Limit Theorems and Estimation

for

Structural and Aggregate Teletraffic Models

by

David A. Rolls

A thesis submitted to

the Department of Mathematics and Statistics

in conformity with the requirements

for the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

January 2003

Abstract

The thesis proposes models for aggregate data network traffic which incorporate the additional randomness arising from the randomness in the number of data sources. A conditionally-Gaussian scale mixture process is shown to be a limit for the cumulative work from a random superposition of alternating on-off processes. Sub-Fractional Brownian Motion is shown to be the limit in a particular case. Queueing and estimation results for processes which are conditionally Fractional Gaussian Noise are included. A model with a superposition of alternating on-off processes with independent lifetimes is also considered.

Acknowledgments

First, I would like to thank my past teachers, both in Kingston and in Waterloo. Through them I obtained the foundations on which to build.

I am grateful to Dr. Glen Takahara for his unending guidance, support and patience.

I am very grateful to Rika Hagihara, who gave me extra motivation to finish this thesis.

Finally, I am most grateful to my parents. Their encouragement, concern and support have been extremely valuable to me.

Without them this effort would not have been possible.

Statement of Originality

All results presented in this thesis are original unless otherwise stated. Results from the literature include references.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The last five years has seen heightened interest in probability models with long-range dependence, self-similarity and heavy power-tails. This interest is motivated by observed data, which appears to show these properties in various applications. For example, packet traffic of modern data networks appears self-similar over many time scales. Associated with this property is long-range dependence, which may adversely affect the behaviour of networks in ways that Markov theory cannot predict.

Accurate modelling requires accurate measurements of model parameters. In particular, measurements of the self-similarity (Hurst) parameter, and the tail characteristic of a power tail are of interest. A typical assumption for statistical inference is that the data are independent and identically distributed (i.i.d.). With long-range dependent data the false assumption of independence may have serious consequences. For example, the sample mean is an estimator for the true mean. Under independence (or even short-range dependence), the variance of this estimator decreases as

n^{-1}, where n is the sample size. Under long-range dependence the variance decreases more slowly, as $n^{-\beta}$, $\beta \in (0,1)$. This has important implications on the size of confidence intervals. For accurate statistical inference long-range dependence must be considered.

The remainder of Chapter 1 presents the mathematical notions of self-similarity, long-range dependence, regular variation and mixing distributions. The presence of these features in observed data is discussed, with particular emphasis on packet data networks. Some current models incorporating these features are mentioned. An outline for the thesis appears in Section 1.6.

1.1 Self-Similarity in Packet Networks

Traditional communication networks (i.e. telephone networks) showed characteristics consistent with the assumptions of a Poisson process or a Markov arrival model. Poisson arrivals for voice connection requests and exponential holding times (durations) for call lengths were standard assumptions and the theory of these networks is well understood.

Within the last decade, measurements of network traffic at the packet level have demonstrated the need to develop queueing theory beyond the assumptions of Markov-type arrivals and exponential holding times. (The AMS Notices [95] provides a very accessible introduction.) Studies of Ethernet local area network (LAN) data [55], [56], [93] and wide area network (WAN) data [72], [19] have revealed interesting features, not easily captured with Markov arrivals and exponential holding times. Data traffic is very bursty- fluctuating between moments of low and high volume. Moreover, this burstiness appears on many time-scales (self-similarity). Attempts to model distributions of key quantities such as packet inter-arrival times and number of packets per unit time suggest they may have infinite variance- consistent with heavy-tails. Analysis and simulation of queueing systems with self-similarity and infinite variance suggest these features can have a large impact on the performance of a queue.

Much of the motivating work in this area concerns high-resolution measurements made of Ethernet LAN traffic. Perhaps the most striking feature of the data is revealed by graphing it at various time scales (Figure 1.1). Ethernet LAN traffic appears bursty on many timescales. In contrast, aggregating traditional models smooths the data.

The idea that the measurements appear similar over many different time-scales is referred to as *self-similarity* and can be placed in a mathematical framework, in which the time index is either continuous or discrete. The continuous-time definition is useful when thinking of the workload of a queue. The discrete-time definition of exactly self-similar is convenient for use with time series data. The intuitive idea is the equivalence of time and space.

To consider continuous time, let $X = \{X_t : t \geq 0\}$ be a continuous-time stochastic process with mean zero and stationary increments.

Definition 1.1.1. [83, p. 311] X is *self-similar* if

$$X_{at} \stackrel{D}{=} a^H X_t$$

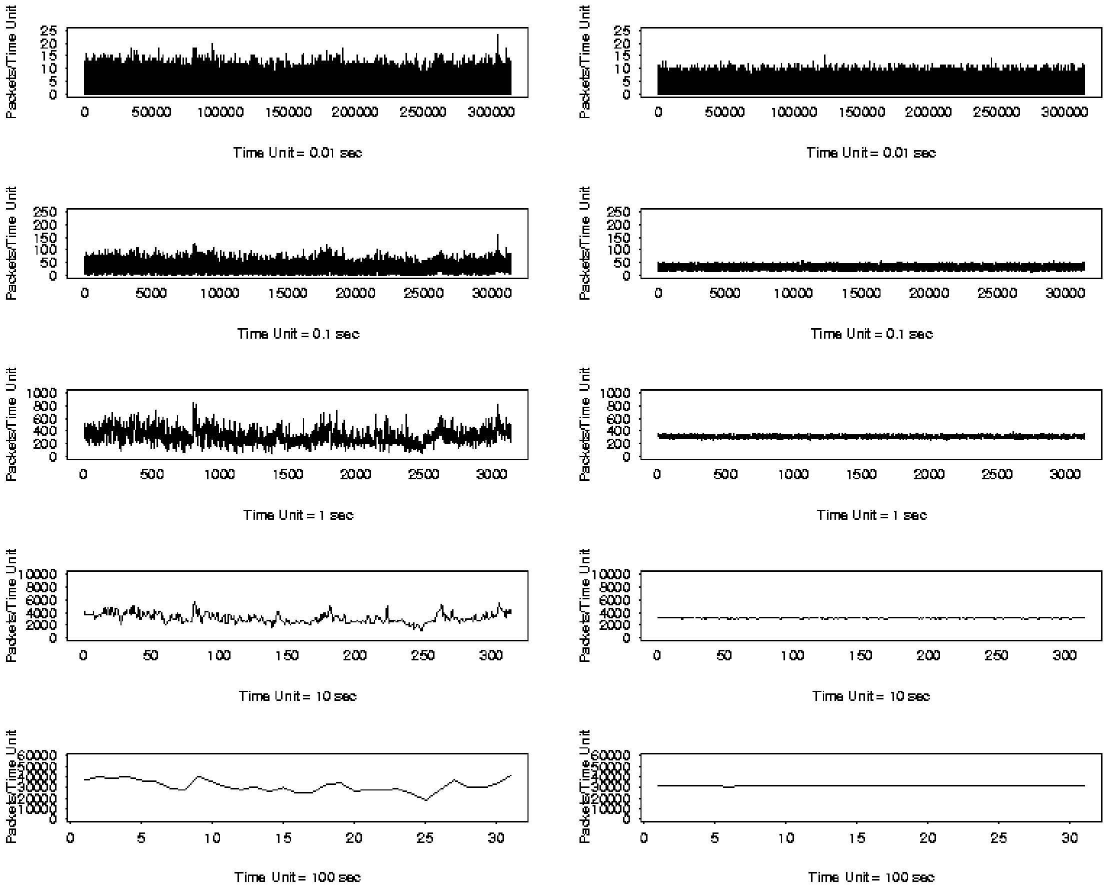

Figure 1.1: Looking for Self-Similarity

Aggregating reveals burstiness over time-scales. The left column is the packet counts from Bellcore's pAug.TL dataset. For contrast the right column simulates the same mean rate with i.i.d Poisson random variables. Longer datasets and a more complicated Markov model would show the same effect.

(Throughout, $\stackrel{D}{=}$ and $\stackrel{D}{\to}$ denote equality and convergence of distributions, respectively, while $\stackrel{P}{\to}$ denotes convergence in probability.) The exponent H is the self-similarity, or *Hurst*, parameter.

For discrete time, let $Y = \{Y_n, n = 0, 1, 2, \ldots\}$ be a covariance stationary stochastic process with mean μ, variance σ^2 and autocorrelation function $r(k)$, $k = 0, 1, 2, \ldots$.

Obtain the block means process by averaging consecutive blocks of size m

$$Y_k^{(m)} = \frac{1}{m}\left(Y_{(k-1)m+1} + \cdots + Y_{km}\right)\ ,\ k = 1, 2, 3, \ldots\ ,\ m = 1, 2, 3, \ldots.$$

Each $Y^{(m)}$ defines a covariance stationary process, with autocorrelation function $r^{(m)}(k)$. Now, for $0 < \beta < 1$ define

$$h(k) = \frac{1}{2}\left(|k+1|^{2-\beta} - 2|k|^{2-\beta} + |k-1|^{2-\beta}\right) = \delta^2(|k|^{2-\beta})/2$$

where

$$\delta(f(x)) = f(x + \frac{1}{2}) - f(x - \frac{1}{2})$$

is the central difference operator. Note that

$$h(k) \sim \frac{1}{2}(2-\beta)(1-\beta)k^{-\beta} \text{ as } k \to \infty.$$

(Throughout, $f(x) \sim g(x)$ means $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.)

Definition 1.1.2. [93, p. 72-73] For $0 < \beta < 1$ the covariance stationary process $Y = \{Y_n, n = 0, 1, 2, \ldots\}$ with mean μ, variance σ^2 and autocorrelation function $r(k)$, $k = 0, 1, 2, \ldots$ is

(a) *self-similar with parameter* $H = 1 - \frac{\beta}{2}$ if for all $m = 1, 2, 3, \ldots$ the process

$$\frac{1}{m^H}\left(Y_{(k-1)m+1} + \cdots + Y_{km}\right) = \frac{1}{m^{H-1}} Y_k^{(m)} \ , \ k = 1, 2, 3 \ldots$$

has the same finite-dimensional distributions as Y,

(b) *exactly second-order self-similar with parameter* $H = 1 - \frac{\beta}{2}$ if for all $m = 1, 2, 3, \ldots$

$$\frac{1}{m^H}\left(Y_{(k-1)m+1} + \cdots + Y_{km}\right) = \frac{1}{m^{H-1}} Y_k^{(m)} \ , \ k = 1, 2, 3 \ldots$$

has the same variance and autocorrelation as Y, that is

$$\mathrm{var}(Y^{(m)}) = \sigma^2 m^{-\beta} \text{ and}$$

$$r^{(m)}(k) = r(k) \ , \ k = 0, 1, 2, \ldots$$

and so, the autocorrelation must be $h(k)$,

(c) *asymptotically second-order self-similar with parameter* $H = 1 - \frac{\beta}{2}$ if

$$r^{(m)}(k) \to h(k) \text{ as } m \to \infty, k = 0, 1, 2, \ldots.$$

For a covariance stationary process, slow power-law decay of the autocorrelation function is significant [16].

Definition 1.1.3. [16, pp. 58-60] Let $\{Y_n\}$ be a covariance stationary sequence with some autocorrelation function $r(k)$ and power spectral density $g(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} r(n)e^{-in\omega}$, $-\pi < \omega < \pi$. The sequence is called *short-range dependent* if

(a) $\sum_{k=1}^{\infty} r(k) < \infty$

(b) $|g(0)| < \infty$, and

(c) $\{Y_k^{(m)}\}$ tends to second-order pure noise

(*i.e.* $\forall k \geq 1$, $r^{(m)}(k) \to 0$ as $m \to \infty$.)

This definition by Cox places few restrictions on the form of the autocorrelation $r(k)$. It includes power-law decay of the autocorrelation $r(k) \sim k^{-\beta}$, as $k \to \infty$ where $\beta > 1$ since this guarantees summability of $\sum r(k)$.

Definition 1.1.4. [16, pp. 59-60] Let $\{Y_n\}$ be a covariance stationary sequence with autocovariance

$$r(k) \sim k^{-\beta} L(k), \text{ as } k \to \infty \tag{1.1}$$

where $0 < \beta < 1$ and L is slowly varying at infinity (i.e. $\lim_{k\to\infty} \frac{L(kx)}{L(k)} = 1 \forall x > 0$. Slowly varying functions are further described in Section 1.4).

The process $\{Y_n\}$ is called *long-range dependent* if

(a) $\sum_{k=1}^{\infty} r(k) \to \infty$,

(b) $g(\omega)$ obeys a power law near 0

(*i.e* $g(\omega) \sim |\omega|^{-\gamma}$ as $\omega \to 0$, $0 < \gamma < 1$), and

(c) $m \operatorname{var}(Y^{(m)}) \to \infty$.

The article by Tsybakov and Georganas [92] provides a comprehensive treatment of the definitions of (asymptotically) second-order self-similarity and long-range dependence. They provide reasons why some definitions in the literature are unnecessarily narrow.

These definitions do not classify all covariance stationary processes. For example, it may be that $\sum r(k)$ is divergent although the asymptotic decay does not obey a power law. This might happen if $r(k) \to \rho$, for some constant $\rho \in [-1, 0) \cup (0, 1]$. A pathological example would have $r(k)$ oscillate between 1 and -1, so that $\sum r(k)$ is bounded but divergent.

An example of a self-similar process is Brownian motion, which has $H = 1/2$. In discrete time, an exactly (second-order) self-similar process is Fractional Gaussian Noise [61] (discussed below.) An example of an asymptotically second-order self-similar process is the Fractional AutoRegressive Integrated Moving Average (FARIMA) process introduced by Granger and Joyeux [32] and Hosking [38].

Several striking features of the short-range dependent processes are that they have summable autocorrelation functions and aggregation smooths the process to pure noise. Long-range dependent processes are distinguished by non-summable autocorrelation functions, a non-degenerate correlation structure, and the variance of a sample mean that decreases more slowly than the reciprocal of the sample size.

Exactly or asymptotically self-similar processes are long-range dependent. Their use in modelling of network traffic is in direct contrast with traditional Markov-type models with exponential (thus summable) functions $r(k) \sim \rho^k$, $0 < \rho < 1$, and spectral densities bounded at the origin.

The significance of long-range dependence in network traffic was addressed by Erramilli, Narayan and Willinger [27]. Their study used the inter-arrival times from Bellcore's Ethernet LAN data to form input to a queue. (Similar results for other data including Wide Area Network (WAN) traffic traces are also claimed.) To examine the effect of short- and long-term correlations in the data, they divided the data into equal size consecutive blocks, and randomly shuffled between blocks (external shuffle) or within blocks (internal shuffle). Plots of average delay versus utilization show that long-term correlations are much more important in determining the shape of the curve. The same result holds for the tail of buffer overflow versus utilization.

