_m}$ or $\mathbf{P_m}$, as these are all (highly) non-linear. The following commutation diagrams summarise the operator relationships.

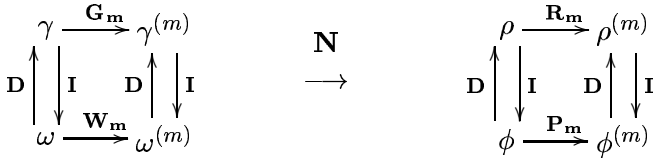

Figure 2.1: Operator relationships. Left: unnormalised operators and functions, Right: normalised.

The fact that the functional relationships are generally valid allows us to split the problem of identifying self-similar processes into two sub-problems. One of them is the identification of those processes that are left unchanged under the operator $\mathbf{R_m}$ and the other is the identification of PSD processes.

Those processes that are left unchanged under the operator $\mathbf{R_m}$ will be called the fixed points (FP) of $\mathbf{R_m}$. These correspond to the ACF of self-similar processes, but are a wider class, since these do not have to satisfy the constraint of positive semi-definiteness. Similarly fixed points of the operator $\mathbf{P_m}$ can also be defined. The formal definition of fixed points is given as follows:

Definition 2.2.1 (Fixed Points of $\mathbf{R_m}$ and $\mathbf{P_m}$)
The function ρ^* is a FP of the operator family $\{\mathbf{R_m}\}$ if, for all $m = 1, 2, 3, \cdots$, $\mathbf{R_m}\{\rho^*\} \equiv \rho^*$.
The function ϕ^* is a FP of the operator family $\{\mathbf{P_m}\}$ if, for all $m = 1, 2, 3, \cdots$, $\mathbf{P_m}\{\phi^*\} \equiv \phi^*$.

Theorem 2.2.1 (Equivalence of Fixed Points)
A function ρ^* is a fixed point of $\mathbf{R_m}$ if and only if $\phi^* \equiv \mathbf{I}\{\rho^*\}$ is a fixed point of $\mathbf{P_m}$.

Proof
Assume that $\mathbf{R_m}\{\rho^*\} \equiv \rho^*$. Operating with $\mathbf{I}$ yields $\mathbf{IR_m}\{\rho^*\} \equiv \mathbf{I}\{\rho^*\} \equiv \phi^*$. Now $\mathbf{IR_m}\{\rho^*\} \equiv \mathbf{ING_m}\{\rho^*\} \equiv \mathbf{NIG_m}\{\rho^*\} \equiv \mathbf{NW_mI}\{\rho^*\} \equiv \mathbf{NW_m}\{\phi^*\} \equiv \mathbf{P_m}\{\phi^*\}$ using Lemma 2.2.1(d) and (e). Thus $\mathbf{P_m}\{\phi^*\} \equiv \phi^*$ as required. A similar proof holds for the other direction. □

According to Theorem 2.2.1 it is possible to find the fixed points of $\mathbf{R_m}$ by first finding the fixed points of the much simpler operator $\mathbf{P_m}$ and then applying operator $\mathbf{D}$.

The simplicity of $\mathbf{P_m}$ can also be utilised during the quest for asymptotically self-similar processes. For this purpose we define limit points as follows:

Definition 2.2.2 (Limit Points of $\mathbf{R_m}$ and $\mathbf{P_m}$)
The function ρ^* is a *limit point* (LP) under $\{\mathbf{R_m}\}$ of some function γ, if $\mathbf{R_m}\{\gamma\} \to \rho^*$ pointwise as $m \to \infty$.
The function ϕ^* is a LP under $\{\mathbf{P_m}\}$ of some function ω, if $\mathbf{P_m}\{\omega\} \to \phi^*$ pointwise as $m \to \infty$.

Functions γ and ω for which $\mathbf{R_m}\{\gamma\} \to \rho^*$ and $\mathbf{P_m}\{\omega\} \to \phi^*$ are said to be in the *domain of attraction*(DoA) of the limit point ρ^* and ϕ^* respectively.

Theorem 2.2.2 (Equivalence of Limit Points)
The LP of γ under $\mathbf{R_m}$ is ρ^* if and only if the LP of $\omega \equiv \mathbf{I}\{\gamma\}$ under $\mathbf{P_m}$ is $\phi^* \equiv \mathbf{I}\{\rho^*\}$.

Proof
Let $\lim_m$ denote $\lim_{m\to\infty}$, and assume that $\lim_m \mathbf{R_m}\{\gamma\} \equiv \rho^*$. Operating with $\mathbf{I}$ yields $\mathbf{I}\lim_m \mathbf{R_m}\{\gamma\} \equiv \mathbf{I}\{\rho^*\} \equiv \phi^*$. Now $\mathbf{I}\lim_m \mathbf{R_m}\{\gamma\} \equiv \mathbf{I}\lim_m \mathbf{NG_m}\{\gamma\} \equiv \lim_m \mathbf{ING_m}\{\gamma\}$ using Lemma 2.2.1(f), and from (d) this becomes $\lim_m \mathbf{NIG_m}\{\gamma\} \equiv \lim_m \mathbf{NW_mI}\{\gamma\} \equiv \lim_m \mathbf{NW_m}\{\omega\} \equiv \lim_m \mathbf{P_m}\{\omega\}$. Thus $\lim_m \mathbf{P_m}\{\omega\} = \phi^*$ as required. A similar proof works for the other direction. □

Now the equivalence of the two formulations has been established. The following theorem proves the connection between fixed and limit points by showing that they are in fact equivalent. It is trivial that all fixed points are limit points, since they are the limit of themselves. The other direction is not trivial, and so it will be proved.

Theorem 2.2.3 (Limit Points are Fixed Points)
(ACF) If ρ^* is a LP of some γ under $\mathbf{R_m}$, then it is also a FP of $\mathbf{R_m}$.
(CTF) If ϕ^* is a LP of some ω under $\mathbf{P_m}$, then it is also a FP of $\mathbf{P_m}$.

Because the equivalence of the ACF and CTF formulations the statements of Theorem 2.2.3 are equivalent, so a single proof suffices.

Proof
We use the ACF definition. We have to show that for each m and n fixed, $\rho^{*(m)}(n) = \rho^*(n)$. Now $\rho^{*(m)} \equiv \mathbf{R_m}\lim_l \mathbf{R_l}\{\gamma\} \equiv \mathbf{NG_m}\lim_l \mathbf{R_l}\{\gamma\} \equiv \mathbf{N}\lim_l \mathbf{G_mR_l}\{\gamma\}$ from Lemma 2.2.1(f). This can be written as $\mathbf{N}\lim_l c_{m,l}\mathbf{R_mR_l}\{\gamma\} \equiv \mathbf{N}\lim_l c_{m,l}\mathbf{R_{lm}}\{\gamma\}$ where the γ-dependent constants $c_{m,l}$ converge: $\lim_l c_{m,l} = c_m$, by the same argument used in the proof of Lemma 2.2.1(f). Thus $\mathbf{N}\lim_l c_{m,l}\mathbf{R_{lm}}\{\gamma\} \equiv \mathbf{N}\{\lim_l c_{m,l}\lim_l \mathbf{R_{lm}}\{\gamma\}\} \equiv \mathbf{N}\{c_m\rho^*\} = \rho^*$, using Lemma 2.2.1(a), and the fact that a subsequence of a convergent sequence is convergent to the same limit. □

This result tells us that the set of fixed points is very important. Any initial function γ either tends to a fixed point under renormalised aggregation, or it does not converge at all.

2.3 Exploring the set of self-similar processes

The aim of this section is to describe the whole set of self-similar processes according to Definitions 2.1.4 and 2.1.5. From a second-order point of view a process is described if its autocovariance function is known. However, one can easily justify that using either definition, the self-similarity of a process is determined by its ACF, it does not depend on the variance of the process. Therefore self-similar processes can be characterised by their ACFs. As it was pointed out in Section 2.2 the correlation time function provides a simpler but equivalent description to the autocorrelation function therefore in this section self-similar processes will be described by their CTFs. In order to have a uniform notation across the two formulations, we introduce the following term.

Definition 2.3.1 (Valid functions)
A function γ is said to be a *valid* (autocovariance or autocorrelation) function if it is positive semi-definite.
A function ω is said to be a *valid* (variance time or correlation time) function if $\gamma = \delta^2\{\omega\}$ is positive semi-definite.

The identification of all SS2 processes will follow the following steps:

1. If a function ϕ' is the CTF of an SS2 process, then basically by definition, it is also a fixed point of the operator $\mathbf{P_m}$. So first the fixed points of $\mathbf{P_m}$ will be identified.
2. Then in the second step the set of candidate functions will be reduced by requiring that they satisfy some constraints introduced by the positive semi-definiteness of the corresponding ACF.
3. As a final step it will be shown that the remaining functions are really valid, and so there exist processes with the given CTF/ACF.

Once SS2 processes have been identified the set will further be reduced to select the entire range of SS1 processes.

2.3.1 Fixed points of $\mathbf{P_m}$

According to Definition 2.2.1 a function ϕ^* is a fixed point of $\mathbf{P_m}$ if $\mathbf{P_m}\{\phi^*\} = \phi^*$, that is

$$\phi(n) = \frac{\phi(nm)}{\phi(m)}, \text{ for all } n, m \in \mathbf{Z}^+. \tag{2.8}$$

Functions with $\phi(m) = 0$ for some m can be excluded, since this would be equivalent to the case of $\mathcal{V}^{(n)} = 0$, which has already been dealt with. (See the discussion after Definition 2.1.6 on page 24) So Equation (2.8) can be written as

$$\phi(mn) = \phi(m)\phi(n), \text{ for all } n, m \in \mathbf{Z}^+. \tag{2.9}$$

Equation (2.9) provides a simple, sufficient and necessary condition for being a fixed point of $\mathbf{P_m}$. Functions satisfying (2.9) already have a name and a well-developed literature.

Definition 2.3.2 (Multiplicative function)
A function $\phi(n)$ defined on $\mathbf{Z}^+$ is *multiplicative* if Equation (2.9) is satisfied for all n, m co-primes. The function is said to be *totally multiplicative* if (2.9) is satisfied for all n, m positive integers.

As it was pointed out in Section 2.1 the only possible solution (under very general conditions on continuity) of $g(xy) = g(x)g(y)$ for real arguments is $g(x) = x^\alpha$, $\alpha \in \mathbf{R}$. One can immediately verify that these simple pure power-law solutions remain valid for discrete arguments.

By prime factorisation it is not difficult to show that for all totally multiplicative functions ϕ:

$$\phi(m) = \prod_{i=1}^{s} \phi(p_i)^{r_i}, \quad \text{for each } m \in \mathbf{Z}^+, \tag{2.10}$$

where the p_i are the s distinct prime factors of n, and r_i is the multiplicity of p_i.

Equation (2.10) shows that totally multiplicative functions are entirely characterised by their values on primes. A way of constructing such a function is assigning the values at primes randomly and use (2.10) to determine the values at non-primes. It is therefore clear that a huge variety of solutions exist which are different, even radically so, from simple power-laws. It is also clear that most such random solutions will not be valid, and further that many will be pathological: with none of ρ_∞, S_∞, or ϕ_∞ existing.

Two simple examples of fixed points will now be constructed. The first example is constructed by assigning $\phi(p) = 0.38$, for each prime p. Values of ϕ at non-primes can be calculated using (2.10). This ϕ function and the corresponding $\rho = \mathbf{D}\{\phi\}$ are depicted on the left side of Figure 2.2. Although $\rho(k) \in (-1, 1) \quad \forall k \in Z, k \neq 0$ it will turn out that this example is not valid. The second example is constructed by assigning $\phi(p) = 1$ for all primes p except $p = 5$, where $\phi(5) = 0.2$. This example is depicted on the right side of Figure 2.2. It will be shown, that in contrast to the previous one, this example is valid. Note that for this latter example none of ρ_∞, S_∞, or ϕ_∞ exist.

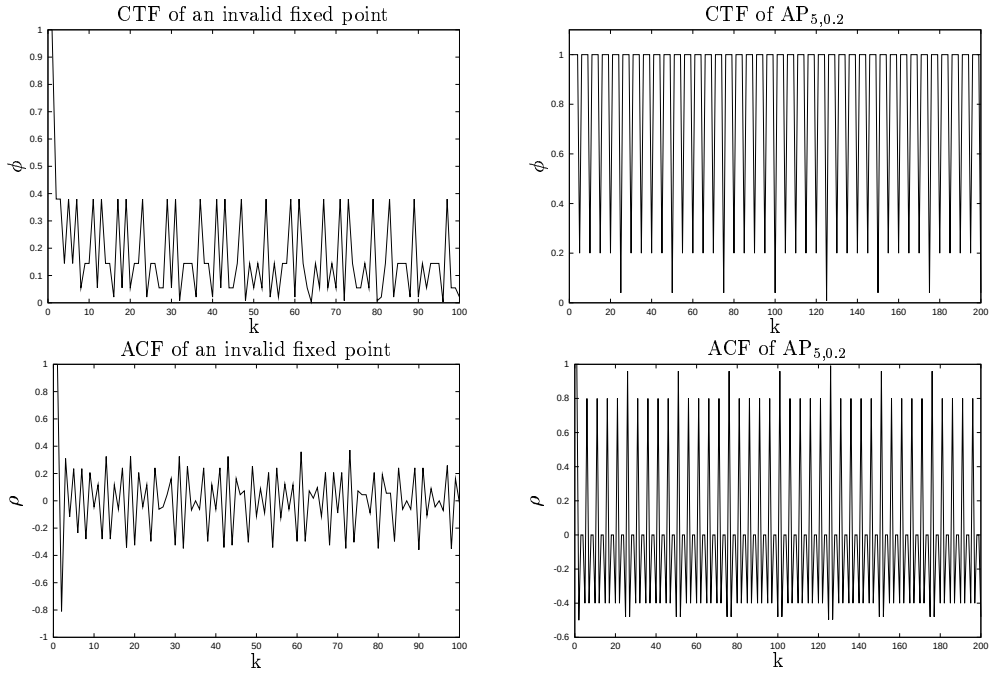

Figure 2.2: Examples of fixed points of $\mathbf{P_m}$ (top row) and the corresponding fixed points of $\mathbf{R_m}$ (bottom row) with $\rho(k) \in (-1, 1)$, $k \neq 0$. Left: 100 lags of an invalid fixed point defined by $\phi(p) = 0.38$ for each prime p. Right: 200 lags of a valid fixed point from the (Almost Periodic) $\mathrm{AP}_{q,c}$ family with $q = 5$, $c = 0.2$.

2.3.2 Valid fixed points

We have now seen that the fixed points of $\mathbf{P_m}$ correspond to totally multiplicative functions. Now we will reduce the set of candidate processes by requiring the positive semi-definiteness of the corresponding ACFs.

Theorem 2.3.1 (Valid fixed points)
Let ϕ be a totally multiplicative function and $\rho = \mathbf{D}\{\phi\}$. If ρ is positive semi-definite then either of the following holds:

1. $\phi(n) = n^\beta$, for all $n = 1, 2, 3, \ldots$ and $\beta \in [0, 2]$.
2. $\phi(q) = 1$ for all primes q, except for exactly one prime p, where $\phi(p) \in (0, 1)$. The value of ϕ at non-primes can be calculated using Equation (2.10).