A measurement study of WAN traffic was described in Molnár and Vidács [64]. Their study measured data in the interior of the Finnish University and Research Network (FUNET). They found that the aggregate traffic exhibited properties of long-range dependence. Queueing results showed quantitative difference from simulated Poisson traffic. Traffic shaped with a “leaky bucket” traffic shaper continued to exhibit long-range dependent characteristics, particularly at large correlations.

1.2 Short-Range Dependent Models

Models of packet traffic with finite variance and short-range dependence were studied intensively in the early 1990’s. For particular applications to network engineering, the tail asymptotics of the probability of buffer overflow were specifically studied. This work was motivated in large part by the results of Anick, Mitra and Sondhi [3]. Subsequent research expanded the class of sources for which tail asymptotics were known. Two main techniques involved either solving an eigenvalue problem [35], [25] or using Cramér’s Theorem and extensions from large deviation theory [20], [44], [28], [46], [29]. Glynn and Whitt [29] considered tail asymptotics under very mild assumptions. The article by Kelly [45] provides a unified framework for these results.

About the same time that the results of Glynn and Whitt [29] attempted to treat most cases of interest, measurements of real traffic data demonstrated the need to revisit the assumptions of Markov-type arrivals and exponential holding times. Models incorporating self-similarity or long-range dependence were needed along with analysis of queues under these models. Assumptions such as those in Glynn and Whitt [29] were not broad enough to consider long-range dependence.

Models that incorporate self-similarity or long-range dependence can generally be classified as either short-range dependent approximations or true long-range dependent models. There are various examples in the first class. In Andersen and Friis Nielsen [2], a superposition of two-state Markov Modulated Poisson Processes (MMPPs) is proposed to model the process of packet counts of LAN traffic. By providing the number of underlying processes d, the number of time scales n, Hurst parameter H, and two other parameters, their algorithm fits the data using second-order properties. In effect, they approximate the autocorrelation of the data between lags 1 and 10^n with the sum of d exponential terms. They used this technique to fit two widely available datasets from Bellcore ("pAug.TL" and "pOct.TL") and generate simulated datasets based on the fitted parameters. Then they fed real data and corresponding synthetic data into a queue and compared the queue length probabilities. These were comparable for the pOct.TL data, but quite different for the pAug.TL data. Their suggestion is that modelling second order properties is insufficient for accurate modelling of queueing behaviour. Ren and Kobayashi [76] use an M-dimensional diffusion process to model the superposition of Markov modulated rate processes (MMRP). The total rate at any time is $\sum_{j=1}^{M} R_j N_j(t)$ where R_j is the rate associated with state j, and N_j is the number of sources in state j at time t. They model only simulated data, and conclude that when state holding times have finite variance their method predicts queue overflow well for utilizations that are fairly high ($\rho \geq 0.8$). To address long-range dependence they mention an extension in which the distribution of state holding times is approximated by a "truncated power-tail" [33].

1.3 Self-Similar Models

Models of network traffic that are truly self-similar or long-range dependent have also been proposed. Tsybakov and Georganas [92] proposed a discrete-time self-similar model. A Gaussian model based on Fractional Brownian Motion [48, 61, 55, 67, 68] has received much attention. A stable non-Gaussian model based on Linear Fractional Stable Motion has been proposed by Karasaridis and Hatzinakos [41]. Throughout much of this work is the desire to model real packet traffic with a minimum number of parameters. Thus, the term *parsimonious modelling* has become fashionable for those researching these models.

1.4 Power-Tails and Regular Variation

The distribution of a random variable X is said to have a power-tail with index α if

$$\mathrm{P}(X > x) \sim cx^{-\alpha} \text{ as } x \to \infty \text{ for some } c, \alpha > 0.$$

An example is the Student's t distribution with ν degrees of freedom, for which

$$\mathrm{P}(X > x) \sim cx^{-\nu} \text{ as } x \to \infty.$$

When $0 < \alpha < 2$ the variance of X is infinite, and the power-tail is also called a "heavy-tail". (A good introduction to these are the papers in [1].) On the other hand, if $\alpha > 2$ the variance is finite. The simplest example of a heavy-tailed distribution is the Pareto distribution with distribution function $F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}$, $x \geq k$, $k > 0$, $0 < \alpha < 2$ and density $p(x) = \alpha k^{\alpha} x^{-\alpha-1}$, $x \geq k$, $0 < \alpha < 2, k > 0$. Another class of heavy-tailed distributions are the stable (also called stable Paretian) distributions with tail index $\alpha < 2$.

The key feature of heavy-tails is the presence of infinite variance which manifests itself as extreme outliers in the data. Plots of data with infinite variance are characterized by almost all of the data compressed into a narrow band with only occasional but extreme outliers outside the band.

In teletraffic applications, heavy-tails have been reported in the distributions of On- and Off-times of Ethernet Local Area Network (LAN) source-destination pairs [94]. It is argued that this causes another phenomenon, long-range dependence. In [19] heavy-tails are reported in the distribution of Web file transmission times, transmitted file sizes, and some related quantities. In particular, infinite variance models attempt to capture large jumps in the measured quantities which appear as extreme outliers.

Computer technology has helped make infinite variance more relevant as larger datasets become more manageable, and applications produce more observations. In small datasets, outliers are not frequent, and arguably may be ignored as "exceptional". However, in datasets with hundreds of thousands or millions of observations it is much harder to ignore, say, twenty outliers. Moreover, their presence may have extreme consequences in the application being considered.

An important application of heavy-tails has been to insurance risk models. In this application the presence of heavy-tailed claims may lead to ruin if sufficient premiums are not collected. The key quantity is the probability of ruin, which can be expressed in terms of the supremum of a random walk. The standard method for considering the probability of ruin requires that the moment generating function of the increments exist in a neighbourhood of the origin (*i.e.* that exponential moments exist.) This technique is due to Cramér and forms the basis of classical large deviation theory. Heavy-tails require a different methodology. In this context the key class of distributions are the sub-exponential distributions $\mathcal{S}$ whose distributions F are defined by:

$$F \in \mathcal{S} \Leftrightarrow \lim_{x \to \infty} \frac{\mathrm{P}(X_1 + \cdots X_n > x)}{\mathrm{P}(\max(X_1, \ldots, X_n) > x)} = 1 \text{ for some (all) } n \geq 2$$

(It has been shown that if the limit above holds for some $n \geq 2$ then it holds for all $n \geq 2$ [31, p. 20].) This class includes the Pareto and stable distributions, for which regularly varying functions are also important.

Definition 1.4.1. A positive, measurable function L is said to be *regularly varying with index* α, $RV(\alpha)$ for $\alpha \in \mathbf{R}$ if

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = t^{\alpha} \; \forall t > 0. \tag{1.2}$$

If $L \in RV(0)$ the function is said to be *slowly varying*.

Examples of slowly varying functions are $L(x) = \log(x)$ and $L(x) = c$ for some constant $c > 0$. It can be shown [9, p. 26] that if $L_1(x) \in RV(\alpha_1)$ and $L_2(x) \in RV(\alpha_2)$ then $(L(x))^p \in RV(\alpha p)$ for $p \in \mathbf{R}$ and $L_1(x)+L_2(x) \in RV(\alpha)$ where $\alpha = \max(\alpha_1, \alpha_2)$.

A function is said to be *locally bounded* on a set A if it is bounded on every compact subset of A. It is *locally integrable* if it is integrable on every compact subset of A. In the tail, all regularly varying functions have these properties.

Proposition 1.4.1. *[9, p.18] If $f \in RV(\alpha)$ for some $\alpha \in \mathbf{R}$ then there exists $X \geq 0$ such that f and $1/f$ are locally bounded and locally integrable on $[X, \infty)$.*

Notice this does not say that the tail of the function need be integrable. Rather, far enough out in the tail the function is integrable on every compact (i.e. closed and bounded) set. The existence of the cutoff X is useful to take advantage of certain theorems.

A distribution function, F, has a regularly varying tail if

$$\bar{F}(x) = \mathrm{P}(X > x) = x^{-\alpha} L(x)$$

where $L(x)$ is slowly varying. These include the Pareto and Stable distribution functions, along with others such as Burr and log-gamma. The key difference between the power tail and the regularly varying tail is the inclusion of the slowly varying function in the latter. While that would seem to pose an inconvenience, at least for integration, the inconvenience is slight. The following propositions are included for their readability.

Proposition 1.4.2. *[9, p. 26] If $L(x)$ is slowly varying, X is so large that $L(x)$ is locally bounded in $[X, \infty)$ and $\alpha > -1$ then*

$$\int_X^x t^\alpha L(t)dt \sim x^{\alpha+1} L(x)/(\alpha+1) \text{ as } x \to \infty. \tag{1.3}$$

Proposition 1.4.3. *[9, p. 27] If $L(x)$ is slowly varying and $\alpha < -1$ then, for some constant $c > 0$, $\int_c^\infty t^\alpha L(t)\,dt$ converges and*

$$\frac{x^{\alpha+1} L(x)}{\int_x^\infty t^\alpha L(t)dt} \to -\alpha - 1 \text{ as } x \to \infty. \tag{1.4}$$

The case $\alpha = -1$ separates the two situations covered by these propositions. With the inclusion of the slowly varying function, $\int_c^\infty t^\alpha L(t)\,dt$ may or may not converge and this becomes a hypothesis for the corresponding proposition.

Proposition 1.4.4. *[9, p. 27] If $L(x)$ is slowly varying and $\int_c^\infty t^\alpha L(t)\,dt < \infty$ for some $c > 0$ then $\int_x^\infty L(t)/t\;dt$ is slowly varying, and*

$$\frac{1}{L(x)} \int_x^\infty L(t)/t\;dt \to \infty \;\text{ as } x \to \infty. \tag{1.5}$$

These propositions are captured and extended in the commonly-cited theorem due to Karamata. As in Bingham, Goldie and Teugels [9], Karamata's Theorem is presented in two halves.

Theorem 1.4.5 (Karamata's Theorem; direct half). *[9, p. 28] Let $f \in RV(p)$ and locally bounded in $[X, \infty)$. Then*

(a) for any $\sigma \geq -(p+1)$,

$$\lim_{x\to\infty} \frac{x^{\sigma+1} f(x)}{\int_X^x t^\sigma f(t)dt} = \sigma + p + 1\,;$$

(b) for any $\sigma < -(p+1)$ (and for $\sigma = -(p-1)$ if $\int^{\infty} t^{-(p+1)} f(t)dt < \infty$)

$$\lim_{x\to\infty} \frac{x^{\sigma+1} f(x)}{\int_x^{\infty} t^{\sigma} f(t)dt} = -(\sigma+p+1)\,.$$

Theorem 1.4.6 (Karamata's Theorem; converse half). *[9, p. 30] Let f be positive and locally integrable in $[X,\infty)$. Then*

(a) if for some $\sigma > -(p+1)$,

$$\lim_{x\to\infty} \frac{x^{\sigma+1} f(x)}{\int_X^{x} t^{\sigma} f(t)dt} = \sigma+p+1\,;$$

then $f \in RV(p)$.

(b) if for some $\sigma < -(p+1)$

$$\lim_{x\to\infty} \frac{x^{\sigma+1} f(x)}{\int_x^{\infty} t^{\sigma} f(t)dt} = -(\sigma+p+1),$$

then again f varies regularly with index p.

As an application to distribution functions, if F is a distribution with regularly varying tail in $RV(\alpha)$ then

$$F_I(x) = \int_0^x \bar{F}(t)\,dt \in RV(-(\alpha-1)).$$

Karamata's Theorem is useful for considering the asymptotic behaviour of integrated tails of distributions such as the Pareto distribution.

There is also a notion of a regular varying sequence [9, p. 53] c_n in which (1.2) is replaced by

$$\lim_{n\to\infty} \frac{c_{\lfloor \lambda n \rfloor}}{c_n} = \lambda^{\alpha}\,\forall \lambda > 0$$

where $\lfloor x \rfloor$ denotes the integer part of x. One can generally regard a regularly varying sequence as a regularly varying function. Since distribution functions are functions of a continuous variable, regular varying sequences will not be considered further.

1.5 Mixtures of Distributions

Let $G(\mathbf{x};\theta)$ be a d-dimensional cumulative distribution function (c.d.f.) depending on an m-dimensional parameter vector θ. Let $H(\theta)$ be an m-dimensional c.d.f. Then

$$F(\mathbf{x}) = \int G(\mathbf{x};\theta)\,dH(\theta) = \mathrm{E}_H[G(\mathbf{x};\theta)]$$

defines a c.d.f. for a new distribution called a *mixture* and H is called the *mixing* distribution [26, p. 4], [75, p. 184]. The distribution F is called a *finite mixture* if the distribution H assigns probability one to a finite number of points, a *countable mixture* if the number of points is countable, and an *arbitrary mixture* for a general mixing distribution.

Let Λ be a class of mixing distributions H and ζ be the class of corresponding mixture distributions. The classes Λ and ζ are called *identifiable* if there is a one-to-one mapping between Λ and ζ (i.e. for each element in one class we can identify one corresponding element in the other class). The practical importance of identifiability is in whether or not the distribution H can be uniquely determined from observations of the distribution F.

Let Z have a distribution function G, U and V be random variables, and define the random variable Y by

$$Y = UZ + V.$$

Then Y is a scale mixture through the effect of U and a location mixture through the effect of V. In short, Y is a scale-location mixture. Special cases involve the degenerate random variables $V = 0$ (scale mixture) or $U = 1$ (location mixture). The c.d.f. of Y can be written as

$$\begin{aligned} \mathrm{P}(Y \leq y) &= \mathrm{P}(UZ + V \leq y) \\ &= \mathrm{P}\left(Z \leq \frac{y - V}{U}\right) \\ &= \mathrm{E}_{(U,V)}\left[G\left(\frac{y - V}{U}\right)\right] \end{aligned}$$

where the last line can be seen through conditioning on (U, V), for example.

In the particular case that Z has a Normal distribution one gets scale-location mixtures of the Normal distribution. It is known that the class of scale-location mixtures of $N(\mu,\, \sigma^2)$ distributions are identifiable for finite mixtures, but not identifiable for arbitrary mixtures [75, p. 209], [89]. It is also known that arbitrary mixtures of $N(\mu,\, 1)$ (i.e. specified variance) distributions are identifiable. Also, the scale mixtures are identifiable, as are the location mixtures.