The following two consequences of positive semi-definiteness will be used (with proofs in the Appendix) to exclude classes of fixed points from consideration. Recall that the correlation sum function is defined as $s(n) = \sum_{i=-n}^{n} \rho(i)$.

CTF–(i) $\phi(m - n) \leq 2(\phi(m) + \phi(n))$ for all $m, n \in \mathbf{Z}^+$, $m - n \geq 1$.

CTF–(ii) For any $n \in \mathbf{Z}^+$, $|s(n)| < 2\sqrt{\phi(n)} + C$, where C is a constant independent of n.

Proof of Theorem 2.3.1
The following concept is at work: for any fixed point ϕ, the powers of each prime p lie along a unique power-law curve given by $f(x) = x^{\alpha_p}$, where $\alpha_p = \frac{\ln \phi(p)}{\ln p}$. Hence if $p > q$ and $\alpha_p > \alpha_q$, then $\phi(p) > \phi(q)$.
Case I: Assume that $\phi(p) > 1$ for some prime p, and that there exist q_1, q_2 such that $\alpha_{q_1} \neq \alpha_{q_2}$. Let $\alpha := \sup_p \alpha_p > 0$.
Case Ia: The supremum α is not attained. Hence there exists an infinite sequence of primes p_i s.t.

$$\alpha_{p_i} > \max_{q<p_i} \alpha_q.$$

Let $\beta_i := \alpha_{p_i}$. According to **CTF**–(ii), and using the fact that $\phi(m) < \phi(p_i)$ for all $m < p_i$,

$$\phi(p_i^2) - \phi(p_i^2 - 1) = s(p_i^2 - 1) < 2\sqrt{\phi(p_i)^2} + C = 2\phi(p_i) + C = 2p_i^{\beta_i} + C. \tag{2.11}$$

For $p_i > 2$ clearly $2|(p_i^2 - 1)$, and using the definition of β_i we have

$$\phi(p_i^2 - 1) = \phi(\frac{p_i^2 - 1}{2})\phi(2) < \left(\frac{p_i^2}{2}\right)^{\beta_i} \phi(2) = p_i^{2\beta_i} \frac{\phi(2)}{2^{\beta_i}}. \tag{2.12}$$

Combining Equations (2.11) and (2.12) we have

$$p_i^{2\beta_i} \left(1 - \frac{\phi(2)}{2^{\beta_i}}\right) < \phi(p_i^2) - \phi(p_i^2 - 1) < 2p_i^{\beta_i} + C.$$

Dropping the middle term, substituting $\gamma = 1 - \frac{\phi(2)}{2^{\beta_i}} > 0$, and dividing by $p_i^{\beta_i}$ we obtain

$$p_i^{\beta_i} < \frac{1}{\gamma}\left(2 + \frac{C}{p_i^{\beta^i}}\right). \tag{2.13}$$

Since $\beta_i > 0$, as $i \to \infty$ the left-hand side of (2.13) diverges, whereas the right-hand side is bounded, which is a contradiction. Such cases are therefore invalid.

Case Ib: The supremum α is attained. Let q be a prime with $\alpha_q = \alpha$, and p another with $\alpha_p < \alpha$. According to Euler's theorem [18], there exists a positive integer j such that $q^j = 1 \bmod p$, and so for any $k \in Z^+$: $q^{kj} = 1 \bmod p$. Using this and the definition of α,

$$\phi(q^{kj} - 1) = \phi(p)\phi\left(\frac{q^{kj} - 1}{p}\right) \leq \phi(p)\left(\frac{q^{kj} - 1}{p}\right)^{\alpha} < q^{kj\alpha}\left(\frac{\phi(p)}{p^{\alpha}}\right).$$

Combining this with Property **CTF**–(ii) we obtain

$$q^{kj\alpha}\left(1 - \frac{\phi(p)}{p^{\alpha}}\right) < \phi(q^{kj}) - \phi(q^{kj} - 1) < 2q^{kj\alpha/2} + C,$$

which, similarly to Equation (2.13), yields a contradiction as $k \to \infty$. Such cases are therefore also invalid.

Case II: $\phi(p) \leq 1$ for all p prime and assume that there exist two different primes q_1 and q_2 for which $\phi(q_i) < 1$.

Using the non-negativity of ϕ and substituting $k = m - n$ in **CTF**–(i), we have

$$0 \leq \phi(k) \leq 2(\phi(n + k) + \phi(n)).$$

For any $\epsilon > 0$ we can choose an $r \in \mathbf{Z}^+$ sufficiently large such that $\phi(q_i^r) < \epsilon$, for $i = 1, 2$. It is easy to see that for any $k \geq 1$ it is possible to find an n such that $q_1^r | n$ and $q_2^r | (n + k)$. Since $\phi(p) \leq 1$ for all p, this implies that for each k

$$0 \leq \phi(k) \leq 2(\phi(n + k) + \phi(n)) \leq 4\epsilon$$

which is only consistent with the fixed point $\phi \equiv 0$, a case already excluded. All such cases are therefore invalid.

Many cases have now been excluded. What remains is the following : $\phi(p) > 1$ for some prime p, and all α_p-s are equal. This corresponds to the case of $\phi(n) = n^{\beta}$, with $\beta > 0$. In 1.2 it was shown that $\omega(m) \in [0, \mathcal{V}m^2]$, which yields that $\beta \leq 2$.

$\phi(p) \leq 1$ for all p primes and there is either none or exactly one prime q such that $\phi(q) < 1$. When there is no such prime then $\phi \equiv 1$. This in fact corresponds to $\phi(n) = n^{\beta}$, with $\beta = 0$ and as such will be treated together with the previous set of processes. □

The first set will be called **fractional noise**, and the second set **almost periodic** processes.

2.3.3 Identification of all self-similar processes

In this subsection it will be shown that the two sets of processes identified in the previous subsection really exist, and their basic properties will be explored.

Fractional noise

Definition 2.3.3 (Fractional noise(FN))

By *fractional noise* we mean a family of processes which have the same autocovariance function as the fractional Gaussian noise ([2, 20]), but the requirement of Gaussianity is dropped. The family therefore has CTF $\phi(n) = n^{2H}$ and so ACF $\rho(k) = \mathbf{D}\{i^{2H}\}(k)$, where $H \in [0, 1]$ is the Hurst parameter [20].

These processes are also self-similar according to Definition 2.1.4, and in fact it follows directly from the definition that these are the only SS1 processes. The definition requires that X and $X'^{(m)} = m^{1-H}X^{(m)}(t)$ have the same autocovariance function, and so as a special case they also have the same variance.

$$\mathrm{Var}(X'^{(m)}) = m^{2(1-H)}\mathcal{V}^{(m)} = \mathcal{V} = \omega(1) \tag{2.14}$$

$$m^{-2H}\omega(m) = \omega(1) \tag{2.15}$$

$$\frac{\omega(m)}{\omega(1)} = \phi(m) = m^{2H} \tag{2.16}$$

We have now shown that it is a necessary condition for $\phi(m)$ to be equal to m^{2H} in order to satisfy Definition 2.1.4. By straightforward calculations it can also be shown that this condition is also sufficient.

Definition 2.1.4 was derived via the increment process of the continuous time H-ss process, giving an alternative definition for fractional noise.

For each value of H within $[0, 1]$ it is well known ([20]) that $\rho(k)$ is PSD, and so there are processes that correspond to it. This includes the special values of $H = \{0, 1\}$, although they are frequently omitted in the literature.

The FN family is important as it plays a central role in discrete self-similarity, as we will show in detail in Section 2.4. Depending on the value of H, several qualitatively different behaviours are found. These will be detailed below.

$\mathbf{H = 1}$

The autocorrelation function of this process is $\phi \equiv 1$, with values perfectly correlated. These processes are non-ergodic. An example is the process $X(t) = X$, for all $t \in \mathbf{Z}$, where X is a random variable with finite, but non zero variance.

$\mathbf{H \in (0.5, 1)}$

These processes are usually called long-range dependent. Although $\gamma_\infty = 0$, the covariance sum $S_\infty = \sum_{i=-\infty}^{\infty} \gamma(i)$ is infinite.

$\mathbf{H = 0.5}$

This is the totally uncorrelated *white noise* process, where $\rho(k) = 0$ for $k \neq 0$.

$\mathbf{H \in (0, 0.5)}$

For these processes the covariance sum S_∞ is zero, and $\rho(k) < 0$ for all $k \neq 0$. These processes will be called *constrained short-range dependent*, since $S_\infty = 0$ is a constraint, and processes where S_∞ is finite are usually called *short-range dependent* (SRD).

H = 0

Let $X(t)$ be a white noise process and $Y(t) := X(t) - X(t-1)$ a moving-average process of order 1 ([5]). Then the process Y has ACF of $\rho(1) = -0.5$, and $\rho(k) = 0$ for $|k| > 1$. This differenced process is FN_0.

The discrete time process $X(t)$ can also be viewed as sampling of a continuous time white noise process which is nothing else but a H-ss process with $H = 0$.

These different subsets of FN processes are summarised in Table 2.2.

H subclass	γ_∞	S_∞	ω_∞	$\rho(k)$, $k \geq 1$	Name/Comment
1	$\mathcal{V}$	∞	∞	constant, $= 1$	Non ergodic ($= \mathcal{NE}_1$)
$(0.5, 1)$	0	∞	∞	decreasing, in $(0, 1)$	Long-range dependent
0.5	0	$\mathcal{V}$	∞	constant, $= 0$	White noise (WN)
$(0, 0.5)$	0	0	∞	increasing, in $(-0.5, 0)$	Constrained SRD
0	0	0	$\mathcal{V}$	$\rho(1) = -0.5$, else $= 0$	differenced white noise

Table 2.2: The Fractional Noise (FN_H) family. Depending on H, qualitatively different behaviour is found.

Almost periodic processes

Definition 2.3.4 (The almost periodic family $\mathbf{AP}_{q,c}$)
The two parameter family defined by $\phi(1) = 1$, $\phi(p) = 1$ for all primes p except $p = q$, where $\phi(q) = c$, $c \in (0, 1)$, will be called *Almost Periodic*, and denoted by $\mathrm{AP}_{q,c}$.

A member of the family is exhibited in the right-hand side plot of Figure 2.2, where its 'almost periodic' nature is readily appreciated.

The next theorem shows that the $\mathrm{AP}_{q,c}$ family is valid.

Theorem 2.3.2 (Validity of $\mathbf{AP}_{q,c}$)
Each member of the $\mathrm{AP}_{q,c}$ family is a valid fixed point.

Lemma 2.3.1 (Limits of processes remain processes)
If a sequence $\{\omega_k\}$ of valid VTFs converges pointwise, then the limit ω is also valid.

This lemma will be proved in the Appendix.

Proof of Theorem 2.3.2
Define the periodic function $\bar{X}_m(t)$ as $\bar{X}_m(t) = 1$ if $m|t$, and 0 otherwise [2] . A stationary process $X_m(t)$ can be defined as $X_m(t) := \bar{X}_m(t - \kappa)$ where κ is a random variable uniformly distributed on 0, 1, ..., $m - 1$. Now consider $Y_m(t) := X_m(t) - X_m(t-1)$. Since $\omega_m(n) := \omega_{Y_m}(n) = \mathbf{E}[(Y_1 + Y_2 + \cdots + Y_n)^2] = \mathbf{E}[(X_n - X_0)^2]$, it is straightforward to show that $\phi_m(n) = 1$ except when $m|n$, where it vanishes.

[2] Here $a|b$ means a divides b, i.e. $b = a * n$, $n \in \mathbf{Z}$.

We now construct a new CTF via a convergent infinite sum of CTFs of the above type:

$$\phi_{q,c}(n) := \frac{1-c}{c} \sum_{k=1}^{\infty} c^k \phi_{q^k}(n). \tag{2.17}$$

It is straightforward to verify that $\phi_{q,c}(n) = c^f$, where $n = aq^f$, $q \not| a$, $a \in \mathbf{Z}^+$, and f a non-negative integer. This, however, is nothing other than the CTF of $\mathrm{AP}_{q,c}$. It remains to show that $\phi_{q,c}(n)$ is valid. From Equation (1.6) it is clear that for finite sums of independent processes the ω simply add, and remain PSD. Lemma 2.3.1, applied to a sequence of partial sums constructed from Equation (2.17), shows that the convergent limit is also PSD, as required. □

2.4 Asymptotically self-similar processes

2.4.1 Comparing definitions for asymptotical self-similarity

In this subsection a different definition for asymptotical self-similarity will be given as promised in Section 2.1.

Definition 2.4.1 (Asymptotic self-similarity, Alternative (AASS))
A process X with ACF ρ is asymptotically self-similar if $\lim_{m\to\infty} \rho^{(m)}(k) = \rho^*(k)$ exists for all $k \in \mathbf{Z}$, where $\rho^*(k)$ is the ACF of a fractional noise process [8, 6, 23].

That this alternative definition dominates the literature can be attributed to the following reasons: As self-similar processes are defined, either directly or indirectly, to be equivalent with the class of fractional noise, it is no surprise that these processes appear in the definition of asymptotical self-similarity. Second, the equivalence of limit points and fixed points as demonstrated in Theorem 2.2.3 has not been investigated before. Therefore it was desirable to explicitly define the limiting process.

All processes satisfying Definition 2.4.1 also satisfy 2.1.6, but for example the almost periodic processes satisfy 2.1.6 but not 2.4.1 so AASS processes form a real subset of ASS processes.

2.4.2 Exploring the set of asymptotically self-similar processes

In this section only valid functions will be considered. Although the domain of attraction of valid fixed points may contain invalid functions they will not be investigated here. Lemma 2.3.1 ensures that all valid functions converge to valid limit/fixed points if they converge at all.

Here Definition 2.1.6 will be used for asymptotic self-similarity. To enable a simpler treatment this definition will be reformulated to use the correlation time function instead of the autocorrelation function. That this is an equivalent definition can be justified by the previously demonstrated equivalence of the ACF and CTF approaches and the equivalence of limit points and fixed points (Theorem 2.2.3).

Definition 2.4.2 ((CTF based) Asymptotic Second-Order Self-Similarity)
A process is asymptotically second-order self-similar if its correlation time function satisfies

$$\lim_{m\to\infty} \mathbf{P_m}\phi \to \phi^*, \quad \text{or} \quad \lim_{m\to\infty} \frac{\phi(nm)}{\phi(m)} = \phi^*(n), \tag{2.18}$$

where ϕ^* is a fixed point.

The different DoA are already well defined and characterised precisely in terms of ϕ. The purpose of this subsection is to describe them in more accessible terms, notably through important families of examples. A fundamental realisation is that, according to Definition 2.1.6, *all* processes which have limit points are asymptotically self-similar. Such a general definition of asymptotic self-similarity is desirable, it highlights the importance of the set of SS processes, and should be compared with the generality of the central limit theorem, which is important precisely because many distributions are in the domain of attraction of the Gaussian distribution.