The scale mixtures of the Normal distribution are particularly interesting. Define the random variables $Z \sim N(0, 1)$, $Z_1 \sim N(0, 1)$, $S \sim \chi^2_{(\nu)}$, and X a stable random variable totally skewed to the right with characteristic exponent $\alpha/2$, $0 < \alpha \leq 2$, all independent of Z. Then the scale mixtures of the Normal distribution include

$$\frac{Z}{Z_1}$$

which has a Cauchy distribution,

$$t_\nu = \frac{Z}{\sqrt{\frac{S}{\nu}}}$$

which has a Student's t distribution with ν degrees of freedom, and

$$X^{1/2}Z$$

which has a symmetric stable distribution with characteristic exponent α. (In the particular case $\alpha = 1$, X would have a Lévy distribution and the product $X^{1/2}Z$ would have a Cauchy distribution.)

Scale-location mixtures of the Normal distribution arise as the limits of random sums. The Normal distribution is a consequence of the Central Limit Theorem, while randomness in the limit of summation leads to additional random variables U and V (possibly degenerate).

1.6 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 introduces several important data network traffic models from the literature in which long-range dependence and/or heavy-tails plays a role. Particular emphasis is placed on so-called 'structural' models and one involving Fractional Brownian Motion (FBM). The structural models connect properties in the aggregate traffic to properties in the constituent source-destination traffic streams. Known results for simulation and estimation of Fractional Gaussian Noise (FGN), a canonical self-similar Gaussian sequence formed by the increments of FBM, are also included.

In Chapter 3 a known simulation technique to approximate sample paths of FGN is improved. Using common statistical techniques, sample paths produced by the unimproved technique can be clearly distinguished by their inappropriate features. With a simple change these features are much less noticeable. With the improvement, the simulation technique can be useful for quickly simulating FGN data to test the performance of statistical estimators.

In Chapter 4 a technique to estimate the variance of FGN is presented. In the presence of long-range dependence, the sample variance is a biased estimator. A different estimator, based on Whittle's Maximum Likelihood Estimator, is developed. While that technique produces an estimate of 'scale', it is not an estimate of the variance, but one can be obtained from the other. Results from numerous simulations, some using the improved simulation technique from Chapter 3, are presented.

In Chapter 5 a known limit result for the aggregate of strictly alternating on-off sources is extended. In the known result, the number of on-off sources is considered to be large and fixed and the limiting process is FBM. In Chapter 5 the number of sources may be large, fixed but still random. Results for random sums are used to obtain the limiting process which involve some random variable U times FBM. In a special case, the limiting result is a stable non-Gaussian, self-similar process called sub-Fractional Brownian Motion. Simulation results are included.

Chapter 6 is also about traffic models. It takes as its starting point not the on-off processes from Chapter 5, but rather the limiting process of U times FBM. Over disjoint time intervals one might have an independent collection of these processes.

The 'Many-Process Scale Mixture Model' formalizes this idea. The ratio of variance to mean, γ, is shown to be a quantity of interest because it is not affected by the randomness from U. Techniques to estimate γ are proposed and these are applied to simulated on-off data assuming various distributions for the number of on-off sources. Variance estimators developed in Chapter 4 are used.

Chapter 7 concerns queueing results for models that incorporate additional randomness from the number of users, such as the Many-Process Scale Mixture Model. Known bounds and asymptotic results for the queue length of FBM-based Gaussian traffic are extended.

Chapter 8 is a more elaborate extension of the strictly alternating on-off model. On-off processes are given random independent 'lifetimes' whose distributions have tails that are regularly varying. The processes arrive according to a Poisson process. Thus the number of such on-off processes is allowed to vary randomly with time. Results for random sums are needed to obtain the limiting processes, which are found to be FBM. The degree of self-similarity is shown to depend not only on the distribution of on and off times, but also on the distribution of the lifetimes.

1.7 Thesis Contributions

This thesis makes contributions to the theory of long-range dependent traffic modelling and to simulation methodology and parameter estimation techniques for Fractional Gaussian Noise data.

A new structural model for data network traffic is proposed as a modification of a current model which involves a superposition of on-off processes. In the proposed model, the number of on-off sources is assumed to be random, although fixed in time. Convergence results to a scale mixture Fractional Brownian Motion are obtained by combining the proof employed for the current model with known results for the growing random sums of independent random variables with finite variance. The results justify the use of scale mixture Fractional Brownian Motion for modelling aggregate data network traffic.

For scale mixture Fractional Brownian Motion traffic, the constant γ, defined to be the ratio of variance to mean, is shown to be a quantity of interest. Two consistent estimators of γ are proposed, and their limiting distributions are found. These estimators are applied to the superposition of a large number of simulated on-off processes to demonstrate their effectiveness.

Queueing results are obtained for the queue length of a first-in first-out queue fed scale mixture Fractional Brownian Motion traffic. These results include a lower bound and asymptotics for the tail of the queue length distribution. It is shown that with the additional randomness, asymptotics of the unconditioned queue length distribution are quite different from the asymptotics when the additional randomness is not present. It is also shown that in certain cases the effect of the additional randomness occurs with long-range, but not short-range dependence.

Another new structural model for data network traffic is proposed. In this model, alternating on-off processes are assumed to arrive at Poisson time points and have independent, random lifetimes. Convergence of the finite-dimensional distributions to

Fractional Brownian Motion is shown, and the Hurst parameter is shown to depend also on the distribution of the lifetimes. The proof technique used here combines the proof for the superposition of on-off processes without lifetimes, known results for the convergence of growing random sums of independent random variables having finite variance, and Karamata's Theorem. The convergence results here provide further support for the use of Fractional Brownian Motion for network traffic modelling by showing this limiting process can be obtained even without the assumption that on-off processes exist for all time. These results also show how factors beyond the distribution of on and Off times may effect the degree of long-range dependence in the superposition.

An improved technique for simulating Fractional Gaussian Noise which yields increased accuracy, with no effect on execution time, is proposed as a modification of an existing method. The improvement is justified theoretically and the increased accuracy is demonstrated with numerous simulations. While the original method was proposed to simulate data network traffic, the additional accuracy obtained here allows new applications, for example in statistics, where long, fast, accurate simulations are needed.

A joint estimator $(\hat{\sigma}_n^2, \hat{H}_n)$ of the variance, σ^2, and the Hurst parameter, H, of Fractional Gaussian Noise is proposed as a modification of Whittle's approximate Maximum Likelihood Estimator. Consistency and limiting distributional properties of $(\hat{\sigma}_n^2, \hat{H}_n)$ are shown to be inherited from Whittle's estimator. The effectiveness of $\hat{\sigma}^2$ as a variance estimator is demonstrated with numerous simulations.

Chapter 2

Literature Review

2.1 Traffic Models

2.1.1 Fractional Brownian Traffic Model

A model for packet traffic often cited is based on Fractional Brownian Motion. This process, discovered by Kolmogorov [48], was described in 1940. It was popularized by Mandelbrot and Van Ness in 1968 [61], motivated by applications to hydrology. Its use for modelling network traffic was proposed in 1993 by Leland, Taqqu, Willinger and Sherman [55], and developed by Norros [67, 68].

There are several ways to define Fractional Brownian Motion. Mandelbrot and Van Ness [61] define its sample paths in terms of a stochastic integral and go on to reveal properties of the process. Samorodnitsky and Taqqu have called this a *Moving Average representation* [83, p. 320]. Alternatively, they define the process via its properties: the unique Gaussian self-similar process with Hurst parameter H,

$0 < H \leq 1$, and stationary increments. An equivalent definition adapted from Norros [67] is more revealing.

Definition 2.1.1. [67, pp. 387-388] Let $Z_H(t)$, $t \in (-\infty, \infty)$ be a process with the following properties:

(a) $Z_H(t)$ has stationary increments,

(b) $Z_H(0) = 0$, $\mathrm{E}Z_H(t) = 0$ for all t,

(c) $\mathrm{E}Z_H^2(t) = \sigma_0^2|t|^{2H}$ for all t,

(d) $Z_H(t)$ has continuous sample paths almost surely,

(e) $Z_H(t)$ is Gaussian (*i.e.* all its finite-dimensional distributions are Gaussian.)

Then $Z_H(t)$ is called *Fractional Brownian Motion (FBM)* and $H \in (0, 1]$ is the *Hurst parameter*.

In the case $H = 0.5$, Fractional Brownian Motion reduces to Brownian motion. The case $\sigma_0^2 = 1$ reduces to *standard* Fractional Brownian Motion. Norros requires all sample paths continuous, and $0.5 \leq H < 1$. These restrictions are not mathematically necessary. The definition given here is consistent with previous definitions [61], [83]. The case $H = 1$ is degenerate [83, p. 310] in the sense that $Z_1(t) = tZ_1(1)$ almost surely. Hereafter it is assumed $0 < H < 1$ unless otherwise stated.

By the properties given in the definition it's clear that $Z_H(t)$ is completely determined by its first and second-order properties, as with standard Brownian motion.

It's then easy to see that $Z_H(t)$ is self-similar since $\mathrm{E}(\alpha Z_h(t)) = 0$, $\alpha > 0$ and

$$\mathrm{E}[Z_H(\alpha t)^2] = |\alpha t|^{2H} = \alpha^{2H}|t|^{2H} = \mathrm{E}[\alpha^H Z_H(t)]^2 \ , \ \alpha > 0 \ , \ 0 < H < 1.$$

Definition 2.1.2. [83, p. 332] The increment process of Fractional Brownian Motion $Z_H(t)$

$$X_H(n) = Z_H(n+1) - Z_H(n) \ , \ n \in \mathcal{Z} \ (\text{or } \mathcal{Z}^+),$$

is a stationary Gaussian sequence called *Fractional Gaussian Noise (FGN)* and again H is the *Hurst parameter*.

It can be shown that

$$r(n) = \mathrm{Cov}(X(0), X(n)) = \frac{1}{2}\left((n+1)^{2H} - 2n^{2H} + (n-1)^{2H}\right) \ , \ n = 1, 2, 3, \ldots$$

so that X is exactly (second-order) self-similar with parameter H and is long-range dependent.

Norros [67, 68] (and also Leland, Taqqu, Willinger and Wilson [56]) proposed a fluid model for network traffic in which the total work arriving in time interval $[0, t)$ is described by the three parameter process

$$I(t) = mt + \sqrt{am} Z_H(t) \ , \ t \in (-\infty, \infty)$$

which he called *Fractional Brownian traffic*. In this model, $m > 0$ is the mean input rate, $a > 0$ is a variance coefficient and $H \in [0.5, 1)$ is the Hurst parameter. Then the variance can be written using the relation

$$\sigma^2 = ma.$$

In Norros [69] it is recognized that the constant a is the index of dispersion and the constant m captures the number of sources, without changing their characteristics.

When Fractional Brownian traffic is fed to a queue with server rate C, the net input traffic (sometimes called the *netput*) is

$$Y(t) = I(t) - Ct = (m - C)t + \sigma Z_H(t) \ , \ t \in (-\infty, \infty)$$

and the requirement $C > m$ is a stability condition. The difference

$$Y(t) - Y(s) = I(t) - I(s) - C(t - s) \ , \ s < t$$

is stationary by virtue of the stationary increments of Fractional Brownian Motion. Norros [67, 68] defined the storage occupancy process, called *Fractional Brownian Storage*, using Reich's formula [5, p. 4], as

$$V(t) = \sup_{s \leq t} \{Y(t) - Y(s)\} \tag{2.1}$$

$$= \sup_{s \leq t} \{I(t) - I(s) - C(t - s)\} \ , \ t \in (-\infty, \infty). \tag{2.2}$$

This is a familiar expression (see, for example, Glynn and Whitt [29]) for modelling the workload in the system at time t.

Since $V(t)$ is stationary and $Z_H(t)$ is time-invertible, the steady-state queue length for Fractional Brownian storage with parameters m, σ^2 and C can be defined as

$$Q^{(m,\sigma^2,C)} = V(0) = \sup_{t \geq 0}\{I(t) - Ct\} = \sup_{t \geq 0}\{\sigma Z_H(t) + (m - C)t\}. \tag{2.3}$$

It is easy to see that $Q^{(m,\sigma^2,C)} \geq 0$ almost surely (a.s.), but it is not clear that Q is a proper random variable (*i.e.* $\mathrm{P}(Q^{(m,\sigma^2,C)} < \infty) = 1$.) Indeed it will not be if $m \geq C$. However, if $m < C$, it is shown [69, p. 104] that $Q^{(m,\sigma^2,C)}$ is a.s. finite.

The exact distribution of $Q^{(m,\sigma^2,C)}$ is not known for general H, although $H = 0.5$ is an exception. Some results are available for $0.5 < H < 1$. Often these are also true for $H = 0.5$. For example, exploiting the self-similarity of Fractional Brownian Motion, one can show the following scaling law [69, p. 105].

Lemma 2.1.1.

$$P\left(Q^{(m,\sigma^2,C)} > x\right) = \bar{F}_Q\left(\frac{(C-m)^{H/(1-H)}}{\sigma^{1/(1-H)}}x\right) \quad \textit{where } \bar{F}_Q(x) = P\left(Q^{(0,1,1)} > x\right) \tag{2.4}$$

Thus it has been said that, since $\mathrm{P}\left(Q^{(0,1,1)} > x\right)$ depends only on H, the Hurst parameter is the only true parameter of the storage system– the others can be removed through scaling.

In the special case $H = 0.5$, the underlying process is Brownian Motion, which is commonly used for diffusion approximations. It is a special case because, for instance, the increments are independent. The net input process is

$$Y(t) = (m - C)t + \sigma Z_{0.5}(t)\ ,\ t \in (-\infty, \infty)$$

which is Brownian Motion with negative drift (under the stability assumption $m < C$) and

$$\mathrm{P}\left(Q^{(m,\sigma^2,C)} > x\right) = \mathrm{e}^{-\frac{2(C-m)}{\sigma^2}x},\ x > 0 \tag{2.5}$$

so that $Q^{(m,\sigma^2,C)}$ has an exponential distribution.

For the more general case $0.5 < H < 1$, Norros [67] proposed a bound on the decay of the complementary distribution function of $Q^{(m,\sigma^2,C)}$.

Theorem 2.1.2. *[67, pp. 394-395]*

$$P(Q^{(m,\sigma^2,C)} > x) \geq \overline{\Phi}\left(\frac{(C-m)^H}{\sigma}\frac{1}{\kappa(H)}x^{(1-H)}\right)$$

where $\overline{\Phi}(y) = \mathrm{P}(Z_H(1) > y)$ is the complementary distribution function of the standard Gaussian distribution and

$$\kappa(H) = H^H(1-H)^{(1-H)}.$$

Proof: See Norros [67]. □

The approximation $\overline{\Phi}(y) \sim \exp(-y^2/2)$ for large y gives

$$\overline{\Phi}\left(\frac{(C-m)^H}{\sigma}\frac{1}{\kappa(H)}x^{1-H}\right) \sim \exp\left(-\frac{(C-m)^{2H}}{2\sigma^2}\frac{1}{\kappa^2(H)}x^{2-2H}\right) \tag{2.6}$$

for large x, which is a lower bound on the tail asymptotics of $Q^{(m,\sigma^2,C)}$.