A simple observation is that each DoA is non-empty, as it contains the fixed point itself. The next natural question is whether they are non-trivial, that is if they contain more than just one point. We address this question in the case of the FN_H family in the remainder of the section. The issue remains open for the $\mathrm{AP}_{q,c}$ family.

First two general methods will be given to construct processes in the DoA of FN_H . Recall that the processes which tend to FN_H are characterised by:

$$\lim_{m\to\infty} \frac{\omega(nm)}{\omega(m)} = \lim_{m\to\infty} \frac{\phi(nm)}{\phi(m)} = n^{2H}, \quad \text{with} \quad H \in [0,1]. \tag{2.19}$$

Let $X_1(t)$ and $X_2(t)$ be independent second-order stationary processes, with variance time functions ω_1 and ω_2. From Equation (1.6) it is clear that the VTF of $X_1 + X_2$ is $\omega = \omega_1 + \omega_2$. Now let $X_1(t)$ be FN_H with $H > 0$ and $X_2(t)$ be FN_{H_2} with $H_2 < H$, each with arbitrary variance. Since ω_1 goes to infinity faster than ω_2, the ratio $\frac{\omega(nm)}{\omega(m)}$ is dominated by ω_1 and so the process is in the DoA of FN_H. The second general method is based on the deterministic mixing of two (or more) independent processes. It is explained in detail during the proof of Property **LRD**–(iii) in Section 3.2.

Following the subclasses defined in Table 2.2, we now give explicit example classes to show that the DoA of each SS process of FN_H type is non-trivial.

H = 1
Let X be a random variable with zero mean and unit variance. Define the process $Y = \{\ldots, X, aX, X, aX, X, \ldots\}$, $a \in [-1, 1]$ where, using a fair coin independent of X, we assign the origin of time to X or aX to ensure stationarity. It is easy to see that $\rho = \{1, b, 1, b, \ldots\}$ is periodic, where $b = 2a/(1 + a^2) \in [-1, 1]$, and for even m (odd is similar) $\phi(m) = \frac{m^2}{2}(b + 1)$. Each member of this family with $b > -1$ ($a > -1$) is in the DoA of FN_1, since $\lim_{m\to\infty} \phi(nm)/\phi(m) = n^2$.

H ∈ (0.5, 1)
By any reasonable definition, the FN_H subclass with $H \in (0.5, 1)$ is long-range dependent, and indeed we define LRD processes as those in the domain of attraction of this class. A full discussion and numerous examples are given in Section 3.2.

H = 0.5

The FN_H with $H = 1/2$ is just white noise, and we call its domain of attraction the set of *short-range dependent* processes. An important subclass is formed by processes for which $S_\infty = \lim_{n\to\infty} S(n) = \sum_{k=-\infty}^{\infty} \gamma(k)$ exists, and lies in $(0, \infty)$, the finite value indicating that the net influence of the past is not too heavy (essentially, any process whose spectral density exists and converges to a finite but non-zero at the origin). For white noise $S_\infty = \mathcal{V}$. This subclass is quite large, for example it includes the frequently used autoregressive moving average (ARMA) models [5].

Recall the definition of *asymptotic equivalence* (Æ): $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) \to 1$.

Lemma 2.4.1

If $S_\infty \in (0, \infty)$ is defined, then $\mathcal{V}^{(m)} \sim \frac{S_\infty}{m}$.

Proof

From Equation (1.6) it is clear that

$$m^2\mathcal{V}^{(m)} = \omega^{(m)} = \sum_{j=0}^{m-1} S(j).$$

Now chose an $\epsilon > 0$ and chose j_0 such that $\forall j \geq j_0 : |S_\infty - S(j)| < \epsilon$.

$$\begin{aligned}
m\mathcal{V}^{(m)} &= \frac{\sum_{j=0}^{j_0-1} S(j)}{m} + \frac{\sum_{j=j_0}^{m-1} S(j)}{m} \\
m\mathcal{V}^{(m)} &< \frac{\sum_{j=0}^{j_0-1} S(j)}{m} + \frac{\sum_{j=j_0}^{m-1} S_\infty + \epsilon}{m} \\
m\mathcal{V}^{(m)} &< \frac{\sum_{j=0}^{j_0-1} S(j)}{m} + \frac{m - j_0}{m}(S_\infty + \epsilon) \\
\lim_{m\to\infty} m\mathcal{V}^{(m)} &< 0 + S_\infty + \epsilon
\end{aligned}$$

Similarly it can be shown that $\lim_{m\to\infty} m\mathcal{V}^{(m)} > S_\infty - \epsilon$. Since this can be shown for every $\epsilon > 0$ it can be concluded that $\lim_{m\to\infty} m\mathcal{V}^{(m)} = S_\infty$. □

An alternative proof can be found in [12]. We can now immediately give

Theorem 2.4.1 (Classic short-range dependence)

If $S_\infty \in (0, \infty)$ is defined, then the process is in the domain of attraction of white noise.

Proof

From Equation (1.6), Property Æ–(ii) (see Appendix), and Lemma 2.4.1, it follows that $\omega(m) \sim mS_\infty$, and so $\lim_{m\to\infty} \phi(nm)/\phi(m) = n$. Thus members of the class are in the domain of attraction of white noise. □

H ∈ (0, 0.5)

The fixed points with $H \in (0, 1/2)$ typically receive little attention. We name the processes in their domains of attraction *constrained short-range dependent*. We show that a large subclass of processes with $S_\infty = 0$ are in the DoA. This means that the covariances conspire to cancel exactly , which can be thought of as a constraint.

H	γ	S_∞	$\omega^{(\infty)}$	$\mathcal{V}^{(m)} \overset{m}{\sim}$	Name/Comment
1	$c_1, c_2, c_1, c_2, \ldots$	∞	∞	$(c_1 + c_2)/2$	Non-ergodic
$(0.5, 1)$	$\gamma(k) \sim c_\gamma(k) m^{2H-2}$	∞	∞	$\frac{c_\gamma(m)}{H(2H-1)} m^{2H-2}$	LRD
0.5	–	$(0, \infty)$	∞	$S_\infty m^{-1}$	SRD
$(0, 0.5)$	$\gamma(k) \sim c_\gamma(k) m^{2H-2}$	0	∞	$\frac{c_\gamma(m)}{H(2H-1)} m^{2H-2}$	CSRD
0	–	0	$(0, \infty)$	$\omega^{(\infty)} m^{-2}$	Differentiated SRD

Table 2.3: Processes in the domain of attraction of FN_H for different values of H. The shaded fields represent a set of sufficient conditions for a process to be in the appropriate DoA. The other values (in the fields with white background) are determined by the values in the shaded fields. $c_\gamma()$ is a slowly varying function, as defined in Section 3.1. A simple example of a slowly varying function is any function that converges to a constant.

Theorem 2.4.2 (Constrained Short-Range Dependence)
If $S_\infty = 0$ is defined, and $\gamma(k) \sim c_\gamma k^{2H-2}$ with $H \in (0, 1/2)$ and $c_\gamma < 0$ constant, then the process is in the domain of attraction of FN_H .

The proof is left to the Appendix as it uses results from the next section. That such processes exist is again easy to see in the frequency domain.

H = 0
Finally we consider the $H = 0$ case, which has been largely ignored. If we assume that $\phi_\infty \in (0, \infty)$ exists, then clearly $\lim_{m\to\infty} \phi(nm)/\phi(m) = 1$, corresponding to $H = 0$. We now exhibit a large class of processes which have this property. Let X be any stationary process such that ρ_∞ exists and $\rho(1) < 1$. Consider the differenced process $Y(i) = X(i+1) - X(i)$. By straightforward calculations one can show that $\phi_Y(m) = (1-\rho(m))/(1-\rho(1))$. Hence $\phi_\infty < \infty$ exists, and since this implies, using Lemma A.2.2 from the Appendix, that $\phi_\infty \geq 1/4 > 0$, Y is in the domain of attraction of FN_0. If X is white noise, then Y is immediately FN_0.

These typical examples of asymptotically self-similar processes are summarised in Table 2.3.

2.5 Conclusion

An operator formalism was developed in this chapter, which proved to be useful in extending the equivalence of the autocovariance and covariance time functions to asymptotic regions. The usage of the covariance time function and its normalised form the correlation time function remarkably simplified the exploration of self-similar and asymptotically self-similar processes.

It was shown that the two definitions of self-similarity 2.1.4 and 2.1.5 are, contrary to the common belief (see e.g. [6] and other papers referring to it), not equivalent. An explicit example was given from the subset SS2\SS1, showing that fractional noise is not the only process that does not change its autocorrelation function during aggregation..

Large classes of typical examples have been presented for asymptotically self-similar processes.

Chapter 3

Long-range dependence

From a mathematical point of view long-range dependent processes are simply a subset of asymptotically self-similar processes converging to FN_H with $H \in (0.5, 1)$ as introduced in Section 2.4. However their practical importance justifies a detailed investigation of these processes. Many of the results to be presented here are also useful when analysing the domain of attraction of FN_H with $H \in (0, 0.5)$. For example the proof of Theorem 2.4.2 utilises results from this chapter.

Here a thorough treatment of long-range dependence will be provided including some important and oft-quoted results for which correct proofs are nonetheless difficult or impossible to find.

LRD is related to the asymptotic properties of the ACVF, which is often taken to be *regularly varying*. It is not commonly appreciated, however, that regular variation in the discrete context does not enjoy all the properties of its well developed continuous parent [3]. Therefore the second aim of this chapter is to clearly define the discrete form, state and prove its relevant properties, and clarify its relation to LRD and ASS in general. Since the covariance sum function (S), the variance time function (ω) and the aggregated variance ($\mathcal{V}^{(m)}$) are derived from the ACVF by cumulative summation, the summation of regular varying functions is also discussed.

3.1 Regular variation in discrete time

We must first define the standard concept of 'continuous' regular variation [3, 9].

Definition 3.1.1 (Continuous Regular Variation (CRV))
A function $\widetilde{f}$ defined on $\mathbf{R}^+$ is *regularly varying* at infinity with index α if

$$\lim_{t\to\infty} \frac{\widetilde{f}(tx)}{\widetilde{f}(t)} = x^\alpha, \quad \alpha \in \mathbf{R} \tag{3.1}$$

for every $x \in \mathbf{R}^+$ (it is sufficient that (3.1) be satisfied on a dense subset of $\mathbf{R}^+$, see [9, P. 275]). If $\alpha = 0$ the function f is also said to be *slowly varying* (CSV).

Examples of slowly varying functions include any function with a positive finite limit at infinity, and any real power of $\ln x$.

Let CRV_α denote the set of regularly varying functions defined on $\mathbf{R}^+$ with index α. Some of their main properties are given in the following list. For completeness, proofs are provided in the Appendix.

CRV–(i) $\widetilde{f} \in \mathrm{CRV}_\alpha \Leftrightarrow \lim_{t\to\infty} \widetilde{f}(tx)/\widetilde{f}(t) = x^\alpha$, $\alpha \in \mathbf{R}$, $x \in \mathbf{R}^+$ (Definition)

CRV–(ii) $\widetilde{f} \in \mathrm{CRV}_\alpha \Leftrightarrow \widetilde{f}(x) = \widetilde{s}(x)x^\alpha$, $\widetilde{s}(x) \in \mathrm{CSV}$ (alternative Definition)

CRV–(iii) $\widetilde{f} \in \mathrm{CRV}_\alpha$ and $\widetilde{g} \sim \widetilde{f} \Rightarrow \widetilde{g} \in \mathrm{CRV}_\alpha$

CRV–(iv) $\widetilde{f} \in \mathrm{CRV}_\alpha \Rightarrow \widetilde{f}(x) \sim \widetilde{f}(x + x_0)$, $\forall x_0$ constant.

Now consider a function $f(k)$, $k \in Z^+$. It is significant that in the literature on LRD where regular variation enters, there is rarely if ever any mention of a discrete definition – the properties of CRV are used without question. In fact, there is no discrete definition for which all the properties in the above list hold true. We propose the following one which preserves the majority of them in a natural way. Our choice agrees with the observations of [3, page 52] , and [10] on discrete regularly varying functions.

Definition 3.1.2 (Discrete Regular Variation (DRV))
A function f defined on $\mathbf{Z}^+$ is *regularly varying* at infinity with index α if there exists a $\widetilde{f} \in \mathrm{CRV}_\alpha$ such that $f(n) = \widetilde{f}(n)$ for all $n \in \mathbf{Z}^+$. Similarly to the continuous case for $\alpha = 0$ the function is also called slowly varying (DSV).

By defining DRV via sampling in this way, we guarantee that most of the properties of CRV are directly inherited. Of the above list, only **CRV**–(i) does not carry over exactly. To avoid any possibility of ambiguity, and to provide a convenient reference, we give the corresponding discrete list explicitly:

DRV–(i) $f \in \mathrm{DRV}_\alpha \Rightarrow \lim_{k\to\infty} f(kn)/f(k) = n^\alpha$, $\alpha \in \mathbf{R}$, $n \in \mathbf{Z}^+$

DRV–(ii) $f \in \mathrm{DRV}_\alpha \Leftrightarrow f(k) = s(k)k^\alpha$, $s(k) \in \mathrm{DSV}$ (alternative Definition)

DRV–(iii) $f \in \mathrm{DRV}_\alpha$ and $g \sim f \Rightarrow g \in \mathrm{DRV}_\alpha$

DRV–(iv) $f \in \mathrm{DRV}_\alpha \Rightarrow f(k) \sim f(k + k_0)$, $\forall k_0$ constant.

Although these results are not difficult to prove, we know of no reference for them. Proofs are therefore provided in the Appendix.

The fact that Property **DRV**–(i), in contrast to **CRV**–(i), only acts in one direction, shows that attempting to define DRV by analogy to Equation (3.1) would not have been fruitful. Because k and n take only integer values, far fewer constraints are placed on the function f, allowing much wilder local behaviour. In contrast, Equation (3.1) ensures that, in the limit, the *relative* size of local variations in $\widetilde{f}$ must be negligible.

Property **DRV**–(i) also plays a vital role in the understanding of ASS, as the limit on the right hand side is nothing other than the definition of a power-law type fixed point. Thus, if

f were such a fixed point (with $\alpha \in [0,2]$), the inability to infer that f is DRV clarifies the important fact that *regular variation has no a priori connection to asymptotic self-similarity.*

We can now show that the class of regularly varying autocovariance functions is invariant under (finite level) aggregation.

Theorem 3.1.1 (Persistence of power-law autocovariance functions)
If an autocovariance function satisfies $\gamma(k) = c_\gamma(k)k^\alpha$, $\alpha > 0$, $c_\gamma(k) \in$ DSV then for each fixed $m \in \mathbf{Z}^+$, $\gamma^{(m)}(k) \overset{k}{\sim} (c_\gamma(k)m^\alpha)k^\alpha$, where $\overset{k}{\sim}$ denotes asymptotic equivalence as $k \to \infty$.