Duffield and O'Connor [23] combine classical large deviation techniques with the scaling properties of Fractional Brownian Motion to show

$$\lim_{x\to\infty} x^{-2(1-H)} \log \mathrm{P}(Q^{(m,\sigma^2,C)} > x) = -\frac{(C-m)^{2H}}{2\sigma^2\kappa^2(H)}. \tag{2.7}$$

Thus, the bound from (2.6) is log-asymptotically correct.

More desirable results concern not the log-asymptotics, but the true asymptotics, of $\mathrm{P}(Q^{(0,1,1)} > x)$. Massoulie and Simonian [62] use extremal properties of Gaussian processes to show

$$\mathrm{P}(Q^{(0,1,1)} > x) = \mathrm{O}\left(x^{-\eta}\mathrm{e}^{-\frac{x^{2-2H}}{2\kappa^2(H)}}\right), \text{ where } \eta = \frac{(1-H)(2H-1)}{H}, \tag{2.8}$$

$$\limsup_{x\to\infty} \frac{\mathrm{P}(Q^{(0,1,1)} > x)}{x^{-\eta}\mathrm{e}^{-\frac{x^{2-2H}}{2\kappa^2(H)}}} \leq L \tag{2.9}$$

for some constant L, and

$$\liminf_{x\to\infty} \frac{\mathrm{P}(Q^{(0,1,1)} > x)}{x^{-\eta} \mathrm{e}^{-\frac{x^{2-2H}}{2\kappa^2(H)}}} > 0. \tag{2.10}$$

To obtain the asymptotics of $\mathrm{P}(Q^{(0,1,1)} > x)$ Narayan [65] converts the problem to the frequency domain. An expression for the asymptotics, consistent with (2.8), is obtained, including a technique to find the leading constants. However, several unproven assumptions are used, and it is recognized that the argument does not provide a strict proof.

Hüsler and Piterbarg [39] also use extremal properties of Gaussian processes to obtain results.

Theorem 2.1.3. *Let $\{Z_H(t),\ t \geq 0\}$ be Fractional Brownian Motion with $0 < H < 1$. Then as $x \to \infty$*

$$P(Z_H(t) > x + Ct^{\beta},\ \textit{for some } t > 0) \sim K_1 x^{\frac{(1-H/\beta)(1-H)}{H}} \overline{\Phi}(Ax^{1-H/\beta}) \tag{2.11}$$

where K_1 is a known constant depending on H, $C > 0$ and $\beta > 0$ are constants, $\beta > H$ and

$$A = \left(\frac{H}{C(\beta - H)}\right)^{-H/\beta} \frac{\beta}{\beta - H}.$$

As a consequence, for Fractional Brownian Storage with $0.5 < H < 1$, $\beta = 1$, and so, as $x \to \infty$

$$\begin{aligned} \mathrm{P}(Q^{(0,1,1)} > x) &= \mathrm{P}(Z_H(t) > x + t \text{ for some } t > 0) \\ &\sim K_1 x^{\frac{(1-H)^2}{H}} \overline{\Phi}\left(\frac{x^{1-H}}{\kappa(H)}\right). \end{aligned}$$

Under the further approximation $\overline{\Phi}(x) \sim \frac{e^{\frac{-x^2}{2}}}{x}$ as $x \to \infty$

$$\mathrm{P}(Q^{(0,1,1)} > x) \sim K_1 x^{-\frac{(1-H)(2H-1)}{H}} \mathrm{e}^{-\frac{x^{2-2H}}{2\kappa^2(H)}} \tag{2.12}$$

which is the same asymptotic form found by Massoulie and Simonian.

The important conclusion from all of these results is that while $Q^{(0,1,1)}$ is exponential for $H = 0.5$, in the long-range dependent case it is asymptotically Weibull with shape parameter $2 - 2H$.

Finally, Norros [69] has obtained an exact expression for $\mathrm{P}(Q^{(0,1,1)} > x)$ when $0.5 < H < 1$. His approach defines the Gaussian martingale

$$M_t = c_1 \int_0^1 s^{1/2-H}(t-s)^{1/2-H} dZ_H(s)$$

where c_1 is a known constant depending on H. It can be shown that

$$\mathrm{P}(Q^{(0,1,1)} > x) = \mathrm{E}\left[\exp\left(-\left(\frac{x}{Z_1^*}\right)^{(1-H)/H} Z_1^* - \frac{c_2^2}{2}\left(\frac{x}{Z_1^*}\right)^{(2-2H)/H}\right)\right] \tag{2.13}$$

where

$$Z_1^* = \sup_{t\in[0,1]} Z_H(t) \text{ and } c_2 = \frac{H(2H-1)(2-2H)}{\beta(H-\frac{1}{2}, 2-2H)}$$

and $\beta(u,v) = \int_0^1 x^{u-1}(1-x)^{v-1}dx$ is the Beta function. While this technique provides an exact expression, the distribution of Z_1^* is currently unknown in a closed form, greatly limiting the applicability of the result.

Justification of the Fractional Brownian Traffic model

There are several factors contributing to the suggestion and widespread discussion of Fractional Brownian traffic as a reasonable model, at least for Ethernet LAN traffic.

The least technical is that it is a simple model to understand and work with. Beyond that, a variety of measurements on real data confirm assumptions of the model. There is also a much desired physical explanation of how Fractional Brownian Motion may arise.

There are three main aspects to the model. Traffic in fixed interval time "bins" is stationary, long range dependent and Gaussian. That the traffic is stationary and long-range dependent go together. Beyond the definition of long-range dependence that links them, certain kinds of non-stationarity can imitate long-range dependence. For the majority of Ethernet measurements, testing for long-range dependence against non-stationary alternatives doesn't reject long-range dependence [94]. Moreover, various inference techniques for estimating the Hurst parameter of packet traffic show $H > 0.5$ in almost all cases.

Assuming that network traffic is Gaussian is more problematic. Norros justifies the assumption by stating [68, p. 954] "...[w]hen a model is built on second-order properties alone, a Gaussian process is often the simplest choice." Fractional Brownian Motion is the only self-similar Gaussian process with stationary increments [83]. Showing that the traffic is Gaussian highly motivates the model. It is also a reasonable assumption in the sense of aggregating a large number of individual connections. With many connections, a central-limit type argument supports a Gaussian process.

A boost for the Fractional Brownian traffic model has come from a possible physical explanation, using a so-called "structural model" (see Section 2.1.2). The underlying mathematical result concerns the superposition of strictly alternating renewal processes, in which the renewal times have either heavy-tails or finite variance [60], [87], [56], [88], [95]. By aggregating the number of sources $M \to \infty$ and then letting the time $T \to \infty$, under the proper rescaling, the centered cumulative packet count process converges to Fractional Brownian Motion, where H depends on the characteristic exponent for both the On- and Off-time distributions.

Interestingly, the order of the limits $M \to \infty$ followed by $T \to \infty$ is essential. Reversing the order gives a very different result. In this case the centered process is not Fractional Brownian Motion, but Lévy Stable motion. This is an entirely different process which is self-similar with increments that are stationary, α-stable, and *independent*. For this process $H = 1/\alpha$. When $\alpha = 2$ it reduces to standard Brownian motion with continuous sample paths. For $1 < \alpha < 2$ the paths are discontinuous. A physical context for this result has not been found, and the result seems more of a curiosity at this point. In practice, both M and T are finite and the issue would seem to be whether $T \ll M$ (Fractional Brownian Motion) or $M \ll T$ (Lévy Stable motion).

Detailed measurements of Ethernet data at the source-destination level [94], [95] support the assumption that On- and Off-times are heavy-tailed. However, the data also suggests that individual sources are not entirely independent since they compete for finite resources. The authors hypothesize the result above holds when the sources are weakly dependent.

2.1.2 Alternating On-Off Structural Model

Structural models for packet traffic have been described previously [10, 11, 58, 60, 87, 88, 94]. One such model is the strictly alternating on-off model with heavy-tailed on or Off times [88, 94]. It provides some theoretical justification for the use of Fractional Brownian Motion to model the variability in aggregate packet traffic. In such a model the traffic characteristics of an individual source *at the source edge* of the network are modelled. Empirical evidence for heavy-tails in the sizes of WWW objects [18] further motivates the model.

As described in Taqqu, Willinger and Sherman [88], let $\{W(t),\, t \geq 0\}$ be a strictly alternating, stationary on-off process that takes the value 0 during an Off-period and 1 during an On-period. Let $\{W^{(m)}(t),\, t \geq 0\}$, $m = 1, 2, \ldots$ be mutually independent copies of $\{W(t)\}$. Let $r(t) = \mathrm{E}(W^2(t)) - [E(W(t))]^2$ denote the autocovariance of $W(t)$, and $\mu_W = \mathrm{E}[W(t)]$.

The on and Off times are non-negative, and may be heavy-tailed. The length of the on periods are independent and identically distributed (i.i.d), as are those of the Off-periods. In particular, for the on times let $f_1(x)$, $F_1(x)$, μ_1, and σ_1^2 be the density, distribution function, mean, and variance (if it exists). Let $f_2(x)$, $F_2(x)$, μ_2 and σ_2^2 denote the respective quantities for the Off times. The on and Off times are assumed independent. Then

$$\mu_W = \frac{\mu_1}{\mu_1 + \mu_2} \text{ and } \sigma_W^2 = \mathrm{Var}\left[W(t) - \mathrm{E}[W(t)]\right] = \frac{\mu_1 \mu_2}{(\mu_1 + \mu_2)^2}.$$

Assume as $x \to \infty$

$$\text{either } \bar{F}_1(x) = 1 - F_1(x) \sim l_1 x^{-\alpha_1} L_1(x) \text{ with } 1 < \alpha_1 < 2 \text{ or } \sigma_1^2 < \infty$$

and

$$\text{either } \bar{F}_2(x) = 1 - F_2(x) \sim l_2 x^{-\alpha_2} L_2(x) \text{ with } 1 < \alpha_2 < 2 \text{ or } \sigma_2^2 < \infty$$

where $L_j > 0$ is a slowly varying function and $l_j > 0$ is a constant, $j = 1, 2$.

When $1 < \alpha_j < 2$ let $a_j = l_j\left(\Gamma(2-\alpha_j)/(\alpha_j - 1)\right)$. When $\sigma_j^2 < \infty$ set $\alpha_j = 2$, $L_j = 1$ and $a_j = \sigma_j^2/2$. Let

$$b = \lim_{t\to\infty} t^{\alpha_2-\alpha_1}\frac{L_1(t)}{L_2(t)}.$$

If $0 < b < \infty$ set $\alpha_{min} = \alpha_1 = \alpha_2$, $L = L_2$ and

$$\sigma_{lim}^2 = \frac{2(\mu_2^2 a_1 b + \mu_1^2 a_2)}{(\mu_1+\mu_2)^3\Gamma(4-\alpha_{min})}. \tag{2.14}$$

If $b = 0$ or $b = \infty$ set $L = L_{min}$ and

$$\sigma_{lim}^2 = \frac{2\mu_{max}^2 a_{min}}{(\mu_1+\mu_2)^3\Gamma(4-\alpha_{min})} \tag{2.15}$$

where *min* is the index 1 if $b = \infty$ and the index 2 if $b = 0$, and *max* denotes the other index.

Under these conditions the following theorem has been proven by Taqqu, Willinger and Sherman.

Theorem 2.1.4 ([88]). *For $H = (3 - \alpha_{min}/2)$ and σ_{lim} as above*

$$\mathcal{L}\lim_{T\to\infty}\mathcal{L}\lim_{\mathcal{M}\to\infty}\frac{\left(\int_0^{Tt}\left(\sum_{m=1}^{M} W^{(m)}(u)\right)du - TM\mu_W t\right)}{T^H L^{1/2}(T) M^{1/2}} = \sigma_{lim} Z_H(t)$$

where $\mathcal{L}\lim$ means convergence of the finite-dimensional distributions, and $Z_H(t)$ is standard Fractional Brownian Motion.

Their proof relies on using the Central Limit Theorem to show that

$$\mathcal{L} \lim_{\mathcal{M}\to\infty} M^{-1/2} \sum_{m=1}^{M} \left(W^{(m)}(u) - \mu_W\right) = G(t) \tag{2.16}$$

where $\{G(t), t \geq 0\}$ is stationary Gaussian process with mean 0 and the same auto-covariance as $\{W(t),\, t \geq 0\}$, namely $r(t)$. Then

$$\begin{aligned} V(t) &= \mathrm{Var}\left[\int_0^t W(u)\,du\right] \\ &= 2\int_0^t \int_0^v r(u)\,du\,dv \\ &\sim \sigma_{lim}^2 t^{2H} L(t) \text{ as } t \to \infty \end{aligned}$$

where the last line follows through use of the extended form of Karamata's Tauberian Theorem [9, p. 43]. The conclusions of the theorem follow from the uniqueness of Fractional Brownian Motion as the only mean zero Gaussian process with stationary increments and variance $\sigma_{lim}^2 t^{2H}$ for fixed t.

As part of a similar argument Brichet, Simonian, Massoulié and Veitch [11] provide conditions under which $r(t)$ is regularly varying at infinity.

Theorem 2.1.5. *[11, pp. 126-127] For the distributions F_1 and F_s assume $1 < \alpha_i \leq 2$, $i = 1, 2$, $\alpha_1 + \alpha_2 \neq 4$, and if $\alpha_i \neq 2$ the slowly varying function $L_i(t)$ obeys*

$$\bar{L}_i \equiv \lim_{t\to\infty} \bar{L}_i(t) \in (0, \infty),\; i = 1,\, 2. \tag{2.17}$$

Then

$$r(t) \sim \frac{\sigma_{lim}^2}{2}(3 - \alpha_{min})(2 - \alpha_{min})\bar{L}_{min} t^{\alpha_{min}-1} \tag{2.18}$$

and, in particular, $r(t) \in RV(\alpha_{min} - 1)$.

Notice that Theorem 2.1.5 covers some, but not all, of the distribution functions with tails that are heavy and regularly varying which are covered by Taqqu [88]. The key difference is that (2.17) prevents slowly varying functions such as $L(x) = \log(x)$. But, for example, the tail of the Pareto distribution with characteristic exponent α, $1 < \alpha < 2$ has a slowly varying function that does satisfy (2.17).

2.1.3 $M/G/\infty$ Input Process

The busy server process of an $M/G/\infty$ queueing system has been proposed as a flexible alternative to model data traffic. Cox and Isham [17] considered the continuous-time situation as an immigration-death model, a queueing model with nearly immediate service and also a textile yarn model. Cox [16] identified the role of heavy-tailed distributions leading to long-range dependence. Parulekar [71] and Makowski and Parulekar [59] developed the discrete-time model for packet data traffic.

Consider a queueing system with infinitely many servers. During time slot $[t, t+1)$, N_{t+1} new customers enter the system. Customer i, $i = 1, \ldots, N_{t+1}$ begins service by slot $[t+1, t+2)$ at its own server. Its service time is the number of time slots until it completes service, and is represented by $S_{t+1,i}$. Let X_t denote the number of busy servers (i.e. the number of customers in service) at the beginning of time slot $[t, t+1)$. Assume X_0 servers are initially busy in slot $[0, 1)$, each having residual service time $S_{0,i}$, $i = 1, \ldots, X_0$. The process $\{X_t, t = 0, 1 \ldots\}$ is referred to as the $M/G/\infty$ input process.

In this model several assumptions are made. It is assumed that $\{N_t, t = 1, \ldots\}$

is a sequence of i.i.d. Poisson random variables with parameter $\lambda > 0$, and that $\{S_{t,i}, t = 1, 2, \ldots; i = 1, 2, \ldots\}$ are i.i.d. with common probability mass function G on $1, 2, \ldots$ that describes a distribution with finite mean. The random variables $\{N_{t+1}\}$ and $\{S_{t,i}, t = 1, 2, \ldots; i = 1, 2, \ldots\}$ are assumed mutually independent.

The busy server process is, in general, not stationary. However, it is stationary under additional assumptions for the initial state of the system. In particular, let S be a generic random variable with probability mass function G, and let $\hat{G}$ be the equilibrium probability mass function with generic random variable $\hat{S}$ given by

$$\mathrm{P}(\hat{S} = k) = \frac{\mathrm{P}(S \geq k)}{\mathrm{E}[S]}, \; k = 1, 2, \ldots.$$

Assume X_0 is Poisson with parameter $\lambda \mathrm{E}[S]$, each having residual service time $\hat{S}_{0,i}$, $i = 1, \ldots, X_0$ distributed according to $\hat{G}$. Under these additional assumptions the following proposition can be established.

Proposition 2.1.6. *[59, p. 220] The stationary and ergodic version* $\{X_t, t = 0, 1, \ldots\}$ *of the busy server process has the following properties:*

(a) For each $t = 0, 1 \ldots$*, the random variable* X_t *is a Poisson random variable with parameter* $\lambda E[S]$*.*

(b) The process is reversible in that

$$(X_0, X_1, \ldots, X_t) \stackrel{D}{=} (X_t, X_{t-1}, \ldots, X_0), t = 0, 1, \ldots.$$

The class of $M/G/\infty$ input processes has several nice features for traffic modelling. For example, the superposition of $M/G/\infty$ input processes is also an $M/G/\infty$

input processes. The dependence structure can be altered by changing the common distribution of the service times. They have also been found as the limit of the superposition of i.i.d. on-off sources with Pareto service times, in which the number of sources increases.

For the stationary $M/G/\infty$ input process, the autocovariance is known to be

$$\Gamma(k) = \lambda \mathrm{E}[(S-k)^+] = \lambda \mathrm{E}[S]\mathrm{P}(\hat{S} > k)$$

and the autocorrelation is

$$r(k) = \mathrm{P}(\hat{S} > k) = \frac{\mathrm{P}(S \geq k+1) + \mathrm{P}(S \geq k+2) + \ldots}{\mathrm{E}[S]}.$$

It is also known that

$$\sum_{k=0}^{\infty} \Gamma(k) = \frac{\lambda}{2}\mathrm{E}[S(S+1)]$$

and so $\{X_t\}$ is short-range dependent if and only if $\mathrm{E}[S^2]$ is finite. On the other hand, $\{X_t\}$ will be long-range dependent (i.e. asymptotically self-similar) if $P(S > s)$ is heavy-tailed (or even regularly varying at infinity) with characteristic exponent $\alpha \in (1,2)$. In this case, $H = (3-\alpha)/2$. If S has the distribution given by

$$\mathrm{P}(S > k) = \frac{|k+2|^{2H} - 3|k+1|^{2H} + 3|k|^{2H} - |k-1|^{2H}}{4(1-2^{2H-2})}, k = 1, 2, \ldots$$

for $H \in (0.5, 1)$ then $\{X_t\}$ is exactly self-similar.

Results have been obtained for the queue length of a single-server first-in first-out queue fed by an $M/G/\infty$ input process. In the case of service times with exponential tails, asymptotics are known. In more general cases, only bounds are known. Assume the queue has a constant server rate of C packets per slot. Let Q_t represent the number of packets remaining in the buffer at the end of slot $[t-1, t)$. The queue is stable if

$$\lambda \mathrm{E}[S] < C$$

and in this case $Q_t \to Q$ as $t \to \infty$ where Q is the steady-state queue length and

$$Q \stackrel{D}{=} \sup_{t>0} \{(X_1 + \cdots + X_t) - ct\}.$$

The random variable Q is known to be almost surely finite when the sequence $\{X_t, t = 1, 2, \ldots\}$ is stationary and ergodic, and $\mathrm{E}[X_1] < \infty$.

If the tail of the service-time distribution G decays at least exponentially fast (e.g. Gamma, geometric, Rayleigh), large deviation techniques can be used to study the asymptotics of $\mathrm{P}(Q > b)$. It can be shown that

$$\lim_{b \to \infty} \frac{1}{b} \log \mathrm{P}(Q > b) = -c$$

for some constant c. This asymptotic form is closely connected with the notion of "effective bandwidths".

2.2 Simulation Techniques

2.2.1 Simulation of Fractional Gaussian Noise

Paxson [73, 74] proposed a technique that produces approximate sample paths from FGN of even length n for Hurst parameter $H \in [0.5, 1)$. The technique begins by approximating the power spectrum of FGN $\{f_1, f_2, \ldots f_{n/2}\}$ where f_j is the power spectrum at frequency $\frac{2\pi j}{n}$, $j = 1, 2, \ldots, n/2$. Each f_j is then "fuzzed" to give $\hat{f}_j$ by multiplying by an independent exponential random value with mean 1. This corresponds to a known result that asymptotically the periodogram is independent at distinct frequencies and is exponentially distributed with mean equal to the spectral density. In other words, $\hat{f}_j$ corresponds to the first half of the frequencies of the "unscaled" periodogram

$$I(\omega_j) = \left| \sum_{k=1}^{n} e^{-i\omega_j k} Y_k \right|^2 \quad \text{where } \omega_j = \frac{2\pi j}{n}, \, j = 1, 2, \ldots, n \tag{2.19}$$

and Y_k is the desired sample path. Then a sequence of complex values $\{z_1, \ldots, z_{n/2}\}$ is generated in which $|z_i| = \sqrt{\hat{f}_j}$ and the phase of z_j are independent and uniformly distributed over the unit circle. The penultimate step is to expand $\{z_1, \ldots, z_{n/2}\}$ to $\{z'_0, \ldots, z'_{n-1}\}$ by forming

$$z'_j = \begin{cases} 0, & j = 0 \\ z_j, & 0 < j \leq n/2 \\ \overline{z}_{n-j}, & n/2 < j < n \end{cases} \tag{2.20}$$

where $\overline{z}_n$ denotes the complex conjugate of z_n. This sequence corresponds to $\tilde{Y}_j$, the Fourier transform of the desired sample path. Symmetry guarantees the sample path will be real-valued. The final step generates Y_k by applying the inverse discrete Fourier transform.

2.3 Mean and Variance Estimation of Fractional Gaussian Noise

2.3.1 Estimation of the Mean

Let $\mathbf{X} = \{X(j), j = 1, \ldots, n\}$ be a consecutive sample of data from FGN with variance σ^2. Let $\mathbf{X}_\mu = \{X(j) + \mu, j = 1, \ldots, n\}$ be FGN shifted to have mean μ. Estimators of the mean and variance for data $\mathbf{X}$ or $\mathbf{X}_\mu$ have been studied extensively (see Beran [7] and the references therein.) Even in the presence of long-range dependence, the sample mean $\bar{\mu}_n = \sum_{j=1}^{n} X(j)/n$ performs well [7, p. 148-151]. It is a consistent estimator, and the relative asymptotic efficiency is at least 98% compared with the best linear unbiased estimator for μ, which is also the maximum likelihood estimator (MLE). Moreover, as the average of Gaussian random variables, $\bar{\mu}_n$ is conveniently Gaussian for all n.

2.3.2 Estimation of the Variance

Unlike estimating the mean, long-range dependence plays an important role in estimating the variance. The sample variance,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is known to be biased and inefficient [7, p. 156-157], [53]. If H is known, it is easy to unbias, although generally H is unknown. More importantly, while S^2 is asymptotically Gaussian for $0.5 \leq H < 0.75$, its distribution is non-Gaussian and complicated for $H > 0.75$ [7, p. 67-73, 156-157]. At $H = 0.75$ there is a known change in the limiting behaviour of the sum of the squares of random variables from FGN. (This is really a consequence of an important result by Taqqu [85] concerning a function of Hermite rank 2 applied to a Gaussian, long-range dependent sequence.) Notice that the presence of a non-Gaussian limit does not contradict the Central Limit Theorem since random variables from FGN are not independent. Instead of using the sample variance, estimators based on parametric Maximum Likelihood Estimate (MLE) methods are $\sqrt{n}$-consistent for all $H \in (0.5, 1)$ and asymptotically Gaussian.

Whittle's Estimate

Denote Whittle's approximate MLE estimator by $\hat{\theta}_n = (\hat{c}_n, \hat{H})$. In this context, $\hat{c}_n$ is an estimator of scale, and $\hat{H}$ is an estimator of the Hurst parameter. Although we're interested here in data from Fractional Gaussian Noise, the technique is useful for the much larger class of stationary processes with long-range dependence [7, p. 42].

Let the spectral density

$$f(\lambda; \underline{\theta}) = \sigma^2 \pi^{-1} \sin(\pi H) \Gamma(2H + 1)(1 - \cos\lambda) \sum_{j=-\infty}^{\infty} |2\pi j + \lambda|^{-2H-1} \tag{2.21}$$

of Fractional Gaussian Noise with Hurst parameter H and variance σ^2 be parameterized by $\underline{\theta} = (\sigma^2, H) \in (\mathbf{R}^+ \times (0, 1))$. Note that the spectral density of standard Fractional Gaussian Noise is $f(\lambda; (1, H))$. For data $X(1), \ldots, X(n)$ from Fractional Gaussian Noise with variance σ_0^2 and Hurst parameter H_0 the goal is to estimate $\theta_0 = (\sigma_0^2, H_0) \in (\mathbf{R}^+ \times (0, 1))$.

The spectral density can be written in other forms. Specifically, we can write

$$f(\lambda; \underline{\theta}) = \sigma^2 f(\lambda; (1, H)) \tag{2.22}$$

$$= \sigma^2 \theta_1 f_1(\lambda; (1, H)) \tag{2.23}$$

where

$$\theta_1 = \theta_1(H) = \exp\left((2\pi)^{-1} \int_{-\pi}^{\pi} \log f(\lambda; (1, H) d\lambda \right) \tag{2.24}$$

is the special scale parameter such that

$$f(\lambda; (1, H)) = \theta_1 f_1(\lambda; (1, H)) \text{ and } \int_{-\pi}^{\pi} \log f_1(\lambda; (1, H)) \, \mathrm{d}\lambda = 0. \tag{2.25}$$

The estimation technique is based on minimizing

$$L_n(\underline{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda; \underline{\theta}) d\lambda + \frac{1}{n} Q_n(\underline{\theta}) \tag{2.26}$$

with respect to $\underline{\theta}$, where $Q_n(\underline{\theta}) = X^t A_n(\underline{\theta}) X$, $A_n(\underline{\theta}) = [\alpha(j-l)]_{j,l=1,\ldots,n}$ and

$$\alpha(j-l) = (2\pi)^{-2} \int_{-\pi}^{\pi} \frac{1}{f(\lambda; \underline{\theta})} \mathrm{e}^{i(j-l)\lambda} d\lambda. \tag{2.27}$$

Let $\underline{\theta}^* = (1, \eta)$ and $Q_n^*(\eta) = Q_n(\underline{\theta}^*)$. A simplification is achieved by choosing a special scale parameter $c = c(\eta) = \exp\left((2\pi)^{-1} \int_{-\pi}^{\pi} \log f(\lambda; (1, \eta) d\lambda \right)$ so that

$$f(\lambda, \underline{\theta}) = c f_1(\lambda, \underline{\theta}^*) \text{ and } \int_{-\pi}^{\pi} \log f_1(\lambda; \underline{\theta}^*) \, d\lambda = 0$$

which achieves the form in (2.23). To accomplish the minimization one simply solves

$$\frac{d}{d\eta} Q_n^*(\eta) = 0$$

to obtain $\hat{\eta}_n$. Let $\hat{c}_n = Q_n^*(\hat{\eta}_n)$. With this technique one obtains the approximate MLE estimator

$$\hat{\theta}_n = (\hat{c}_n, \hat{H}_n) = (Q_n^*(\hat{\eta}_n), \hat{\eta}_n)$$

of $\theta_0 = (\sigma_0^2\theta_1(H_0), H_0)$. Note that $\hat{c}_n$ is an estimator of the pre-factor $\sigma_0^2\theta_1(H_0)$ from the form in (2.23).

Lemma 2.3.1. *Let θ_0 and $\hat{\theta}_n$ be defined as above. Then*

$$\hat{\theta}_n \to \theta_0 \text{ almost surely, and} \tag{2.28}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\left((0,0)\ ,\ \begin{pmatrix} 2\sigma_0^4\theta_1^2(H_0) & 0 \\ 0 & c_{22}(H_0) \end{pmatrix} \right) \tag{2.29}$$

where

$$c_{22}(H_0) = \left(\frac{1}{\pi} \int_{-\pi}^{\pi} \left[\frac{\partial}{\partial \eta} \log f_1(\lambda; (1,\eta)) \right]^2 d\lambda \Big|_{\eta = H_0} \right)^{-1}.$$

Note in particular that $\hat{\theta}_n$ is $\sqrt{n}$-consistent and that $\hat{c}_n$ and $\hat{H}_n$ are asymptotically independent as $n \to \infty$.