Proof
Take Equation (1.10). According to **DRV**–(iv) we can replace each $\gamma(km \pm m \pm i)$ term by $\gamma(km)$. Since m is fixed, so is the number of terms in the summation, and using **Æ**–(iii) from the appendix (page 66) we have

$$\begin{aligned}\gamma^{(m)}(k) &\overset{k}{\sim} \frac{1}{m^2}\left[m\gamma(km) + 2\sum_{i=1}^{m-1} i\gamma(mk)\right] = \frac{1}{m^2}\left[m\gamma(km) + m(m-1)\gamma(mk)\right] \\ &= \gamma(mk) = c_\gamma(mk)m^\alpha k^\alpha \overset{k}{\sim} \left(c_\gamma(k)m^\alpha\right)k^\alpha.\end{aligned}$$

□

Another example of the utility of the discrete definition is that cumulative sums of functions with DRV can be conveniently derived from the corresponding continuous results.

Theorem 3.1.2 (Integration of regularly varying functions)
Let $\widetilde{Z}(t) \in \mathrm{CRV}_\alpha$, and let $\widetilde{L}(t)$ and $\widetilde{U}(t)$ be defined as:

$$\widetilde{L}(t) := \int_0^t \widetilde{Z}(y)\, dy \qquad\qquad \widetilde{U}(t) := \int_t^\infty \widetilde{Z}(y)\, dy.$$

$$\text{(a) If } \alpha \geq -1 \quad \text{then} \quad \frac{t\widetilde{Z}(t)}{\widetilde{L}(t)} \to (1+\alpha), \quad \text{and } \widetilde{L} \in \mathrm{CRV}_{\alpha+1}.$$

$$\text{(b) If } \alpha < -1 \quad \text{then} \quad \frac{t\widetilde{Z}(t)}{\widetilde{U}(t)} \to -(1+\alpha), \quad \text{and } \widetilde{U} \in \mathrm{CRV}_{\alpha+1}.$$

A proof can be found in [19, page 17], and also [9, pages 279–281] □.

Theorem 3.1.3 (Summing Discrete Regular Varying Functions)
Let $K(n) \in \mathrm{DRV}_\alpha$, and let $L(t)$ and $U(t)$ be defined as

$$L(m) := \sum_{n=0}^{m-1} K(n), \qquad\qquad U(m) := \sum_{n=m}^{\infty} K(n).$$

$$\text{(a) If } \alpha \geq -1 \quad \text{then} \quad \frac{mK(m)}{L(m)} \to (1+\alpha), \quad \text{and } L \in \mathrm{DRV}_{\alpha+1}.$$

$$\text{(b) If } \alpha < -1 \quad \text{then} \quad \frac{mK(m)}{U(m)} \to -(1+\alpha), \quad \text{and } U \in \mathrm{DRV}_{\alpha+1}.$$

The proof is left to the Appendix.

The two theorems above essentially tell us that regular variation is preserved under integration. Unfortunately the same is not true when differencing a function or sequence. Regular variation ensures that the local variations are, compared to the function values, small enough not to significantly influence the asymptotic behaviour of the sum. During differencing it is not the relative size, but the rate of change which matters. The value of the function vanishes, and local variations determine the nature of the difference. There is no guarantee that they are well-behaved in any sense. Assuming that RV (CRV or DRV) is preserved under differencing is another misconception one finds in the applied literature (see e.g. [13]). This is perhaps because it does hold in certain special cases.

3.2 Review of long-range dependent processes

3.2.1 Definitions of long-range dependence

Aggregation limit point based definitions

In any reasonable sense, processes in the FN_H subclass with $H \in (0.5, 1]$ are long-range dependent. In keeping with our aggregation based approach to processes, a very broad and natural definition for LRD is therefore given by

Definition 3.2.1 (LRD)
Long-range dependent processes are those in the union of the DoA of FN_H with $H \in (0.5, 1]$.

This definition will now be compared to the following alternatives from the literature.

Autocovariance function based definitions

The definitions in this part are based on the shape of the autocovariance function.

Definition 3.2.2 (LRD1, Slow power-law decrease)
LRD1 processes are those whose ACVFs obey $\gamma(k) \sim c_\gamma k^{2H-2}$, $H \in (0.5, 1)$, $c_\gamma \in \mathbf{R}^+$.

This definition is the most frequently encountered one. For example it is used in [6, 2, 1].

Definition 3.2.3 (LRD2, Slow regular varying decrease)
LRD2 processes are those whose ACVFs obey $\gamma(k) = c_\gamma(k)k^{2H-2}$, $H \in (0.5, 1)$, $c_\gamma \in \mathrm{DSV}$.

This choice generalises LRD1 in a natural way by replacing the constant c_γ, a particular slowly varying function, with a general DSV function. It often appears with $\sim$ replacing $=$, but according to **DRV**–(iii) this is superfluous. In the literature LRD2 is often introduced but then immediately specialised to LRD1 (eg. [6, 15]).

A number of additional 'alternatives' to LRD2 can be generated by using different and, we would argue, flawed definitions of DRV. We choose not to discuss such cases. Since LRD1 $\subset$ LRD2 we will not discuss LRD1 further.

Definition 3.2.4 (LRD3, Infinite sum)
LRD3 processes are those whose covariance sums obey $S_\infty = \infty$.

This definition, used for example in [20], nicely captures the idea of LRD, when the sum of the past has a strong impact.

Definitions based on estimator convergence rate

For independent samples the variance of the mean estimate is asymptotically equivalent to C/m, where $C > 0$ is a constant. A process is considered long-range dependent if the variance converges to zero at a slower rate than $O(1/m)$.

Definition 3.2.5 (LRD4, Slow convergence)
A process will be called LRD4 if $m\mathcal{V}^{(m)} \to \infty$.

.

This is the basic definition that can be found in [12] and [11].

Definitions 3.2.2 to 3.2.3 can and will be compared to our definition (3.2.1) in section 3.2.2.

Before proceeding to the comparison a few extension to LRD4 will be shown. The following definitions are not directly comparable to ours.

The first approach to extend Definition 3.2.5 is given by Heyde and Yang ([12]). They provide a definition of long-range dependence which applies for non-stationary time series and also for processes without second moments. Their definition reduces to 3.2.5 in case of second-order stationary processes.

Definition 3.2.6 (LRD4a, LRD for non-stationary processes)
Assume that the process $X(t)$ has zero mean and $\forall t \in \mathbf{Z} : \mathbf{E}[X(t)^2] \neq 0$ exists. Now define

$$Y^{(m)}(t) := \frac{\sum_{l=tm-m+1}^{tm} X(l)}{\sum_{l=tm-m+1}^{tm} \mathbf{E}[X^2(l)]}.$$

The process $X(t)$ is LRD4a if

$$\left(\sum_{l=tm-m+1}^{tm} \mathbf{E}[X^2(l)] \right) \mathrm{Var} Y^{(m)}(t) \to \infty, \tag{3.2}$$

as $m \to \infty$.

Although the authors didn't specify explicitly it is assumed that Equation (3.2) applies for all $t \in \mathbf{Z}$.

In this definition $Y^{(m)}(t)$ corresponds to $X^{(m)}(t)$, the aggregated process. Because the variance of the samples differ the sum of the $X(t)$ values is normalised by the sum of their variances and not simply by the number of the aggregated samples, (m). It can easily be justified that in the case when all variances are equal this definition reduces to LRD4.

This definition can further be extended to include processes without second moments. In this case the variance is replaced by the empirical variance.

Definition 3.2.7 (LRD4b, LRD for processes without finite second moments)
A process $X(t)$ is LRD4b if

$$\frac{(\sum_{l=tm-m+1}^{tm} X(l))^2}{\sum_{l=tm-m+1}^{tm} X(l)^2} \xrightarrow{p} \infty, \tag{3.3}$$

as $m \to \infty$, where $\xrightarrow{p}$ means convergence in probability.

In case of stationary processes the expected value of the left-hand side of (3.3) is $m\mathcal{V}^{(m)}$, so it is an unbiased estimator for it.

The other direction to extend, or rather restate Definition 3.2.5 is given by Hall ([11]). Definition 3.2.5 was based on the convergence rate of the mean estimate . Hall argues that the mean is just one of the properties one might be interested in estimating, and so there is no particular reason that long-range dependence should be linked to the convergence speed of the mean estimation. According to Hall long-range dependence is not a property of the time series itself, but of both the time series and the quantity being estimated. He suggests to define a process LRD if the statistics to be analysed converges at a slower rate than it would in the context of independence.

Definition 3.2.8 (LRD4x, Application oriented definition)
The time series $X(t)$ is LRD4x if and only if the mean square convergence rate of the estimator is slower than in the context of independence.

To be consistent we suggest to complement this definition with another one which declares time series to be *very short range dependent* if the mean square convergence rate is faster than in the case of independence. Such behaviour might arise in the case of anti-correlated samples, like for example for FN_H, with $H < 0.5$.

The author of [11] presents an example of a non-parametric density estimation. According to his definition in this context processes that satisfy $\gamma(k) \sim ck^{2H-2}$ are long-range dependent if $H > 3/5$.

3.2.2 Properties of LRD processes

This subsection deals with the nature of LRD2 and LRD3, their relationship to our definition of LRD, and related results.

The following four properties will be proved and discussed. Each of the last three provide classes of examples of LRD processes, as promised in Section 2.4. Recall that $\mathcal{V}^{(m)}$ stands for the variance of the m-aggregated process.

LRD–(i) For LRD2 processes, $\mathcal{V}^{(m)} \sim \frac{c_\gamma(m)m^{2H-2}}{H(2H-1)}$.

LRD–(ii) LRD2 $\subset$ LRD.

LRD–(iii) LRD2 $\neq$ LRD.

LRD–(iv) LRD2 $\subseteq$ LRD3.

LRD–(v) LRD3 $\subseteq$ LRD4.

Proof of Property LRD–(i)
Define $S_1(i) := \sum_{k=0}^{i-1} \gamma(k)$. According to Theorem 3.1.3, $S_1 \in \mathrm{DRV}_{2H-1}$ and $S_1(i) \sim \frac{c_\gamma(i)}{2H-1} i^{2H-1}$. It is easy to see that because $2S_1(i) \to \infty$, $2S_1(i) \sim 2S_1(i) - \mathcal{V} + \gamma(i) = S(i)$. Since Æ is transitive (see **Æ**–(i) in Appendix), we have

$$S(i) \sim 2\frac{c_\gamma(i)}{2H-1} i^{2H-1},$$

which by **DRV**–(iii) implies $S(i) \in \mathrm{DRV}_{2H-1}$. Now applying Theorem 3.1.3 again for S, we obtain $\omega(m) = \sum_{i=0}^{m-1} S(i) \sim \frac{c_\gamma(m)}{H(2H-1)} m^{2H}$, and because of (1.6) and **Æ**–(ii):

$$\mathcal{V}^{(m)} \sim \frac{c_\gamma(m)}{H(2H-1)} m^{2H-2},$$

so $\mathcal{V}^{(m)}$ is also regularly varying with index $2H-2$. □

This result appears in surprisingly few places, as typically the slowly varying prefactor, $\frac{c_\gamma(m)}{H(2H-1)}$, is omitted. One finds $\mathcal{V}^{(m)} \sim m^{2H-2}$ (implying an abuse of the Æ notation). Knowledge of the slowly varying part is however essential for purposes such as the determination of confidence intervals for mean estimates. The result for the special case of $c_\gamma(k)$ a constant is stated in [2], although not proved there, and in many places it is given with the prefactor omitted. Often [6] is cited in lieu of proof, although no formal proof appears there. Invalid proofs, and proofs relying on unproved lemmas, are not difficult to find.

Proof of Property LRD–(ii)
In the proof of **LRD**–(i) it was shown that $\omega \in \mathrm{DRV}_{2H}$ for processes in LRD2, so by **DRV**–(i)

$$\lim_{m\to\infty} \frac{\phi(nm)}{\phi(m)} = \lim_{m\to\infty} \frac{\omega(nm)}{\omega(m)} = n^{2H}$$

for every $n \in \mathbf{Z}^+$, so that ϕ is in the DoA of FN_H. □

The simplicity of the proof is owed to the natural and complementary nature of the definitions of LRD and of DRV, together with the directness of the CTF approach, based on ϕ. Proofs in the literature are typically based on the ACF approach to ASS. The resulting complexity creates many opportunities for subtle errors.

Proof of Property LRD–(iii)
To show that LRD2 is a strict subset of LRD, a process will be constructed that is a member of LRD\LRD2. Let X_1 and X_2 be independent copies of an FN_H process with ACVF γ^*, $H \in (0.5, 1]$. We define $Y(t)$, $t \in \mathbf{Z}$ by deterministically alternating between the two copies:

$$Y(t) := \begin{cases} X_1(t/2), & t \text{ even} \\ X_2(\frac{t-1}{2}), & t \text{ odd} \end{cases}$$

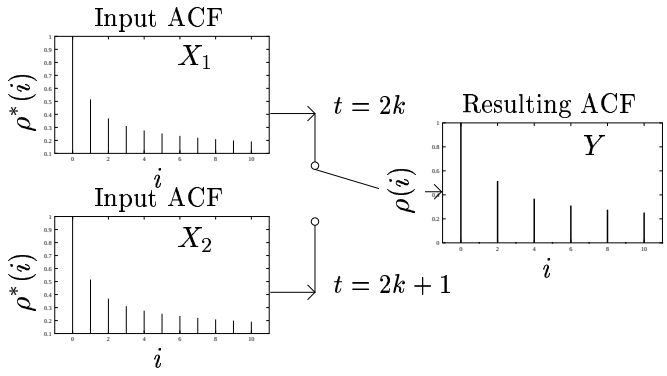

Figure 3.1: Example of a process in LRD\LRD2. Odd lags compare different processes and are uncorrelated.

as illustrated in Figure 3.1. This process will be called *alternating fractional noise* (AFN). It is not difficult to see that Y is a second-order stationary process, with ACVF

$$\gamma_Y(k) = \begin{cases} \gamma^*(k/2), & k \text{ even} \\ 0, & k \text{ odd} \end{cases},$$

and so

$$\omega_Y(m) = \begin{cases} 2\omega^*(m/2), & m \text{ even} \\ 2\omega^*((m-1)/2) + S^*((m-1)/2), & m \text{ odd} \end{cases}. \tag{3.4}$$

If m is even, then it follows trivially that Equation (2.19) is satisfied for ω_Y for any n. If m is odd, then recalling that $\omega(m) = \omega(m-1) + S(m-1)$ and exploiting the monotonicity of ω^*, the asymptotically equivalent upper and lower bounds of $\frac{\omega_Y(nm)}{\omega_Y(m)}$ given in

$$\underbrace{\frac{\omega^*(n\frac{m-1}{2})}{\omega^*(\frac{m-1}{2})}}_{\downarrow n^{2H}} \underbrace{\frac{\omega^*(\frac{m-1}{2})}{\omega^*(\frac{m+1}{2})}}_{\downarrow 1} = \frac{\omega^*(n\frac{m-1}{2})}{\omega^*(\frac{m+1}{2})} \leq \underbrace{\frac{\omega_Y(nm)}{\omega_Y(m)}}_{\downarrow n^{2H}} \leq \frac{\omega^*(n\frac{m+1}{2})}{\omega^*(\frac{m-1}{2})} = \underbrace{\frac{\omega^*(\frac{m+1}{2})}{\omega^*(\frac{m-1}{2})}}_{\downarrow 1} \underbrace{\frac{\omega^*(n\frac{m+1}{2})}{\omega^*(\frac{m+1}{2})}}_{\downarrow n^{2H}}$$

confirm that Equation (2.19) is again satisfied.
□

In this example the monotonicity of the CTF of FN_H was exploited. With a slightly more sophisticated proof any LRD process could have been selected, showing that the set LRD\LRD2 is quite large.