The calculation of Whittle's approximate MLE estimator is computationally intensive. For example, the calculation of $[\alpha(j-l)]_{j,l=1,\ldots,n}$ requires the computation of n integrals for each trial value of η. A discrete version of the estimator is also possible which has computational advantages. Form the Fourier frequencies $\lambda_{j,n} = \frac{2\pi j}{n}$, $j = 1, \ldots, n^*$ where n^* is the integer part of $(n-1)/2$. Observe that $L_n(\underline{\theta})$ in (2.26) can be written as

$$L_n(\underline{\theta}) = \frac{1}{2\pi} \left[\int_{-\pi}^{\pi} \log f(\lambda; \underline{\theta})\, d\lambda + \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda)}\, d\lambda \right] \tag{2.30}$$

where $I_n(\lambda)$ is the periodogram for $X(1), \ldots, X(n)$. Using a Riemann sum, minimizing (2.30) amounts to minimizing

$$\tilde{Q}_n(\eta) = \sum_{j=1}^{n^*} \frac{I(\lambda_{j,n})}{f(\lambda_{j,n}; \theta^*)}$$

with respect to η. The advantage of this technique is that the periodogram can be calculated very quickly using the Fast Fourier transform. This discrete version of the estimator has been implemented in S-Plus with the function `whittle` by Beran [7, pp. 223-233] and Teverovsky [90].

Finally, note that the asymptotic distribution of the discrete version of the estimator $\hat{\theta}_n$ is the same whether the mean μ is known or replaced by $\bar{\mu}_n$ [7, p. 103]. It is easy to see that $\bar{\mu}_n = I_n(0)$ while $\hat{c}_n$ depends on $I_n(\lambda_{j,n})$, $j = 1, \ldots, n^*$. Yajima [96, p. 377] showed that $I_n(0)$ is asymptotically independent of $I_n(\lambda_{j,n})$, $\lambda_{j,n} \neq 0 \pmod{2\pi}$ for $j = 1, \ldots J$ for any finite J as $n \to \infty$. (See also Kunsch [53] and Beran [6].)

Chapter 3

Improved Fast Approximate Synthesis of Fractional Gaussian Noise

3.1 Introduction

A technique to generate approximate sample paths from Fractional Gaussian Noise (FGN) for teletraffic modelling was proposed by Paxson [74]. It is quite fast and has low storage requirements. Especially for Hurst parameter $H \leq 0.85$ the method produces sample paths that behave like FGN when judged with various statistical techniques such as Whittle's estimator, Beran's goodness of fit test, and the variance-time plot.

However, other simple statistical techniques clearly reveal a weakness in the method. The sample mean is indistinguishable from 0 rather than displaying true variability around 0, converging to 0 as the sample size increases. The sample variance of the sample mean doesn't fall like n^{2H-2}, where n is the sample path length. This would happen with true FGN. Generating a large number of sample paths all with the same length and Hurst parameter reveal more curiosities. The paths do not look like those of FGN when considering the Kolmogorov-Smirnov distance. Finally, for H very close to 1, every path is almost identically 0, where each path should have low variability around a value chosen from a Normal distribution. All of these problems can be traced, in large part, to one cause that is easily fixed.

The goal of this chapter is to propose a small fix for the benefit for those requiring fast, accurate sample paths of FGN. In fact, it will be demonstrated with simulations that sample paths behave much more like those obtained with the Cholesky decomposition of the covariance matrix. The latter technique, which will serve as the reference for accuracy, is essentially "exact" although much slower. Moreover limitations on the size of the matrix that can be decomposed tend to restrict the length of sample paths generated with that technique.

The rest of this chapter is organized as follows. Section 3.2 provides background to the simulation technique. Section 3.3 explains the proposed change. In Section 3.4 simulation results are presented to demonstrate the effectiveness of the change in better producing sample paths from FGN.

It is important to note that the periodogram used by Paxson

$$I(\omega_j) = \left| \sum_{k=1}^{n} e^{-i\omega_j k} Y_k \right|^2 \quad \text{where } \omega_j = \frac{2\pi j}{n},\ j = 1, 2, \ldots, n, \tag{3.1}$$

lacks the $1/n$ term often found in the definition [12, p. 332]. Thus, the method uses $I(\omega_j)$ to simulate Y_k with marginal $N(0, n)$ distributions. Never mentioned by Paxson, to obtain $N(0, 1)$ marginal distributions one would actually want to use $I(\omega_j)/n$ to find $X_k = Y_k/\sqrt{n}$. This can be achieved either by applying a transformation to the periodogram, or to the simulated data $\{Y_k\}$ at the very end to achieve the same effect, namely obtaining sample paths $\{X_k\}$ with marginal $N(0, 1)$ distributions. This notation for sample paths $\{Y_k\}$ and $\{X_k\}$, with marginal distributions from $N(0, n)$ and $N(0, 1)$, respectively, will be maintained throughout.

3.2 Proposed Change

True realizations of FGN should have variability in the sample mean. Moreover a sample path generated for $H = 0.5$ should behave as if it were a random sample from i.i.d data. This thesis tries to bring Paxson's method closer to this standard.

Throughout, a marginal $N(0, 1)$ distribution (standard FGN) is assumed. Extension to $N(0, \sigma^2)$ is straightforward. Consider the sample mean $\hat{\mu} = \sum_{k=1}^{n} X_k/n$ for n consecutive data points $X_1, \ldots, X_n$ from FGN with Hurst parameter $H \in [0.5, 1]$. It may be tempting to say that since FGN is self-similar with mean zero, the simulated sample paths should also have mean zero. But, for the *sample* mean of FGN, $\sum_{k=1}^{n} X_k/n^H \sim \mathrm{N}(0, 1)$ and thus $\hat{\mu} \sim \mathrm{N}(0, n^{2H-2})$.

The algorithm should have

$$\tilde{Y}_0 = \frac{\sum_{k=1}^{n} Y_k}{n} = \frac{\sum_{k=1}^{n} X_k}{\sqrt{n}} \sim \mathrm{N}(0, n^{2H-1}).$$

so that $z_0' \sim \mathrm{N}(0, n^{2H-1})$. In its current form the algorithm is different. In (2.20) $z_0' = 0$ which corresponds to the statement $\tilde{Y}_0 = 0$. But this fails to capture the appropriate randomness. In fact, it imposes a condition on the sample paths that $\sum_{k=1}^{n} X_k = 0$. So, for example, in the case $H = 0.5$ the familiar Kolmogorov-Smirnov distance clearly reveals that $\{X_k\}$ does not behave like a random sample of $N(0,1)$ data.

Thus the proposed change is simply to set $z_0' = n^{H-1/2}X$ where X is generated from a $\mathrm{N}(0,1)$ distribution, independent of z_j', $j > 0$. This explicitly creates $\tilde{Y}_0 \sim N(0, n^{2H-1})$ independent of other frequencies. In the S-plus implementation [73, 74] this can be easily implemented by changing one line in the function `ss.gen.fourier` from

```
zprime <- c(0, z, Conj(rev(z)[-1]))
```

to include z_0' with

```
z0 <- as.complex(rnorm(1, 0, (2*n)^(H - 0.5)))
zprime <- c(z0, z, Conj(rev(z)[-1])).
```

Note that `n` in the above code is actually the *half-length*, so it must be doubled.

This change could also be used in the faster version of Paxson's technique [54] which uses a linear approximation in the calculation of the power spectrum but includes the same problem in the sample mean.

3.3 Simulation Results

All results were produced using the S-plus implementation of the method in Paxson [73], or the modified version presented here. These simulations were performed on a Duron 600 PC with 512 MB of memory under Windows 98 running S-plus 2000. Unless otherwise noted, data was rescaled by $1/\sqrt{n}$ so the marginal distributions are approximately $N(0,1)$.

To demonstrate the effect of the modification on the sample mean, 1000 sample

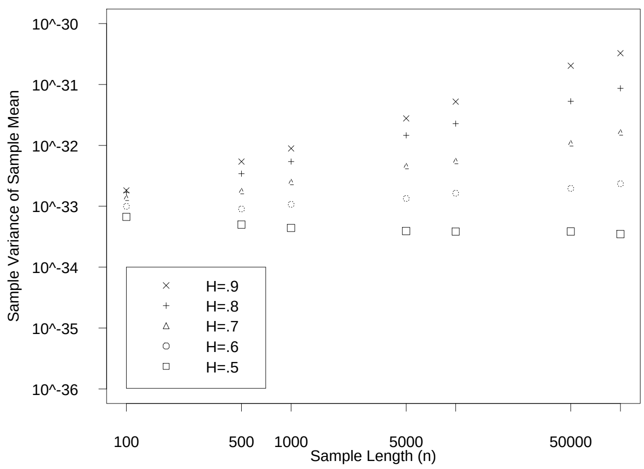

Figure 3.1: Sample variance of 1000 sample means using original code.

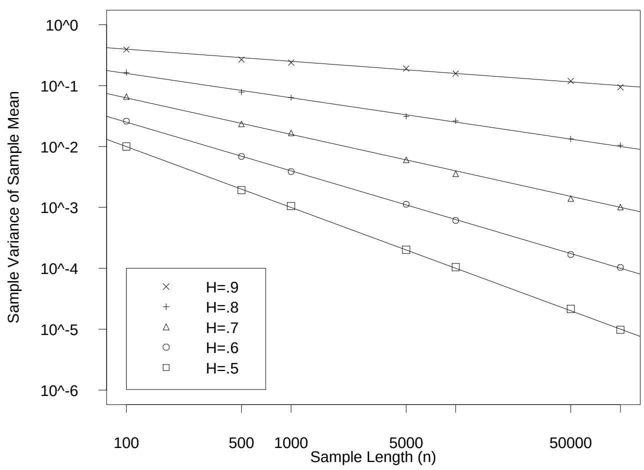

Figure 3.2: Sample variance of 1000 sample means using modified code.

paths of length n were generated, where n=100, 500, 1000, 5000, 10000, 50000, and 100000. The sample variance of these 1000 sample means was then calculated for each value of n. This procedure was performed for H=0.5, 0.6, 0.7, 0.8, and 0.9. The results from the original algorithm are shown in a log-log plot in Figure 3.1. Figure 3.2 shows analogous results using the modified algorithm. On a log-log plot the points should fall on lines of slope $2H - 2$. These lines are also plotted in Figure 3.2 for reference. The agreement is quite good. Notice that the sample variance of the sample mean is about 30 orders of magnitude too small. As well, for $H > 0.5$ the sample variance *increases* with larger sample sizes.

To investigate the effect of the change on the sample paths themselves, both techniques were used with the same initial random seed. Since the only difference is one additional random number generation at the end of the modified code, it is very easy to isolate the effect of the change. This technique makes it very clear to see the only change to the sample paths is the addition of the mean to every point of the sample path. Thus, there is no effect on the dependence structure of a sample path.

As another check on the long-range dependence properties, a variance-time plot analogous to Figure 2 in Paxson [74] was created. It was virtually identical, including the same small difference between actual and theoretical values.

To further demonstrate the effectiveness of this change Kolmogorov-Smirnov (KS) distances from sample paths of length n were considered. As above, paths with $N(0, 1)$ marginal distributions were generated. For each n, 5000 sample paths were generated, producing 5000 KS distances. Each KS distance was created using the S-

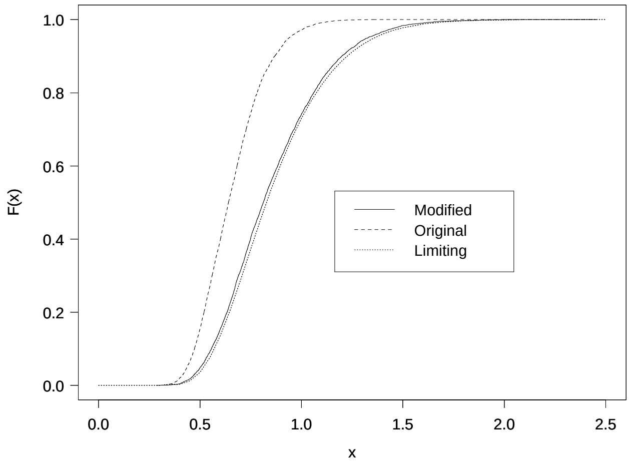

Figure 3.3: Kolmogorov-Smirnov distance distribution for i.i.d. data (n=100) from original and modified code and the known limiting distribution.

plus function `ks.gof` and a $N(0,1)$ distribution for the null hypothesis. Each distance was multiplied by n^{2-2H} [7, p. 199]. (Note that no parameters were estimated.) This technique produces 5000 points to generate the empirical Kolmogorov-Smirnov distance distribution.

Figure 3.3 shows a typical comparison for H=0.5 and n=100; larger n produces similar results. In this special case the data should be i.i.d. The limiting distribution as $n \to \infty$ is the well-known asymptotic Kolmogorov-Smirnov distance distribution [21, p. 467]. The empirical distribution from the original code is quite different from the known limit. This suggests, from the point of view of the Kolmogorov-Smirnov test, the sample paths do not look like i.i.d. $N(0,1)$ data. On the other hand, the empirical distribution from modified code is very close to the limiting distribution. The sample paths behave more like i.i.d. $N(0,1)$ data.

Figure 3.4 shows a typical comparison for H=0.7 and n=100. Empirical KS

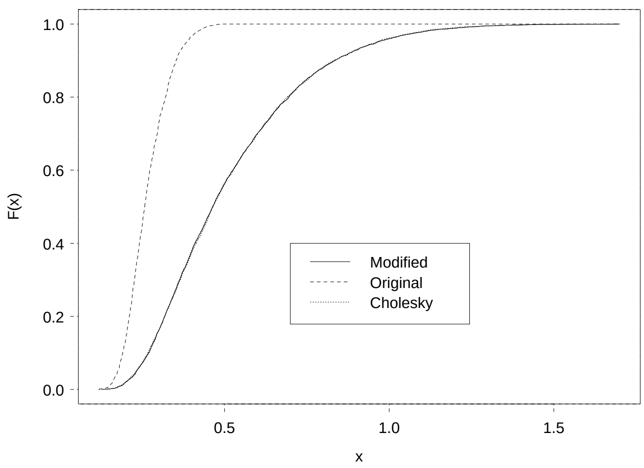

Figure 3.4: Kolmogorov-Smirnov distance distribution for data (n=100, H=0.7) from original and modified code and from Cholesky decomposition.

distributions for both the original and modified code are shown. In this case the limiting distribution is different and not shown. Instead sample paths of standard FGN were generated using the Cholesky decomposition of the FGN covariance matrix. These sample paths should be "exact"- no approximations are used. These sample paths were used to create another empirical KS distance distribution, which is shown. This is a more useful reference than the theoretical limiting distribution because of slow convergence. Again, in the context of the KS distance, sample paths from the original code can clearly be distinguished from Cholesky data, while the modified code appears very similar.

As an example of the advantages of the modified code, convergence of sample p-quantiles $\hat{x}_{p,H}$ for FGN data with Hurst parameter H is considered. As above, sample paths with lengths between 100 and 10000 and a range of Hurst parameters are used. Results for the .25-quantile are shown here, and are typical.

Figure 3.5 shows the empirical distribution for 5000 sample .25-quantiles from

FGN with length 1000 and $H = 0.7$ produced with the original and modified code, and the Cholesky technique. The vertical line corresponds to the true value of the quantile, and the horizontal line corresponds to the probability 0.5. So, all three techniques show a sample .25-quantile centered at the true quantile. However the *distribution* of the sample quantiles are different. Again, the Cholesky and modified Paxson techniques agree, while Paxson's original technique produces a different result.

A closer examination of the variance of the sample quantiles is more revealing. It can be shown that for $H = 0.5$, $\mathrm{Var}(\hat{x}_{p,0.5}) \sim 1/n$ as $n \to \infty$ [84] and more generally $\mathrm{Var}(\hat{x}_{p,H}) \sim n^{2H-2}$ as $n \to \infty$ for $0.5 < H < 1$ [85]. Figure 3.6 shows the sample variances for 5000 values of $\mathrm{Var}(\hat{x}_{.25,0.5})$ for various sample paths on a log-log plot. A reference line of slope -1 is included. All three sets of data show the same rate of decrease, although the data from Paxson's original code is displaced from the others. Figure 3.7 shows the sample variances for 5000 values of $\mathrm{Var}(\hat{x}_{.25,0.8})$. This plot is typical of quantiles in the long-range dependent case $0.5 < H < 1$. A reference line of

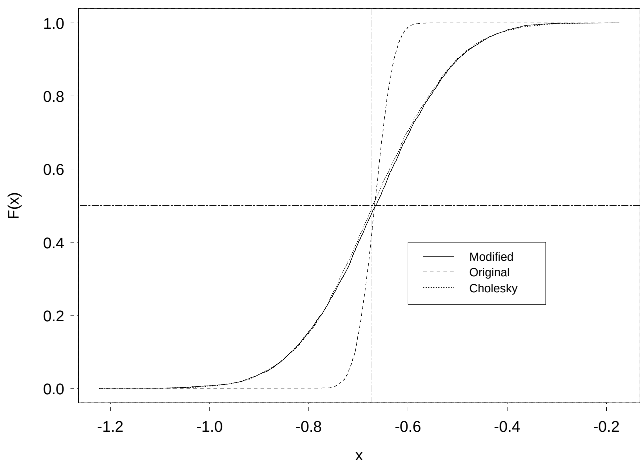

Figure 3.5: Empirical distribution of 5000 sample .25-quantiles from FGN of length 1000, $H = 0.7$

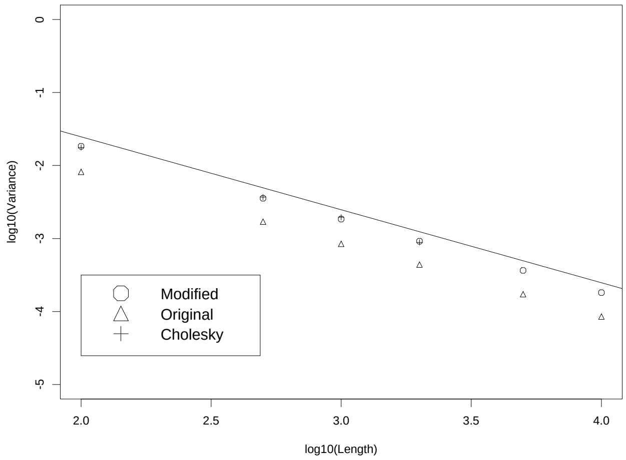

Figure 3.6: Sample variances of 5000 sample .25-quantiles for $H = 0.5$ and various lengths.

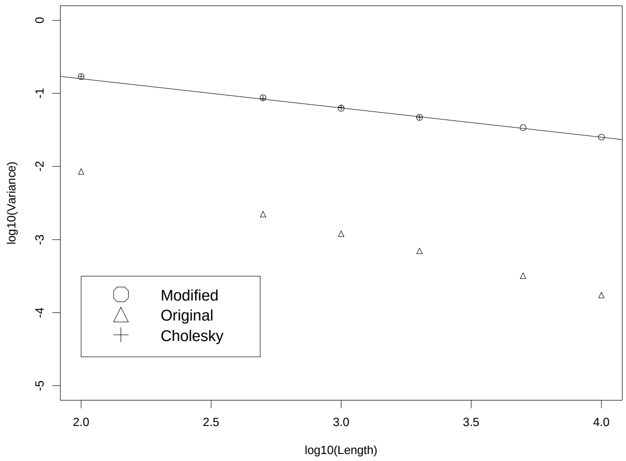

Figure 3.7: Sample variances of 5000 sample .25-quantiles for $H = 0.8$ and various lengths.

slope -0.4 is included. While Cholesky and modified Paxson show a rate of decrease in agreement with theory, data from Paxson's original code clearly decays at a different rate. This discrepancy grows from $H = 0.5$ with increasing H.

Chapter 4

Variance Estimation of Fractional Gaussian Noise

The goal of this chapter is to propose two estimators for the variance of Fractional Gaussian Noise. While interesting in their own right, these estimators are appropriate for parameter estimation in models proposed in the sequel. Two estimators are proposed. In Section 4.1 a variance estimator based on Whittle's maximum likelihood estimator technique is proposed. In Section 4.2, a variance estimator is proposed that uses an estimate of the Hurst parameter to unbias the sample variance. Section 4.3 presents simulation results.

4.1 Whittle's Approximate Maximum Likelihood Estimator

As described in Section 2.3.2, Whittle's approximate maximum likelihood estimator described in Beran [7, pp. 116-119, 156-157] provides a joint estimate

$$\hat{\theta}_n = (\hat{c}_n, \hat{H})$$

of both a scale and Hurst parameter for long-memory Gaussian processes such as Fractional Gaussian Noise. However, the estimated scale parameter $\hat{c}_n$ is not an estimator of the variance, but a function of it. To estimate the variance, σ_0^2, the joint estimator

$$\hat{\Theta}_n = \left(\frac{\hat{c}_n}{\theta_1(\hat{\eta}_n)}, \hat{\eta}_n \right),$$

is proposed. In particular, we have $\hat{\sigma}_n^2 = \frac{\hat{c}_n}{\theta_1(\hat{\eta}_n)}$. The properties of $\hat{\Theta}_n$ are a consequence of the properties of $\hat{\theta}_n$.

Lemma 4.1.1. *Let $\Theta_0 = (\sigma_0^2, H_0)$ and $\hat{\Theta}_n$ be defined as above. Then*

$$\hat{\Theta}_n \to \Theta_0 \text{ almost surely, and} \tag{4.1}$$

$$\sqrt{n}(\hat{\Theta}_n - \Theta_0) \xrightarrow{D} N\left((0,0)\ ,\ \begin{pmatrix} 2\sigma_0^4 & 0 \\ 0 & c_{22}(H_0) \end{pmatrix} \right) \tag{4.2}$$

where

$$c_{22}(H_0) = \left(\frac{1}{\pi} \int_{-\pi}^{\pi} \left[\frac{\partial}{\partial \eta} \log f_1(\lambda; (1, \eta)) \right]^2 d\lambda \bigg|_{\eta = H_0} \right)^{-1}.$$

Note in particular that $\hat{\Theta}_n$ is consistent, and that $\hat{\sigma}_n^2$ and $\hat{H}_n$ are asymptotically independent as $n \to \infty$.

Proof. By Theorems 5.1, 5.3 and the discussion thereafter in Beran [7, p. 105-106, 112],

$$\hat{\theta}_n \to \theta_0 \text{ almost surely, and}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}\left((0,0), C(\sigma_0^2\theta_1(H_0), H_0)\right)$$

where

$$C(\sigma_0^2\theta_1(H_0), H_0) = \begin{pmatrix} 2\sigma_0^4\theta_1^2(H_0) & 0 \\ 0 & c_{22}(H_0) \end{pmatrix}.$$

To prove almost sure convergence first define the events

$$A = [\lim_{n\to\infty} \hat{\eta}_n = H_0],$$

$$B = [\lim_{n\to\infty} (\hat{c}_n, \hat{\eta}_n) = (\sigma_0^2\theta_1(H_0), H_0)],$$

$$C = [\lim_{n\to\infty} \theta_1(\hat{\eta}_n) = \theta_1(H_0)], \text{ and}$$

$$D = [\lim_{n\to\infty} (\frac{\hat{c}_n}{\theta_1(\hat{\eta}_n)}, \hat{\eta}_n) = (\sigma_0^2, H_0)].$$

By definition, from (2.28) $\mathrm{P}(A) = 1$ and) $\mathrm{P}(B) = 1$. Thus, by continuity of $\theta_1(H)$ on $(0,1)$, $\mathrm{P}(C) = 1$. Then $\mathrm{P}(D) = \mathrm{P}(C \cap D) = 1$ and $\hat{\Theta}_n \to \Theta_0$ almost surely.

To prove convergence in distribution, note $\sqrt{n}(\hat{c}_n - \sigma_0^2\theta_1(H_0)) \xrightarrow{D} Z_1$ as $n \to \infty$ where $Z_1 \sim \mathrm{N}(0, 2\sigma_0^2\theta_1(H_0))$ and $\theta_1(\hat{\eta}_n) \xrightarrow{D} \theta_1(H_0)$ as $n \to \infty$. By Slutsky's theorem, $\frac{\hat{c}_n}{\theta_1(\hat{\eta}_n)} \xrightarrow{D} \frac{Z_1}{\theta_1(H_0)}$ as $n \to \infty$. This can be used to prove the bivariate result. $\square$

A simple analytic formula for $c_{22}(H)$ isn't known. Figure 4.1 shows a plot for $0.5 < H < 1$, obtained using the function `CetaFGN` written in S-Plus by Beran [7, pp. 224-225]. Note that $c_{22}(H)$ is increasing, and it increases by about 21% over the interval.

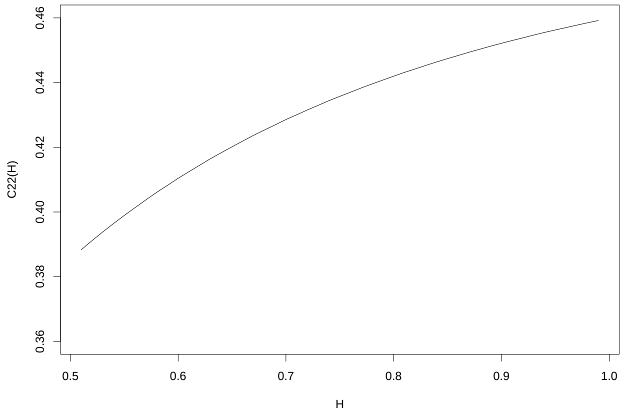

Figure 4.1: Values of $c_{22}(H)$

One could obtain the estimator $\hat{\sigma}_n^2$ by dividing by θ_1, but this requires evaluating the integral in (2.25). It is more convenient to calculate $\theta_{1,n}$ using a Riemann sum such as

$$\theta_{1,n} = \frac{4}{n} \sum_{j=1}^{n^*} \log f\left(\lambda_{j,n}; (1, \eta)\right)$$

where $\lambda_{j,n} = \frac{2\pi j}{n}$, $j = 0, \ldots, n^*$ are the Fourier frequencies and $n^* = \lfloor \frac{n-1}{2} \rfloor$. Let the resulting estimator be $\hat{\Theta}_{n,n} = (\frac{\hat{c}_n}{\theta_{1,n}}, \hat{\eta}_n)$.

Lemma 4.1.2. *The limiting properties for $\hat{\Theta}_{n,n}$ are the same as for $\hat{\Theta}_n$.*

Proof. By the convergence of the Riemann sum to the corresponding integral, $\theta_{1,n}(\eta) \to \theta_1(\eta)$ as $n \to \infty$. The proof using $\theta_{1,n}(\eta)$ then follows the proof using $\theta_1(\eta)$. $\square$

A discrete version of the estimator $\hat{\theta}_n$ is already implemented in S-Plus with the function `whittle` by Beran[7, pp. 223-233] and Teverovsky[90]. To implement the estimator of variance proposed here a change is necessary. Their technique uses a parameterization of the spectral density that yields $\hat{\theta}_n = (\hat{c}_n,\ \hat{H})$, with the two elements called 'sigma2' and 'eta', respectively. We actually want an estimator $\hat{\Theta}_{n,n} = (\hat{\sigma}_n^2,\ \hat{H})$, where $\hat{\sigma}_n^2$ is the proposed estimator of the variance. In fact,

$$\hat{\sigma}_n^2 = \frac{\hat{c}_n}{\theta_{1,n}} \text{ where } \theta_{1,n} = \frac{4}{n}\sum_{j=1}^{n^*} \log f\left(\lambda_{j,n}; (1,\eta)\right),$$

$\lambda_{j,n} = \frac{2\pi j}{n}$, $j = 0, \ldots, n^*$ are the Fourier frequencies and $n^* = \lfloor \frac{n-1}{2} \rfloor$. The properties of $\hat{\Theta}_{n,n}$ follow directly from those of $\hat{\theta}_n$.

4.2 Sample Variance

For comparison it is interesting to include a sample variance-based estimator. The usual sample variance S^2 is a biased estimator of the variance σ^2 of Fractional Gaussian Noise with Hurst parameter $H \in (0,1), H \neq 0.5$ [7, p. 156-157]. (See Appendix A for a simple algebraic proof of the bias.) If H is known an unbiased estimator is

$$S_{n,H}^2 = \frac{n-1}{n - n^{2H-1}} S^2.$$

This is similar to the unbiased estimator suggested by Beran [7, p. 156], although he suggests the reciprocal of the pre-factor. However, the reciprocal of the pre-factor does not produce an unbiased estimator for FGN. This is easy to see heuristically. By Beran [7, p. 9] the positive correlations of FGN mean that S_n^2 *underestimates* the variance. Therefore, an unbiased estimator requires a multiplicative pre-factor *greater* than 1.

For unknown $H \in (0.5, 1)$, $H \neq 0.75$ the estimator

$$S^2_{n,\hat{H}_n} = \frac{n-1}{n - n^{2\hat{H}_n - 1}} S^2.$$

is proposed, where $\hat{H}_n$ is the Whittle estimate of H. Beran [7, p. 156] has suggested simply that H be a consistent estimator. The estimation of H introduces some bias.

The case $H = 0.75$ is specifically excluded. (Since H is imagined to run continuously between 0.5 and 1 this exclusion is not severe.) As $n \to \infty$, for $H \in (0.5, 0.75)$, S^2 is asymptotically Gaussian, but for $H \in (0.75, 1)$ the limit is non-Gaussian with a complicated distribution [6, p. 67-73].