The question whether LRD2 is a strict subset of LRD is not usually asked explicitly, however the opposite is sometimes implicitly assumed. Again, this can be linked to the misconception that **CRV**–(i) applies, and/or an unjustified differencing of a regularly varying CTF.

The definition LRD3 raises new issues, as using the value of S_∞ to define a class can be viewed as a shift from a pointwise definition of convergence under aggregation, to one based

on integrals of γ, of 'weak' convergence. We first offer a simple result on the behaviour of this alternative.

Theorem 3.2.1
Assume S_∞ exists. If $S_\infty < \infty$ then $S_\infty^{(m)} = \frac{S_\infty}{m}$. If $S_\infty = \infty$ then $S_\infty^{(m)} = \infty$.

Proof
The statement is a direct consequence of (1.9). □

Thus the three classes of S_∞ zero, finite non-zero, or infinite, are invariant under (finite level) aggregation. We now prove Property **LRD**–(iii).

Proof of Property LRD–(iv)
From the proof of Property **LRD**–(ii), $S_\infty = \infty$ for processes in LRD2, hence LRD2 $\subset$ LRD3. □

It is tempting to conclude that, since by Theorem 3.2.1 $S_\infty = \infty$ is preserved under aggregation, and since $S_\infty = \infty$ for each of the LRD fixed points, but no others, that LRD3 $\supset$ LRD, i.e. that every LRD process has infinite covariance sum. Similarly it may seem obvious that no process with $S_\infty = 0$ can be LRD, or that no SRD process can have $S_\infty = \infty$. However none of these statements necessarily hold, as the pointwise convergence of $\rho^{(m)}$ does not control the finiteness or otherwise of the $S_\infty^{(m)}$. Not surprisingly however, under additional conditions such as regularly varying ACVF, they can be made to hold.

Proof of Property LRD–(v)
In Lemma 2.4.1 it was shown that if S_∞ is finite then $m\mathcal{V}^{(m)} \to S_\infty$. The proof of this theorem is very similar and therefore not repeated. □

The question whether the above theorem can be reversed is not trivial. It is known that when S_∞ is finite then $\lim_{m\to\infty} m\mathcal{V}^{(m)}$ is also finite. The question is whether there exists a process for which S_∞ does not exist but $\lim_{m\to\infty} m\mathcal{V}^{(m)} = \infty$. From Equation (1.2) it is clear that $m\mathcal{V}^{(m)} = \omega^{(m)}/m$ is a "cumulative average" of $S(i)$ terms. Assume that $S(i)$ essentially converges to ∞, but a few zero terms are "inserted" every one and then. If the frequency of these zero terms is low enough then it wouldn't change the overall behaviour of the average. However one should not forget that the function S should correspond to a positive semi-definite autocovariance function. This issue makes this question really complicated and its investigation is out of the scope of this document.

3.3 Conclusions

Discrete time regular variation was defined, its basic properties stated and proved based on the corresponding results from continuous time regular variation. This has enabled a thorough and precise analysis of long-range dependent processes. Different definitions found in the literature for LRD have been presented and compared to each other as well as to a novel definition. This definition was proposed to capture the idea of being in the domain of attraction of FN_H with $H \in (0.5, 1)$.

The relation of the different LRD definitions is depicted in Figure 3.2.

The two main contributions presented in this chapter are the following:

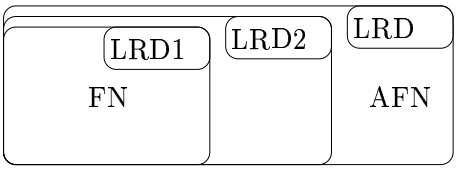

Figure 3.2: Relation of the different long-range dependent definitions, LRD, LRD1 and LRD2. The two examples fractional noise (FN) and alternating fractional noise (AFN) are depicted in the figure.

1. It was proved that if the autocovariance function satisfies $\gamma(k) = c_\gamma(k)k^{2H-2}$, with $H \in (0.5, 1)$ and c_γ slowly varying, then the sequence of aggregated variances satisfies $\mathcal{V}^{(m)} \sim \frac{c_\gamma(m)m^{2H-2}}{H(2H-1)}$. (Property **LRD**–(i))

2. It was shown that processes for which $\gamma(k) = c_\gamma(k)k^{2H-2}$, with $H \in (0.5, 1)$ form a strict subset of processes that converge to FN_H with the corresponding H. (Property **LRD**–(iii))

Chapter 4

Direct consequences

The results presented in the previous chapters contributed to the understanding of self-similarity and long-range dependence from a mathematical point of view. In this chapter the practical implications of the new results will be highlighted focusing on estimating the parameters of long-range dependence. Estimation methods can be analysed from different points of views. Their statistical properties like efficiency and bias are important to know and are commonly investigated. Also issues related to the calculation method, like finding the appropriate region, or cut-off point where the graph should be analysed is important. These issues however will not be treated here. This chapter focuses on some theoretical aspects. In the view of the new results presented in the previous chapters it will be investigated for what type of processes these estimators are suitable, and what conclusions may be drawn from their results, even if one assumes that all the practical issues mentioned above are perfectly solved.

4.1 Estimating long-range dependence

Several tests and estimators have been developed to test the presence and estimate the parameters of long-range dependence in a measured time series. An extensive overview of these methods can be found in [2] and also in [22]. These tests include both parametric and semi-parametric tests.

Full parametric maximum likelihood estimators (MLE) or approximate maximum likelihood estimators are not affected by the new results. These estimators assume a complete characterisation of the covariance structure and so it has to be known in advance what type of autocorrelations may appear.

The heuristic, non-parametric tests are based on some property of long-range dependent processes. The **aggregated Whittle** method for example is a non-parametric extension of the Whittle estimator that is based on the asymptotic self-similarity of LRD processes. The **periodogram plot** operates in the frequency domain and analyses the behaviour of the spectral density function near the origin. The **variance time plot** method is a simple, non-parametric time-domain investigation method that is based on the behaviour of the aggregated variance, as described by Property **LRD**–(i). A more detailed and extensive overview of the various LRD tests and estimators can be found in [2].

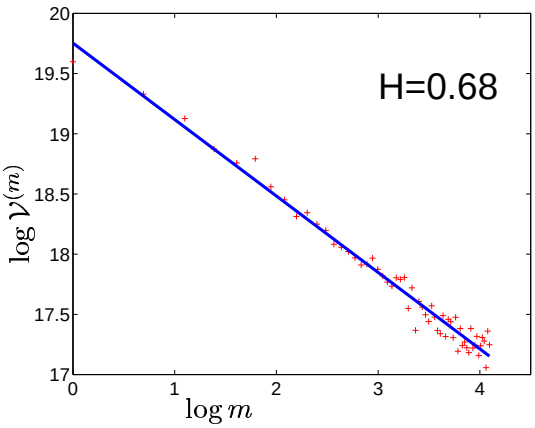

Figure 4.1: Hurst parameter estimation using the Variance Time Plot method

The variance time plot method is selected to demonstrate the effects of the new results. Although this method is neither very reliable nor exact it is simple but at the same time suitable to demonstrate the effects of the new results on LRD estimation.

Variance time plot is based on **LRD**–(i). For LRD1 processes

$$\mathcal{V}^{(m)} \sim \frac{c_\gamma m^{2H-2}}{H(2H-1)}.$$

For a measured trace $\mathcal{V}^{(m)}$ is estimated and plotted against m on a log-log scale. If the trace was generated by an ergodic LRD1 process (and it is long enough) the values at tail of the plot should align a straight line with slope $2H-2$. The level of the straight line indicates the value of c_γ. So first the straightness of the tail is judged. If the tail is found to be straight then its slope is measured and H is calculated. This is depicted in Figure 4.1.

Apart from some practical considerations, like finding the lower cut-off scale for the line fitting, or determining the confidence interval of the result, this estimator can be used as described for LRD1 processes. Care should however be taken with the interpretation of the result when the assumption of LRD1 is not verified.

Case I Assume that the time series under investigation was generated by an LRD2, but not LRD1 process. In this case regardless of the length of the sample the tail of the slope does not converge to a straight line. This estimator is not suitable to detect the presence of LRD2.

Case II Variance time plot investigates LRD indirectly through the sequence of aggregated variances. It has to be noted however that **LRD**–(i) can not be reversed, that is a regular variation of the aggregated variance does not ensure the regular variation of the autocovariance function. Therefore the straightness of the slope is not sufficient in itself to conclude the presence of LRD1.

Now a simple scenario taken from the field of computer networking will be shown to demonstrate how **Case II** might arise in a practical situation.

Similarly to the proof of Property **LRD**–(iii) two independent stochastic processes will be mixed. Assume that two independent traffic streams arrive at a deterministic multiplexer

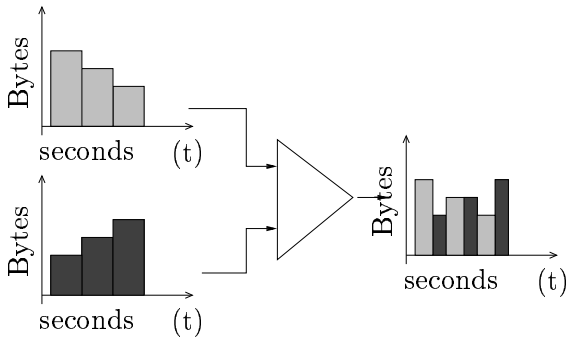

Figure 4.2: Deterministic multiplexing of two independent traffic streams.

as shown in Figure 4.2. In this simple scenario the data rate on the two inputs are the same and the output rate is the double of the input rate. We assume that the two processes at the inputs are both LRD1 with the same parameters (H, c_γ). Although from a theoretical point of view it is not likely that such a case appears, in practice it is sufficient that the corresponding parameters are "close enough" to each other, i.e. their differences are significantly smaller than the "resolution" of the estimator. In fact this estimator already makes a strong assumption, namely the stationarity, which is at least that questionable from a theoretical point of view.

The output process will not be LRD1 or LRD2 according to Definitions 3.2.2 and 3.2.3 respectively, since $\gamma(k) = 0$ for odd values of k. Now let $\mathcal{V}^{(m)}$ be estimated by calculating the empirical variance of the sequence

$$Z_i = \frac{\sum_{j=i}^{i+m-1} X(j)}{m},$$

for $i = 1, 2, \ldots, n-m+1$, where n is the sample size. If the estimated value of $\mathcal{V}^{(m)}$ is plotted against m on a log-log scale then a straight line appears on the slope even though the traffic intensity is not LRD1.

4.1.1 Other estimation methods

Periodogram method

This estimation method operates in the frequency domain. It is based on the behaviour of the spectral density near the origin. According to the Tauberian theorems ([3]) a regularly varying autocovariance function yields a regularly varying spectral density at the origin[1]. If the autocovariance function had a simple power-law decay (i.e. LRD1 process) then the spectral density will also follow a power-law near the origin.

Therefore if the process is assumed to be LRD1 then the spectral density is estimated, and plotted on a log-log diagram. If the process is LRD1 then the values should align a straight line with slope $1 - 2H$, as depicted on Figure 4.3.

[1]Regular variation at a finite value can be defined similarly to regular variation at infinity. The limit at ∞ should be replaced by the finite value in Equation (3.1)

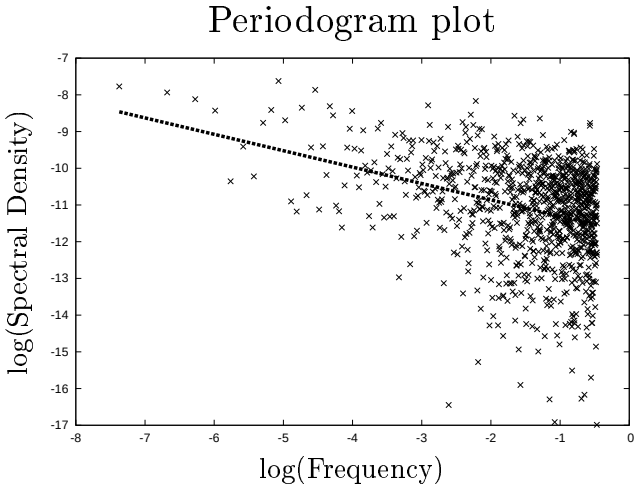

Figure 4.3: Periodogram plot of $FN_{0.7}$ dataset

This estimation method has the same limitations as the variance time plot method, since it works only for simple power-law LRD processes, and since the spectral density is calculated by weighted cumulative summation of the autocovariance coefficients, the presence of power-law relationship in the lower region of the spectral domain does not guarantee a regularly varying autocovariance function.

Wavelet based AV estimator

The wavelet based AV estimator [24] is based on the frequency domain behaviour of LRD1 processes near the origin. Therefore similarly to the previous methods the AV estimator also assumes "simple" LRD1 behaviour. In practice, however, based on the type of the wavelet this estimator is less sensitive to deviations from pure power-law ACVF structure.

Aggregated Whittle estimator

The Whittle estimator is an approximation to the parametric exact maximum likelihood estimator [2]. The approximation is used to reduce the computational complexity, and so the time needed to calculate the result.

Long-range dependent processes are in the domain of attraction of fractional noise. Therefore for "sufficiently large" m the m-aggregate of the process can be approximated by fractional noise and using the parametric Whittle estimator the parameters of long-range dependence (for example H) can be estimated. The choice of m is an important practical issue not investigated here.

Case I Since LRD2 processes also converge to the fractional noise this estimation method works not only for LRD1, but also for the wider set of LRD2 processes. There is however a difference. For LRD1 processes the so-called scale parameter, c_γ of Definition 3.2.2 can be estimated since it can be calculated using H and the variance of the aggregated process which can both be estimated. In contrast, the difficult task of estimating the whole function $c_\gamma(k)$ for LRD2 processes is not possible using this method.

Case II Since this method only analyses the aggregated process and aggregation cannot be reversed, it is not possible to infer details of the original process. Without any prior knowledge of the process this method can only be used to give the value of H, provided the process is in the domain of attraction of any fractional noise at all.

4.2 Conclusion

In this chapter some possible hazards were identified that can occur during the interpretation of the results of LRD estimators.

Estimation methods are usually based on the behaviour of some function that can be calculated by (weighted) cumulative sum of the autocovariance function, like the variance time function, or the spectral density. However it is not possible to conclude asymptotic properties of the autocovariance function, based on the asymptotic properties of these cumulative sums. This is similar to the issue of differencing regular varying functions.

There are also some estimation methods that were developed for LRD1 processes, where the slowly varying factor converges to a constant. These estimation methods are usually not suitable for general LRD processes, or even general LRD2 processes where the slowly varying factor is not constrained.

One should however not conclude that the presented estimation methods are wrong. It must be checked whether the required assumptions are satisfied and care should be taken with the conclusions based upon the results of the estimators.

Chapter 5

Further consequences

In the previous chapter several examples have been shown where the new results are directly applicable mainly during the interpretation of the outcomes of already existing estimators. We believe though that the implications of these results are more fundamental.

The results presented in this work provide solid base, proper tools and an approach, which we believe is the appropriate one in many cases for the study of second-order processes over different time-scales. These together, we believe, will foster the development of new algorithms.

This chapter is dedicated to illustrate the advantages of this novel approach through general considerations and simple examples.

5.1 General Considerations

Let us consider a system that is fed with a time series for example a packet buffer in a packet switched network. The behaviour of the system is influenced by the statistics of the time series and is in most cases significantly different when fed with LRD input compared to the case of independent or weakly correlated input series.

This different behaviour however is not triggered because $\gamma(k)$ is "large" for a selected set of "large" k-s, but it is attributed to the scaling nature of the input, which means that even on larger time scales the aggregated series does not smooth out, it remains strongly correlated. Thus the system behaviour is not determined by how the autocovariance function behaves, but rather how the covariance time function behaves.

Also as we have seen in Section 4.1 most of the estimators of long-range dependence operate with the variance or covariance time function or with some other function that involves summation of the autocovariance and not directly with the autocovariance.

Now recalling that aggregation can much simpler be described using ω than γ we believe that while the ACVF is useful to describe processes at a given time scale, the scaling nature of a process can more naturally be described and analysed using the VTF.

5.2 Analysing the speed of convergence, estimating $\gamma^{(m)}$

Another issue which illustrates the advantage of the above results is analysing the speed of convergence of asymptotically self-similar time series. Here we will concentrate on the domain of attraction of the fractional noise and will further assume that the processes under analysis belongs to a subset of the domain of attraction, where ω is restricted to

$$\omega \sim Cm^{2H} \tag{5.1}$$

with $C > 0$ and $0.5 < H < 1$.

One might be interested in how fast the correlation structure (ACF or CTF) converges to its asymptotical limit (either pointwise or using some other distance metric) or how well does the aggregated variance correspond to the value given by the asymptotical formula as given in **LRD**–(i) or how well can one predict $\gamma^{(m)}(k)$ or $\omega^{(m)}(k)$ for given values of m and k from the asymptotical descriptors of the process: C and H of (5.1).

Here we will focus mainly but not exclusively on the last question.

If the constants C and H are known (or estimated) it is meaningful to estimate $\omega^{(m)}(n)$ and $\gamma^{(m)}(k)$ as $C(mn)^{2H}/m^2$ and $C\mathbf{D}\{(mn)^{2H}/m^2\}$ respectively. These values for $\omega^{(m)}$ and $\gamma^{(m)}$ will be called the asymptotical estimates because they are estimated using the asymptotical descriptors.

Two LRD process with the same C and H parameters will now be compared to each other. The processes X_g and X_w will be defined by their autocovariance functions as:

$$\gamma_g(k) := \begin{cases} \mathbf{D}\{i^{1.6}\}(k/2) & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd} \end{cases}$$

and

$$\gamma_w(k) := \begin{cases} 2^{-0.6}\mathbf{D}\{i^{1.6}\}(k) & \text{for } k \neq 0 \\ \gamma_g(0) & \text{for } k = 0 \end{cases}$$

That such processes exist can be justified as follows: During the proof of property **LRD**–(iii) a process similar to X_g was constructed, so its existence is therefore guaranteed. As far as X_w is concerned first it has to be noted that

$$\gamma(k) := 2^{-0.6}\mathbf{D}\{i^{1.6}\}(k) \tag{5.2}$$

is the ACVF of a fractional noise, so it is positive semi-definite. It is not difficult to check that $\gamma_g(0) = 1 > 2^{-0.6}$, so the process X_w can be constructed by adding a white noise process of variance $\mathcal{V} = \gamma_g(0) - 2^{-0.6}$ to the fractional noise process, since adding two processes results in the summation of their ACVFs.

The "g" in X_g refers to the periodic "gap" in the ACVF, while the "w" in X_w indicates that a white noise process has been added to a fractional noise process. The constants have been selected such that $\omega_g \sim \omega_w$ ensuring the same asymptotical parameters and so the same limit point. For comparison we also define the pure fractional noise process X, with ACVF $\gamma(k)$ as defined in Equation (5.2), which also satisfies $\omega \sim \omega_g$. For all three of these processes $C = 2^{-0.6}$ and $H = 0.8$.

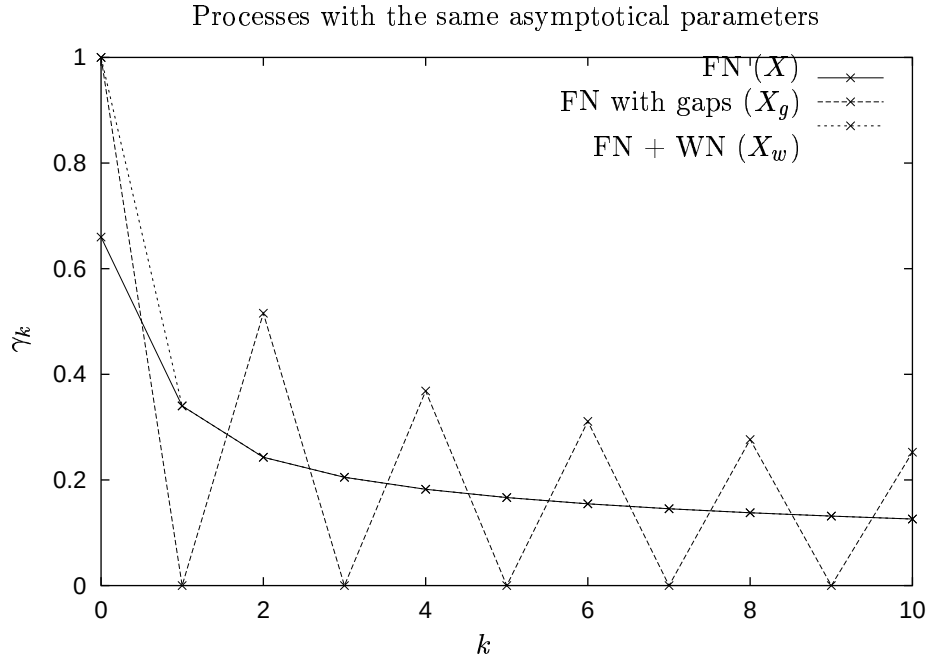

Figure 5.1: Autocovariance structures of three different processes which have the same asymptotical parameters C and H, so $\gamma^{(m)}(k) \overset{m}{\approx} \gamma_g^{(m)}(k) \overset{m}{\approx} \gamma_w^{(m)}(k)$ and also $\omega^{(m)}(k) \overset{m}{\approx} \omega_g^{(m)}(k) \overset{m}{\approx} \omega_w^{(m)}(k)$. The autocovariances are significantly different.

Figure 5.1 shows the first few lags of γ_g, γ_w and γ, while in Figure 5.2 we see the autocovariances of the 5-aggregated versions of the same three processes.

For every single lag k: $|\gamma(k)-\gamma_w(k)| \leq |\gamma(k)-\gamma_g(k)|$, that is the autocovariance coefficients of X_w are closer to that of X. However it takes only 5 level of aggregation (see Figure 5.2) and $\gamma_g^{(5)}$ approaches the value suggested by the asymptotical parameters ($\gamma^{(5)}$) reasonably well, why for the same level of aggregation $\mathcal{V}_w^{(5)} = \gamma_w^{(5)}(0)$ is still far from the asymptotical value. This little example illustrates that the shape of the ACVF does not have a direct impact on its convergence speed.

The shape of the variance time function on the other hand has a much more direct impact on the speed of the convergence. Since $\omega^{(m)}(n)$ can be expressed as $\omega(mn)/m^2$ it is clear that for the variance time function aggregation means selecting higher lag values from the same function. So if for two processes X_1 and X_2, which satisfy $w_1 \sim w_2$ there exists a fractional noise process X^* such that $\omega^* \sim w_1$ then if say $\omega_1(n)$ converges faster to $\omega^*(n)$ in any reasonable sense than $\omega_2(n)$ then so will $\omega_1^{(m)}(n)$ converge faster in the same sense to $\omega^{(m)}(n)$ for n fixed as $m \to \infty$.

Another important point to note is that if one expresses $\gamma^{(m)}(k)$ in terms of $\omega(n)$ then higher lags (k) of $\gamma^{(m)}(k)$ require higher lags of ω, so the higher the value of k the closer $\gamma^{(m)}(k)$ is to the asymptotical value. This means that it will be the variance that has the slowest convergence. So as we could see in Figure 5.2 for the process $X_w^{(m)}$ all autocovariance

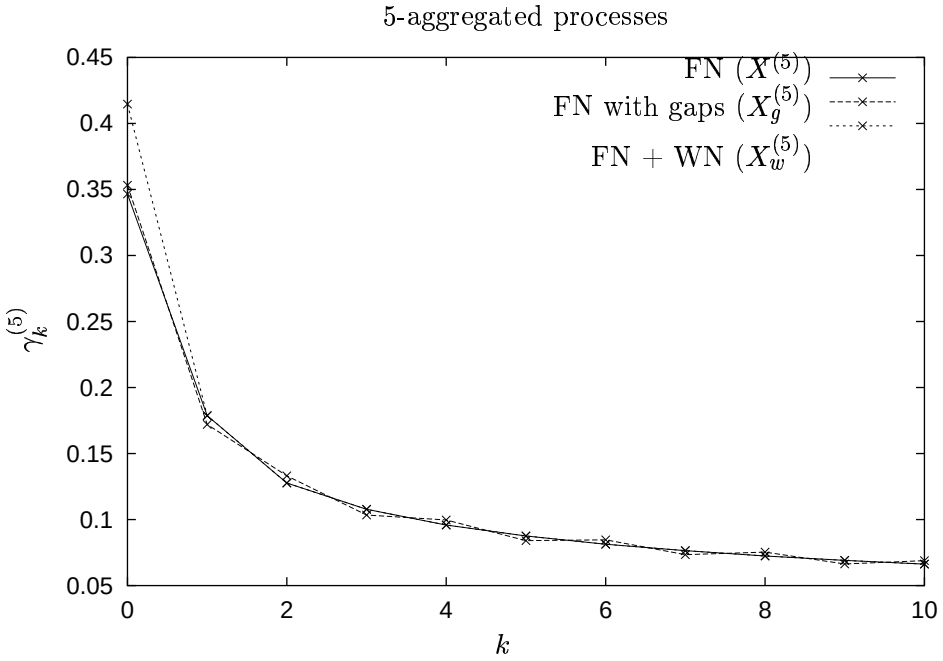

Figure 5.2: The 5-aggregated versions of γ, γ_g and γ_w. Although the difference of the individual lags of γ_g and γ were much bigger than the difference between the corresponding lags of γ_w and γ the difference between $\gamma^{(5)}$ and $\gamma_g^{(5)}$ has vanished, these functions are almost indistinguishable, while the difference of $\gamma^{(5)}(0)$ and $\gamma_w^{(5)}(0)$ is still significant.

coefficients are close to their asymptotic estimate except for the lag 0 value. Calculating the ACF is achieved by normalising by the variance, therefore the slow convergence of the variance to the asymptotic value spoils the convergence of the ACF for every lag k, $k \neq 0$.

It has to be noted that the process X_w is just one example, where the variance has a very slow convergence rate compared to all other lags. In other cases the difference of the convergence speed might be less, but, aside from very extreme examples, the variance will always have the slowest convergence of all values of the ACVF.

The processes in the example might seem artificial but they clearly show the difference between the ACVF and VTF approach which is process independent. In the next subsection the same approach will be used to analyse an other set of processes. The convergence speed of the widely used fARIMA processes will be investigated, to show how a new approach can reveal unknown details of this, otherwise well known, class.

5.2.1 Convergence speed of fARIMA processes

Another example which illustrates the benefits of the variance time function approach is the case of fARIMA processes ([5]). Without going into details the following properties of fARIMA processes are mentioned: fARIMA processes are used to generate models of asymptotically self-similar processes with known covariance structure. fARIMA models offer

much more flexibility than fractional noise since in addition to the asymptotic behaviour of the variance time function one can also "modulate" the process with the autoregressive and moving average coefficients. One of the main usage of such processes is testing estimators and algorithms developed for long-range dependent processes to see how these algorithms perform for "general" ie. non FN long-range dependent processes ([22]). Since the covariance structure of the processes is exactly known it is possible to compare the actual output of the algorithm with the desired one.

The goal of this section is to investigate how fARIMA processes behave under aggregation, how much do they differ from the fractional noise in this regard. A single member of the family of fARIMA processes is selected for investigation, but it will be sufficient to show general characteristics of the whole family. Similarly to the previous subsection 3 processes, a fractional noise (X) and two other processes will be compared to each other. One of these processes (X_f) will be a fARIMA defined by the following equation:

$$\Phi(B)\nabla^d X_f(t) = \Theta(B)Z(t),$$

where B is the backward shift operator, ∇ is the fractional differencing operator, $d = H - 1/2 = 0.3$, $\Phi(B) = 1 - 0.3B$, $\Theta(B) = 1 - 0.7B$ and $Z(t)$ is a unit variance white noise. This process can also be regarded as an ARMA process driven by a so-called fractionally integrated noise ($Y(t)$):

$$X_f(t) - 0.3X_f(t-1) = Y(t) - 0.7Y(t-1).$$

(See [5] for more details on defining ARMA and fARIMA processes.)

The other process will be a sum of the reference fractional noise process (X) and a short-range dependent process (X_s) of ARMA type, defined as

$$X_s(t) - 0.5X_s(t-1) - 0.2X_s(t-2) + 0.1X_s(t-3) = Z(t) + 0.4Z(t-1) - 0.8Z(t-2) - 0.2Z(t-3),$$

where $Z(t)$ is a unit variance white noise. The actual parameters selected for the ARMA process are not special in any way, and in fact their choice does not influence the qualitative behaviour of the aggregated processes.

All three of these processes (X, X_f and $X' := X + X_s$) have the same H parameter and have been scaled to have $C = 1$. (See Equation 5.1.) It has to be noted that the existence of the parameter C (i.e. it is a constant and not a slowly varying function) in itself shows that the process behaves very similarly to the fractional noise. Our investigations show, that this behaviour is common to all members of the fARIMA family. From this we can conclude that fARIMA can only be used to model a small subset of all processes in the domain of attraction of the fractional noise.

fARIMA processes are traditionally defined and analysed in the ACVF domain. Figure 5.3 shows the autocovariance functions of the three processes.

As it can be seen in the figure the ACVF of the fARIMA differs significantly from that of the fractional noise and it looks more similar to that of X'. So one can expect that it will behave similarly to X' under aggregation rather than to X. If, however, the variance time functions of the processes are compared (Figure 5.4) one can clearly see that while the difference of ω and ω' increases monotonically, the difference of ω and ω_f stays constant. This

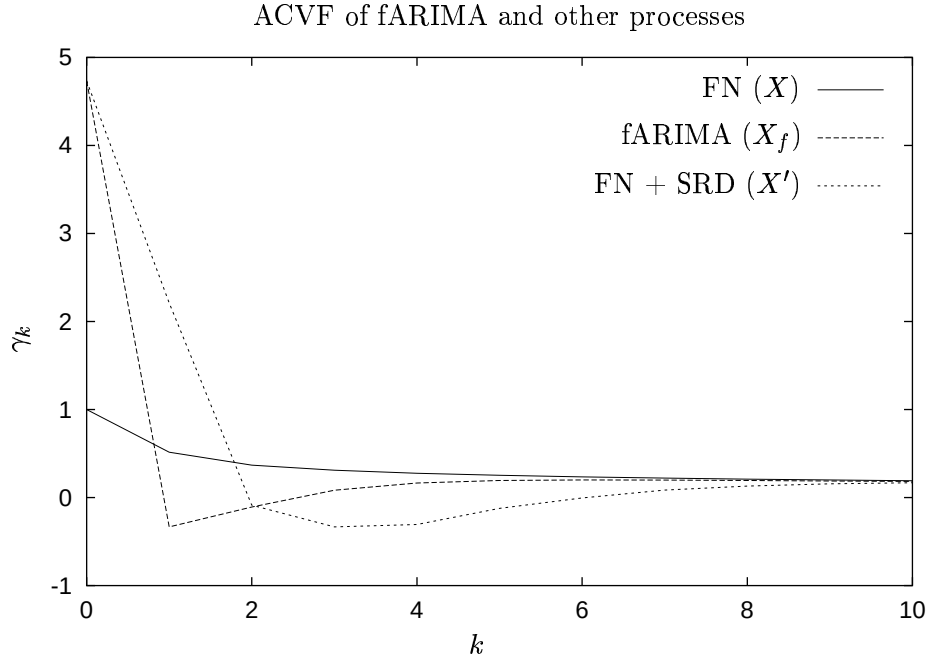

Figure 5.3: Autocovariance functions of fARIMA, fractional noise and modified fractional noise processes with the same C and H asymptotical parameters.

results in a very fast convergence speed of the aggregated autocovariance time function of X_f to that of X as can be seen in Figure 5.5.

Although in this example only a single fARIMA process was analysed our analytical results suggest that this fast speed of convergence is common to all fARIMA processes. Thus these results tell us, that as far as the scaling behaviour is concerned fARIMA processes are not generic enough and should not be used to test the robustness of LRD estimators and algorithms.

This result is more than obvious if one looks at the process from the variance time function perspective (Figure 5.4), but was hidden from the researchers for long time, since the shape of the autocovariance function does not suggest a fast speed of convergence.

5.3 Conclusions

The results presented in this work provide a new approach and a new framework for analysing the scaling behaviour of stochastic processes. In this chapter the advantages of the variance time function based approach to analyse scaling behaviour has been illustrated with examples. Within this framework the nature of scaling becomes simpler to describe, analyse and understand, and therefore foster the achievement of new results within this field. The scaling properties of the well-known fARIMA has also been analysed. From the autocovariance perspective (see Figure 5.3) the similarity between fARIMA and fractional noise is not so

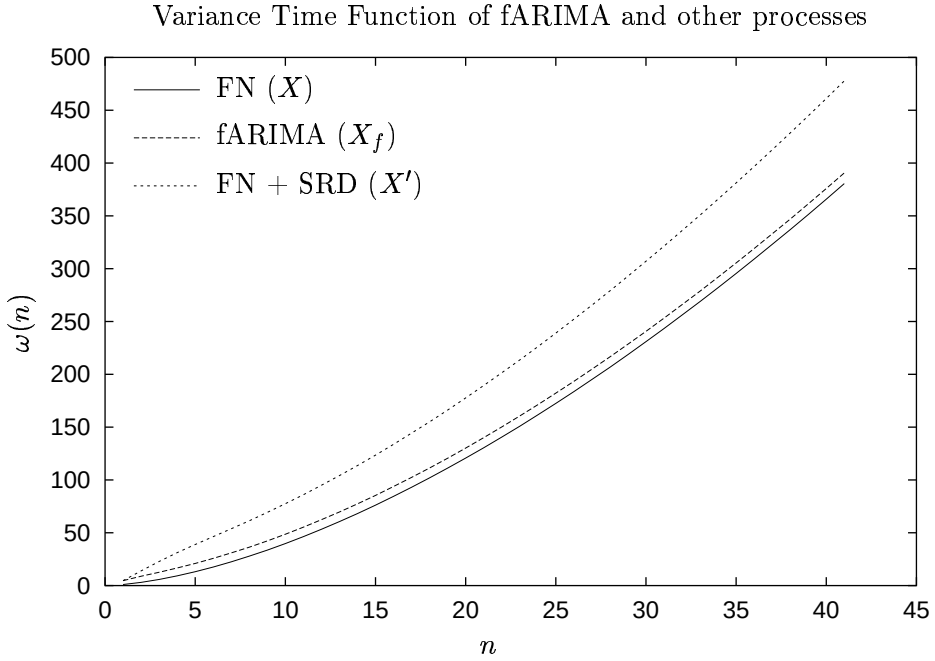

Figure 5.4: Variance time functions of fARIMA, fractional noise and modified fractional noise processes with the same C and H asymptotical parameters. The difference between ω and ω' increases monotonically while the difference between ω and ω_f stays constant.

apparent. The variance time function (see Figure 5.4) clearly shows that these two types of functions are asymptotically very similar and so the family of fARIMA processes is not suitable to represent generic LRD processes. This important result shows how the variance time function based approach can be used to reveal unknown details of known processes.

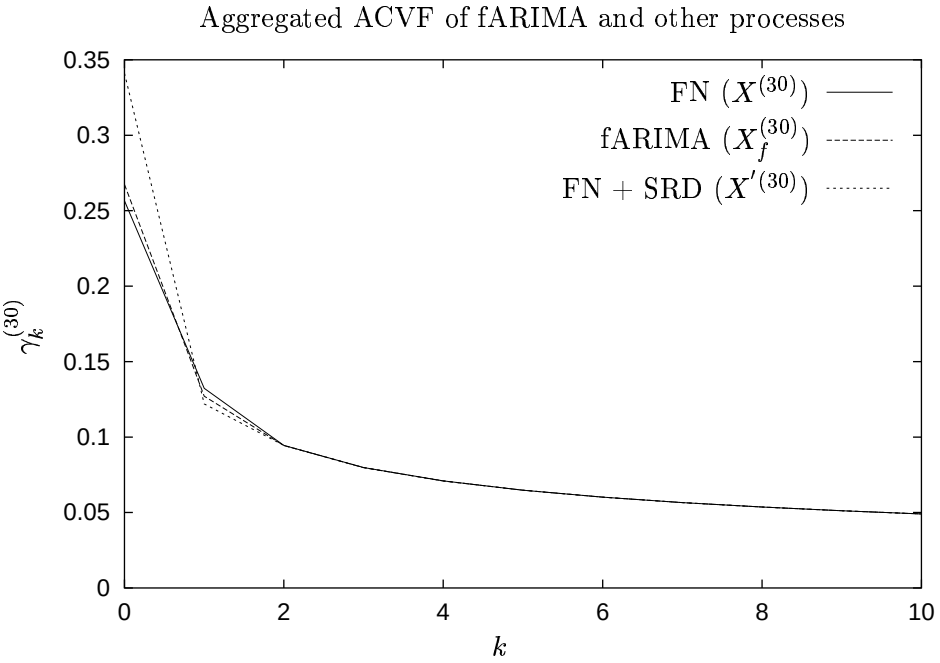

Figure 5.5: Autocovariance functions of the 30-aggregated processes. The aggregated ACVF of fARIMA ($\gamma_f^{(30)}$) is closer to the aggregated ACVF of FN ($\gamma^{(30)}$) than the aggregated ACVF of the modified fractional noise ($\gamma'^{(30)}$). The similarity of the aggregated autocovariances correspond to the similarity of the variance time functions (Figure 5.4).

Publications

[P1] Gefferth András. WWW információs rendszer az Austria Telecomnál (WWW Information system at Austria Telecom). *Magyar Távközlés (Hungarian Telecommunications)*, February 1996.

[P2] A. Gefferth, S. Molnár, and D. Veitch. Discrete Self-Similarity. In *IFIP WG6.7 Workshop and EUNICE Summer School on Adaptable Networks and Teleservices*, Trondheim, Norway, September 2-4 2002.

[P3] A. Gefferth, D. Veitch, I. Maricza, S. Molnár, and I. Ruzsa. The Nature of Discrete Second-Order Self-Similarity. *Advances in Applied Probability*, 35(2), June 2003.

[P4] A. Gefferth, D. Veitch, I. Ruzsa, I. Maricza, and S. Molnár. A New Class of Second Order Self-Similar Processes. *Stochastic Models*, 20(3):381–389 September 2004.

[P5] S. Molnár and A. Gefferth. On the Scaling and Burst Structure of Data Traffic. In *8th International Conference on Telecommunication Systems, Modelling and Analysis*, Nashville, Tennessee, USA, March 2000.

Bibliography

[1] P. Abry and D. Veitch. Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, January 1998.

[2] J. Beran. *Statistics for Long-Memory Processes.* Chapman and Hall, New York, 1994.

[3] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Cambridge University Press, Cambridge England, 1987.

[4] E. Brockmeyer, F. L. Halstrom, and A. Jensen. The Life and Works of A. K. Erlang. *Acta Polytechnica Scandinavia*, 1960.

[5] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer, 1996.

[6] D. R. Cox. *in Long-range dependence: a review,* H.A. David and H.T. David editors, chapter Statistics : an Appraisal, pages 55–74. Iowa State University Press, Ames (IA), 1984.

[7] I. Csabai. 1/f Noise in Computer Network Traffic. *Journal of Physics*, 1994.

[8] D. L. Jagerman and B. Melamed and W. Willinger. Stochastic Modeling of Traffic Processes. In *Frontiers in Queueing: Models, Methods and Problems.* CRC Press, 1996.

[9] W. Feller. *An Introduction to Probability Theory and Its Applications.* John Wiley and Sons, 1971.

[10] J. Galambos and E. Seneta. Regularly Varying Sequences. *Proceedings of the American Mathematical Society*, 41(1):110–116, November 1973.

[11] Peter Hall. Defining and Measuring Long-Range Dependence. *Fields Institute Communications*, 11:153–160, 1997.

[12] C. C. Heyde and Y. Yang. On Defining Long-Range Dependence. *Journal of Applied Probability*, 34:939 – 944, 1997.

[13] Reiner Kriesten, Ulrich Kaage, and Friedrich Jondral. A Unifying View To Fractional Modeling. In *Global Telecommunications Conference*, 1999.

[14] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the Self-Similar Nature of Ethernet Traffic (extended version). *IEEE/ACM Transaction on Networking*, 2(1):1–15, February 1994.

[15] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic. *Computer Communications Review*, 23:183–193, 1993.

[16] P. Major. *Multiple Wiener-Itô Integrals*, volume 849 of *Springer Lecture Notes in Mathematics*. Springer-Verlag, New York, 1981.

[17] S. Molnár and A. Gefferth. On the Scaling and Burst Structure of Data Traffic. In *8th International Conference on Telecommunication Systems, Modelling and Analysis*, Nashville, Tennessee, USA, March 2000.

[18] Melvyn B. Nathanson. *Additive Number Theory. The Classical Bases.* Graduate Texts in Mathematics 164. Springer, 1996.

[19] Sidney I. Resnick. *Extreme Values, Regular Variation and Point Processes.* Springer, New York, 1987.

[20] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman and Hall, 1994.

[21] Ya. G. Sinai. Self-Similar Probability Distributions. *Theory of Probability and its Applications*, 21:64–80, 1976.

[22] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(4):785–798, 1995. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

[23] Boris Tsybakov and Nicolas D. Georganas. On Self-Similar Traffic in ATM Queues: Definitions, Overflow Probability Bound, And Cell Delay Variation. *IEEE/ACM Transactions on Networking*, 5(3):397–409, June 1997.

[24] Darryl Veitch and Patrice Abry. A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory special issue on "Multiscale Statistical Signal Analysis and its Applications"*, 45(3):878–897, April 1999.

Appendix A

Proofs

A.1 Asymptotic equivalence and regular variation

Recall that two functions f, g are asymptotically equivalent, or $f \sim g$, if $f(x)/g(x) \to 1$ as $x \to \infty$.

A.1.1 Properties of asymptotic equivalence

In the following we will assume that $f_i(x) \sim g_i(x)$ for all i.

Æ–(i) Æ is symmetric ($f \sim g \Leftrightarrow g \sim f$), transitive ($f \sim g, g \sim h \Rightarrow f \sim h$), and reflexive ($f \sim f$).

Æ–(ii) $f_1(x)f_2(x) \sim g_1(x)g_2(x)$.

Æ–(iii) If $\alpha_i > 0$, $i < \infty$, constant then $\sum_i \alpha_i f_i(x) \sim \sum_i \alpha_i g_i(x)$ if all the f_i, g_i functions have the same sign near ∞.

Æ–(iv) $f_1(Cx) \overset{x}{\sim} g_1(Cx)$, where C is a positive constant.

Only **Æ**–(iii) will be proved as the others are obvious.

Proof of Æ–(iii)
We first show that if f_1 and f_2 have the same sign near ∞ then $(f_1+f_2) \sim (g_1+g_2)$. As $f_1 \sim g_1$ and $f_2 \sim g_2$, for all $\epsilon > 0$ there exists an x_0 such that for all $x > x_0$, $f_1(x) < (1+\epsilon)g_1(x)$, and $f_2(x) < (1+\epsilon)g_2(x)$, and so $(f_1+f_2)(x) < (1+\epsilon)(g_1+g_2)(x)$. Similarly, there exists x_1 such that for all $x > x_1$ $(f_1+f_2)(x) > (1-\epsilon)(g_1+g_2)(x)$, which means that $(f_1+f_2) \sim (g_1+g_2)$. Now using **Æ**–(ii) with $f_2(x) \equiv g_2(x) \equiv C$ the result follows from induction. □

A.1.2 Properties of continuous regular variation

CRV–(ii) $\widetilde{f} \in \mathrm{CRV}_\alpha \Leftrightarrow \widetilde{f}(x) = \widetilde{s}(x)x^\alpha$, $\widetilde{s}(x) \in \mathrm{CSV}$ (alternative Definition)

Proof
($\Rightarrow$) We must show that $\widetilde{s}(x) := \widetilde{f}(x)/x^\alpha \in$ CSV. But $\frac{\widetilde{s}(tx)}{\widetilde{s}(t)} = \frac{\widetilde{f}(tx)}{\widetilde{f}(t)}\frac{1}{x^\alpha} \to \frac{x^\alpha}{x^\alpha} = 1$.
($\Leftarrow$) If $\widetilde{s} \in$ CSV then for $\widetilde{f}(x) := \widetilde{s}(x)\, x^\alpha$ $\frac{\widetilde{f}(tx)}{\widetilde{f}(t)} = \frac{\widetilde{s}(tx)}{\widetilde{s}(t)}x^\alpha \to x^\alpha$ as required. □

CRV–(iii) $\widetilde{f} \in \mathrm{CRV}_\alpha$ and $\widetilde{g} \sim \widetilde{f} \Rightarrow \widetilde{g} \in \mathrm{CRV}_\alpha$

The statement is a direct consequence of **Æ**–(iv) and **Æ**–(ii).

CRV–(iv) $\widetilde{f} \in \mathrm{CRV}_\alpha \Rightarrow \widetilde{f}(x) \sim \widetilde{f}(x + x_0)$, $\forall x_0$ constant.
Proof
In [9] page 282 and [19] page 35 it was shown that if $\widetilde{s} \in$ SV then it has an integral representation of the form:

$$\widetilde{s}(x) = a(x)e^{\left(\int_1^x \frac{\epsilon(y)}{y}dy\right)}, \tag{A.1}$$

where $\epsilon(x) \to 0$ and $a(x) \to c$, $0 < c < \infty$ as $x \to \infty$. Using this representation

$$\frac{\widetilde{f}(x + x_0)}{\widetilde{f}(x)} = \left(\frac{x + x_0}{x}\right)^\alpha \frac{a(x + x_0)}{a(x)} \exp\left(\int_x^{x+x_0} \frac{\epsilon(y)}{y}dy\right).$$

We know that $\left(\frac{x+x_0}{x}\right)^\alpha \to 1$, $\frac{a(x+x_0)}{a(x)} \to 1$ because $a(x) \to c$, $0 < c < \infty$. We also have that $\int_x^{x+x_0} \frac{\epsilon(y)}{y}dy \to 0$ because $\epsilon(y) \to 0$. □

A.1.3 Properties of discrete regular variation

Properties **DRV**–(i) and **DRV**–(iv) follow directly from their continuous equivalents, since subsequences of convergent functions converge to the same limit.

DRV–(ii) $f \in \mathrm{DRV}_\alpha \Leftrightarrow f(k) = s(k)k^\alpha$, $s(k) \in$ DSV (alternative Definition)
Proof
($\Leftarrow$) Since $s \in$ DSV, by definition there exists a function $\widetilde{s} \in$ CSV s.t. $\widetilde{s}(k) = s(k)$, $\forall k \in Z^+$. Define $\widetilde{f}(x) := \widetilde{s}(x)x^\alpha$. According to **CRV**–(ii) $\widetilde{f} \in \mathrm{CRV}_\alpha$ and clearly $\widetilde{f}(k) = s(k)k^\alpha$. Hence $f \in \mathrm{DRV}_\alpha$ by definition. The proof of the opposite direction is very similar. □

The following lemma will be used for the proof of **DRV**–(iii) and Theorem 3.1.3:

Lemma A.1.1 (Step function interpolation)
If $f \in DRV_\alpha$ then the continuous time function $\widetilde{f}_s(x) := f(\lfloor x \rfloor) \in CRV_\alpha$, where $\lfloor x \rfloor$ in this paper denotes the integer that is the closest to x but not larger in absolute value.

Proof
Because $f \in \mathrm{DRV}_\alpha$ there exists a continuous function $\widetilde{f} \in \mathrm{CRV}_\alpha$ s.t. $\widetilde{f}(n) = f(n)$, $\forall n \in \mathbf{Z}^+$. If one can show that $\widetilde{f}(x) \sim \widetilde{f}(\lfloor x \rfloor) = \widetilde{f}_s(x)$, then according to **CRV**–(iii) $\widetilde{f}_s \in \mathrm{CRV}_\alpha$. To show this we again use the integral representation (A.1) of slowly varying functions, to obtain

$$\frac{\widetilde{f}(x)}{\widetilde{f}(\lfloor x \rfloor)} = \left(\frac{x}{\lfloor x \rfloor}\right)^\alpha \frac{a(x)}{a(\lfloor x \rfloor)} \exp\left(\int_{\lfloor x \rfloor}^x \frac{\epsilon(y)}{y}dy\right).$$

The first two terms on the right hand side converge to 1, and because the size of the interval $(\lfloor x \rfloor, x)$ is bounded and $\epsilon(x) \to 0$, the third term also converges to 1. □

DRV–(iii) $f \in \mathrm{DRV}_\alpha$ and $g \sim f \Rightarrow g \in \mathrm{DRV}_\alpha$.

Proof
Let $\widetilde{f}_s(x) := f(\lfloor x \rfloor)$ and $\widetilde{g}_s(x) := g(\lfloor x \rfloor), \forall x \in \mathbf{R}^+$. According to Lemma A.1.1 $\widetilde{f}_s(x) \in \mathrm{CRV}_\alpha$. It is clear that $\widetilde{f}_s \sim \widetilde{g}_s$, and so from **CRV**–(iii) $\widetilde{g}_s \in \mathrm{CRV}_\alpha$, and because $g(n) = \widetilde{g}_s(n)$, $\forall n \in \mathbf{Z}^+$ we have $g \in \mathrm{DRV}_\alpha$. □

Theorem A.1.1 (Summing Regular Varying Sequences, (Theorem 3.1.3))
Let $K(n) \in \mathrm{DRV}_\alpha$, and let $L(t)$ and $U(t)$ be defined as

$$L(m) := \sum_{n=0}^{m-1} K(n), \qquad U(m) := \sum_{n=m}^{\infty} K(n).$$

(a) If $\alpha \geq -1$ then $\dfrac{mK(m)}{L(m)} \to (1+\alpha)$, and $L \in DRV_{\alpha+1}$.

(b) If $\alpha < -1$ then $\dfrac{mK(m)}{U(m)} \to -(1+\alpha)$, and $U \in DRV_{\alpha+1}$.

Proof
Define the continuous function $\widetilde{K}(x) := K(\lfloor x \rfloor)$. According to Lemma A.1.1 $\widetilde{K} \in DRV_\alpha$, and so according to Theorem 3.1.2

$$\frac{x\widetilde{K}(x)}{\int_0^x \widetilde{K}(y)dy} \to (1+\alpha) \qquad \text{or} \qquad \frac{x\widetilde{K}(x)}{\int_x^\infty \widetilde{K}(y)dy} \to -(1+\alpha)$$

depending on the value of α. Subsequences of convergent functions converge to the same limit so the above equation remains valid if $x \in [0, \infty)$ is replaced by $n = 0, 1, 2, \ldots$. But in this case we know that $\widetilde{K}(n) = K(n)$ and also that $\int_0^n \widetilde{K}(y)dy = \sum_{i=0}^{n-1} K(n)$ and $\int_n^\infty \widetilde{K}(y)dy = \sum_{i=n}^{\infty} K(n)$. So it follows that

$$\frac{nK(n)}{\sum_{i=0}^{n-1} K(n)} \to (1+\alpha) \qquad \text{or} \qquad \frac{nK(n)}{\sum_{i=n}^{\infty} K(n)} \to -(1+\alpha)$$

depending again on the value of α. Because L and U are the sampled equivalents of $\widetilde{L}$ and $\widetilde{U}$ of Theorem 3.1.2 respectively they are both regular varying. □

A.2 Fixed points, PSD, and domains of attraction.

Lemma A.2.1 (Limits of processes remain processes (Theorem 2.3.1))
If a sequence $\{\omega_k\}$ of PSD VTFs converges pointwise, then the limit ω is also a PSD VTF.

Proof
Let $\gamma_k :\equiv \mathbf{D}\omega_k$ and $\gamma :\equiv \mathbf{D}\omega$. Since $\gamma_k(i)$ for any finite k is a finite linear combination of ω values, the pointwise convergence of ω_k implies that of γ_k. For any finite k, and any vector $\mathbf{a}$ of finite length m, $R_{\mathbf{a},k} := \sum_{j,i} \mathbf{a}_j \gamma_k(j-i)\mathbf{a}_i \geq 0$. As $R_{\mathbf{a},k}$ is a finite linear combination of $\gamma_k(i)$

values the summation and limit operation commute, that is $\lim_{k\to\infty} R_{\mathbf{a},k} = \sum_{j,i} \mathbf{a}_j \gamma(j-i)\mathbf{a}_i$. But as $R_{\mathbf{a},k} \geq 0$ for all k its limit cannot be negative, proving the positive semi-definiteness of γ and hence of ω. □

CTF–(i) $\phi(m-n) \leq 2(\phi(m)+\phi(n))$ for all $m, n \in \mathbf{Z}^+$, $m-n \geq 1$.

Proof

The statement is equivalent to $\omega(m-n) \leq 2(\omega(m)+\omega(n))$. Introduce the sums $A := \sum_{i=1}^{m} X(i)$ and $B := \sum_{i=m-n+1}^{m} X(i)$. It follows that $\omega(m-n) = E[(A-B)^2] = E[A^2] + E[B^2] - 2E[AB]$, $\omega(m) = E[A^2]$, $\omega(n) = E[B^2]$. Using $E[(A+B)^2]$ and combining, the result quickly follows. □

CTF–(ii) For any $n \in \mathbf{Z}^+$, $|s(n)| < 2\sqrt{\phi(n)} + C$, where C is a constant independent of n.

Proof

For any $k > 0$ define

$$b = \phi(n+k) - \phi(n) - ks(n) = \phi(n) + \sum_{i=n}^{n+k-1} s(i) - \phi(n) - ks(n) = \sum_{i=n}^{n+k-1} (s(i) - s(n)).$$

Since $|\rho(j)| \leq 1$ for any j, $|s(i) - s(n)| \leq 2(i-n)$, and using the triangle inequality we obtain $|b| \leq |k(k-1)|$. Similarly $|b| \leq |k(k+1)|$ for $k < 0$, so $|b| \leq |k(k+1)|$ for all k. It follows that

$$0 \leq \phi(n+k) \leq \phi(n) + ks(n) + k^2 + |k|, \text{ for any } k \in Z.$$

Setting $k := -\lfloor s(n)/2 \rfloor$ it then follows that

$$0 \leq \phi(n) - \frac{s(n)^2}{2} + \frac{|s(n)|}{2} + \frac{s(n)^2}{4} + \frac{|s(n)|}{2} = \phi(n) - \frac{s(n)^2}{4} + |s(n)|. \tag{B.2}$$

The largest possible $d := |s(n)|$ that satisfies (B.2) for a given $\phi(n)$ is the larger solution of the quadratic equation

$$\phi(n) - \frac{d^2}{4} + d,$$

that is $|s(n)| \leq |s(n)|_{\max} = 8 + \sqrt{1+4\phi(n)} \leq 2\sqrt{\phi(n)} + 9$. Choosing any $C > 9$ completes the proof. □

Lemma A.2.2 (Most valid ϕ are eventually bounded above zero.)
Given a PSD γ with $\mathcal{V} > 0$, if $\liminf_{m\to\infty} S(m) \geq 0$, *then* $\liminf_{m\to\infty} \phi(m) \geq 1/4$.

Proof

Let Γ_m be a PSD covariance matrix, $\mathbf{1}$ be the identity vector, and $\mathbf{e} = \{0, 0, \cdots, 0, \epsilon\}^T$ where ϵ is positive. Letting $\mathbf{a} = \mathbf{1} - \mathbf{e}$ and recalling Figure 1.1, we have

$$\begin{aligned} \mathbf{a}^T\Gamma_m\mathbf{a} &= \mathbf{1}^T\Gamma_m\mathbf{1} - \mathbf{e}^T\Gamma_m\mathbf{1} - \mathbf{1}^T\Gamma_m\mathbf{e} + \mathbf{e}^T\Gamma_m\mathbf{e} \\ &= w(m) - \epsilon[\mathcal{V} + S(m-1)] + \epsilon^2\mathcal{V} \geq 0 \end{aligned} \tag{B.3}$$

by PSD. Since for m sufficiently large $S(m-1)$ is either positive or negative but arbitrarily small, $w(m)$ becomes lower bounded in the limit by $\mathcal{V}\epsilon(1-\epsilon)$. As this quadratic in ϵ is maximised at $\epsilon = 1/2$, we have $\liminf_{m\to\infty} w(m) \geq \mathcal{V}/4$ from which the result follows. □

Lemma A.2.3 (Lemma 2.4.1)
If $S_\infty \in (0,\infty)$ is defined, then $\mathcal{V}^{(m)} \sim \frac{S_\infty}{m}$.

Proof
In view of Equation (1.6) it suffices to show that

$$\frac{\sum_{i=0}^{m-1} S(i)}{m} \to S_\infty. \tag{B.4}$$

For any $\epsilon > 0$ there exists a m_0 s.t. $\forall m \geq m_0$, $S_\infty - \epsilon \leq S(m) \leq S_\infty + \epsilon$. Now rewriting Equation (B.4)

$$\frac{\sum_{i=0}^{m-1} S(i)}{m} = \frac{\sum_{i=0}^{m_0-1} S(i)}{m} + \frac{\sum_{i=m_0}^{m-1} S(i)}{m}$$

the first term goes to 0 and the second term stays between $\frac{m-m_0}{m}(S_\infty - \epsilon)$ and $\frac{m-m_0}{m}(S_\infty + \epsilon)$, showing that the limit stays between $S_\infty - \epsilon$ and $S_\infty + \epsilon$. As this holds for any $\epsilon > 0$ the result follows. □

Theorem A.2.1 (Constrained short-range dependence (Theorem 2.4.2))
If $S_\infty = 0$ is defined, and $\gamma(k) \sim c_\gamma k^{2H-2}$ with $H \in (0, 1/2)$, then the process is in the domain of attraction of FN_H.

Proof
It will be shown that $\omega \in RV_{2H}$, which according to **DRV**–(i) and (2.19) yields the result.

Define the function $S_1^-(n) := \sum_{k=n}^{\infty} \gamma(k)$. We know that $\gamma \in RV_\alpha$ with $\alpha < -1$, so according to Theorem 3.1.3b

$$S_1^-(n) \sim -\frac{c_\gamma(n)}{2H-1} n^{2H-1} \in RV_{2H-1}.$$

It will now be shown how to express $\omega(m)$ in terms of S_1^-. $0 = S_\infty = S(n) + 2\sum_{k=n+1}^{\infty} \gamma(k)$ and so $S(n) = -2\sum_{k=n+1}^{\infty} \gamma(k) = -2S_1^-(n+1)$. Hence

$$\omega(m) = \sum_{n=0}^{m-1} S(n) = -2 \sum_{n=0}^{m-1} S_1^-(n+1).$$

We know that according to **DRV**–(iv) $S_1^-(n+1) \sim S_1^-(n)$ and so according to **DRV**–(iii) $S_1^-(n+1) \in RV_{2H-1}$ and so Theorem 3.1.3 can be used to show that

$$\omega(m) \sim \frac{c_\gamma(m)}{H(2H-1)} m^{2H} \in RV_{2H}.$$

as required, since $\phi(m) = \omega(m)/\omega(1)$. □