It is not intended that the sample variance-based estimator proposed here should be considered superior to the estimator based on Whittle's approximate MLE. It is presented mainly as an alternative for comparison. As discussed in Section 2.3.2, Whittle's technique has much nicer asymptotic properties. Also, with that technique both the variance and the Hurst parameter can be estimated at once. With the sample variance-based estimator, the Hurst parameter must be estimated first so the estimation of the variance is a two step process.

The following lemma is useful for considering $S^2_{n,\hat{H}_n}$.

Lemma 4.2.1. *If $\hat{H}_n$ is a consistent estimator of H, then for the constants $a > 0$, $b > 0$*

$$n^{a(\hat{H}_n - H - b)} \xrightarrow{P} 0$$

Proof. Without loss choose $0 < \varepsilon < \min\{b, 1\}$. By definition of consistency, there exists n_0 such that

$$\mathrm{P}\left(|\hat{H}_n - H| > \varepsilon\right) < \varepsilon \quad \forall n > n_0.$$

Since $0 < \varepsilon < 1$, $\ln \varepsilon < 0$ and so $\left(b + \frac{\ln \varepsilon}{a \ln n}\right) \nearrow b$ as $n \to \infty$. Let $n_1 = \varepsilon^{\frac{1}{a(\varepsilon - b)}}$ so that $\varepsilon = b + \frac{\ln \varepsilon}{a \ln n_1}$. Then for $n > \max\{n_0, n_1\}$

$$\begin{aligned}
\mathrm{P}\left(n^{a(\hat{H}_n - H - b)} > \varepsilon\right) &= \mathrm{P}\left(\hat{H}_n - H > b + \frac{\ln \varepsilon}{a \ln n}\right) \\
&< \mathrm{P}\left(\hat{H}_n - H > \varepsilon\right) \\
&\leq \mathrm{P}\left(|\hat{H}_n - H| > \varepsilon\right) \\
&< \varepsilon
\end{aligned}$$

and the result is proven. $\square$

Theorem 4.2.2. *If $H \in (0.5, 1)$, $H \neq 0.75$*

$$S^2_{n,\hat{H}_n} \xrightarrow{P} \sigma^2 \ \textit{as}\ n \to \infty$$

Proof. By Slutsky's theorem and Lemma 4.2.1 with $a = 2$ and $b = 1 - H$,

$$\frac{n-1}{n - n^{2\hat{H}_n - 1}} = \frac{1 - n^{-1}}{1 - n^{2\hat{H}_n - 2}} \xrightarrow{P} 1.$$

For the sample variance,

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i^2,$$

and since FGN is ergodic, X_i^2 is ergodic with a $\sigma^2 \chi^2_{(1)}$ distribution with mean σ^2. Thus $S_n^2 \xrightarrow{P} \sigma^2$ by ergodicity, and the result follows. $\square$

4.3 Simulation Results for Variance Estimators

In this section, simulation results are presented for the two proposed estimators of the variance. All simulations were again performed on a Duron 600 computer with 512 MB of memory running Windows 98 in S-plus 2000. Sample paths of FGN were simulated using two techniques. The first technique, referred to as "Cholesky", transforms i.i.d. Normal random variables using the Cholesky decomposition of the covariance matrix of FGN [7, p. 215]. It should be exact, although there can be a problem with precision with extremely small values. Storage limitations preclude sample paths larger than about 4000. Results presented here are for sample paths of length $n = 50$, 100, 500, and 1000.

The second simulation technique is referred to as "Modified Paxson". It is the technique described in Chapter 3. Because of its higher speed and low storage requirements, longer sample paths are possible. Although not presented here, paths at least as long as 5x10^6 have been simulated with this technique. Here results are presented for paths of length $n = 50$, 100, 500, 1000, 5000, 10000.

For each technique and sample path length, 1000 sample paths for $H = 0.5$, 0.6, 0.7, 0.8, and 0.9 were simulated- so 50 000 sample paths in total. For each sample path a collection of summary data including the sample mean $\bar{\mu}_{n,i}$, the Whittle estimator of H, $\hat{H}_{n,i}$ and the variance estimators $\hat{\sigma}^2_{n,i}$ and $S^2_{n,\hat{H}_{n,i},i}$, $i = 1, \ldots, 1000$ were calculated. Simulation results using the sample mean and sample variance estimators (biased for FGN) were presented in Chapter 3 and are not shown here. The underlying assumptions on the periodogram for the Whittle estimate are exactly those made

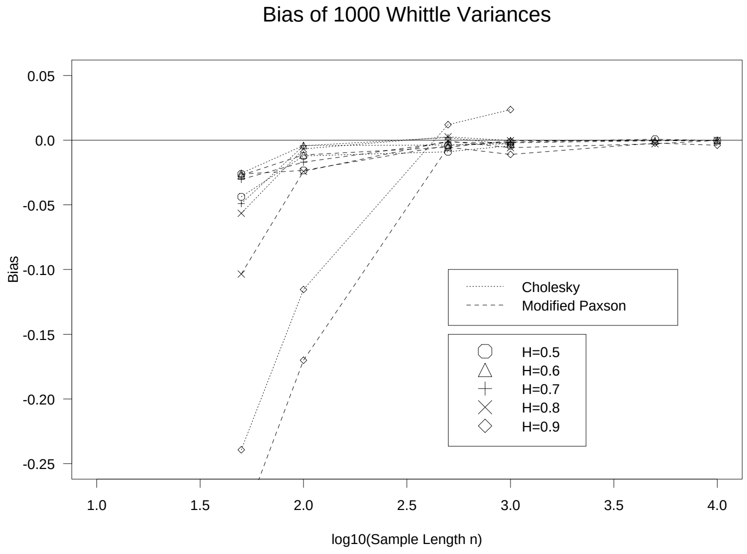

Figure 4.2: Bias of 1000 Whittle estimates $\hat{\sigma}_n^2$ of variance. The sample paths were generated using the Cholesky and Modified Paxson techniques.

with Modified Paxson. Thus, it is really a best case situation when the combination is used together.

Figure 4.2 shows the bias of $\hat{\sigma}_n^2$, defined as

$$\frac{1}{1000}\sum_{i=1}^{1000}\hat{\sigma}_{n,i}^2 - 1.$$

In this plot the horizontal axis has $\log_{10}$ scale. The horizontal line corresponds to the ideal of no bias. One value is below the graph at -0.30. As expected, shorter sample paths have greater bias. For $H = 0.9$ the bias is more extreme and converges slowly. Apart from the Cholesky case with $H = 0.9$, the bias is decreasing and negligible- for paths longer than 500 the bias is less than 2%.

Figure 4.3 shows similar results for the bias of $S^2_{n,\hat{H}_n}$, defined as

$$\frac{1}{1000}\sum_{i=1}^{1000} S^2_{n,\hat{H}_{n,i},i} - 1.$$

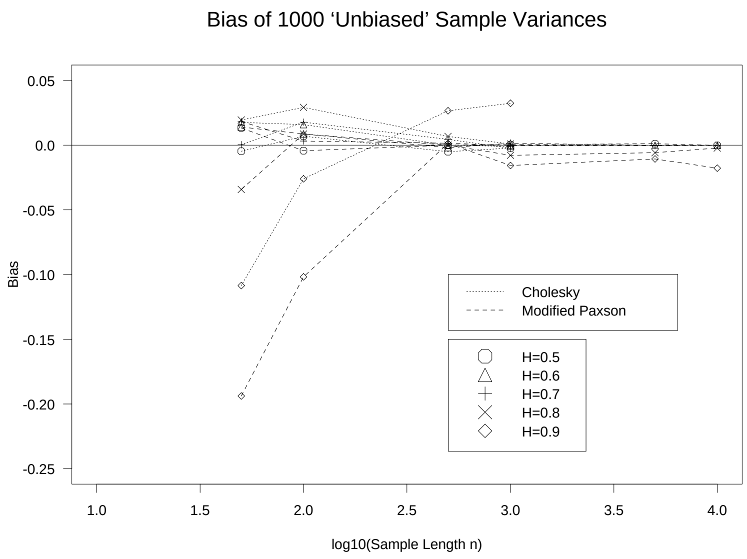

Figure 4.3: Bias of 1000 'unbiased' sample variance estimates $S^2_{n,\hat{H}_n}$ of variance. The sample paths were generated using the Cholesky and Modified Paxson techniques.

These results are similar to those for the bias of $\hat{\sigma}^2_n$ above. Apart from the combination of large H and small sample paths, the bias is quite small.

Figure 4.4 shows simulation results for the sample variance of the Whittle estimator of variance, defined as

$$\frac{1}{999}\sum_{i=1}^{1000}\left(\hat{\sigma}^2_{n,i}\right)^2 - \left(\frac{1}{1000}\sum_{i=1}^{1000}\hat{\sigma}^2_{n,i}\right)^2$$

on a $\log_{10}$-$\log_{10}$ plot. The solid line corresponds to a value of $\frac{2}{n}$ which is suggested by the asymptotic result for the variance of $\hat{\sigma}^2_n$ in Theorem 2.29. The agreement between the Cholesky and Modified Paxson techniques is quite good. Although the use of Modified Paxson is best-case, the results are not exceptional. Note that, apart from the typical slow convergence for $H = 0.9$, the variance appears to decay at a rate proportional to n^{-1} in all cases. For $H = 0.5$, 0.6, and 0.7, the agreement with

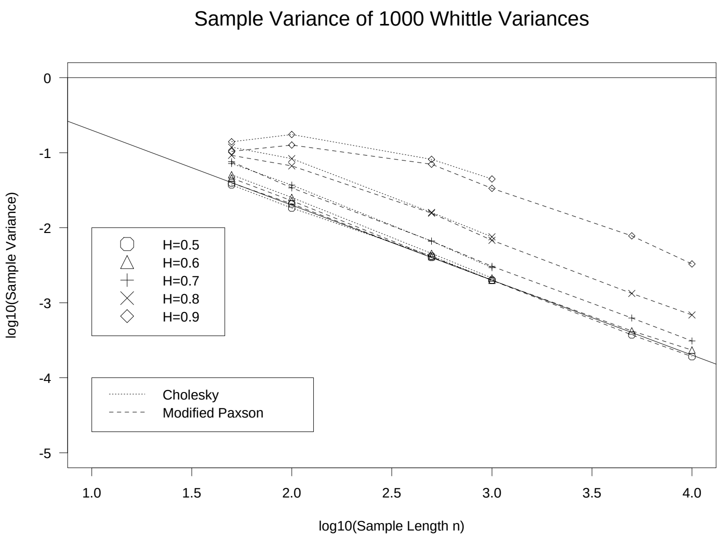

Figure 4.4: Sample variance of 1000 Whittle estimates $\hat{\sigma}_n^2$ of variance. The sample paths were generated using the Cholesky and Modified Paxson techniques.

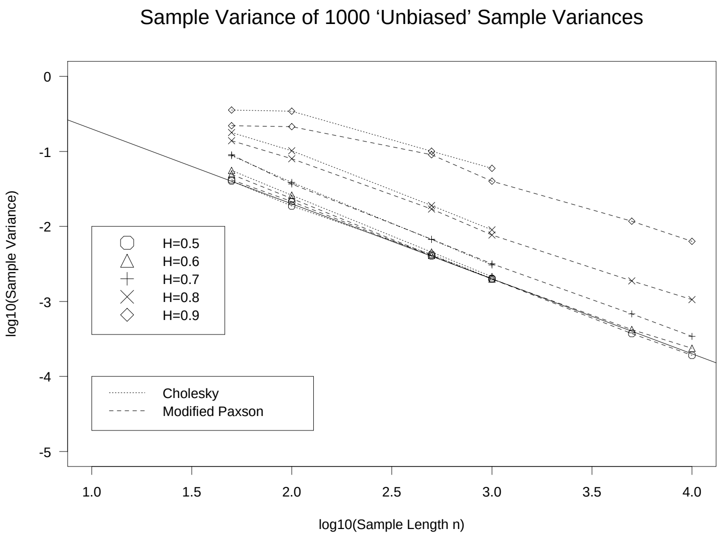

Figure 4.5: Sample variance of 1000 'unbiased' sample variance estimates $S^2_{n,\hat{H}_n}$ of variance. The sample paths were generated using the Cholesky and Modified Paxson techniques.

$2/n$ is quite good. For the biggest differences, the variance is abut 3.5 times larger for $H = 0.8$, and 16 times larger for $H = 0.9$.

Figure 4.5 shows results for the sample variance of the 'unbiased' sample variance estimator of variance, defined as

$$\frac{1}{999}\sum_{i=1}^{1000}\left(S^2_{n,\hat{H}_{n,i},i}\right)^2 - \left(\frac{1}{1000}\sum_{i=1}^{1000} S^2_{n,\hat{H}_{n,i},i}\right)^2,$$

also on a $\log_{10}$-$\log_{10}$ plot. The results are quite similar to those for the Whittle estimator above. Careful inspection reveals, especially for $H = 0.9$, the rate of decay differs from n^{-1}, as suggested by theory.

Chapter 5

Random Superposition of Alternating On-Off Sources

5.1 Introduction

This chapter extends the results of Taqqu, Willinger and Sherman [88] to consider the case in which, a priori, the number of sources is large and random. For example, in any block of time (e.g. afternoons) the number of sources may be essentially constant, the traffic appears stationary, the previously mentioned results apply, and Fractional Brownian Motion is the appropriate limit. But, considered across separate, independent blocks of time (e.g. afternoons from day to day), differing numbers of sources can introduce additional variability, even with all other factors the same. Modelling this situation, limiting processes that are scale mixtures of Fractional Brownian Motion with random drift are obtained.

The technique used here relies on limit theorems for "growing" random sums of independent random variables. Initial limit theorems for growing random sums are due to Robbins [77] and Dobrushin [22]. Korolev [49, 51] provides a unifying treatment for random sums of one dimensional random variables. For multi-dimensional random variables, analogs to those for one dimension are proven in the context of linear operators [50]. Necessary and sufficient conditions for the convergence of distributions of random sums of independent random variables to strictly stable laws are proven by extension [52].

5.2 Finite-Dimensional Distributions

The key departure from Taqqu *et. al.* [88] is the consideration of a "growing" random sum of on-off processes following results from Korolev [51, 52]. Notation for the on-off processes is the same as in Section 2.1.2. As in [52], let N_n be a sequence of positive integer-valued random variables, always assumed independent of $\{W^{(m)}(t), t > 0\}$ for each n, such that $N_n \xrightarrow{P} \infty$ as $n \to \infty$. Let U be a non-negative valued random variable, such that $\mathrm{P}(U \geq 0) = 1$ and

$$\frac{N_n}{d_n^2} \xrightarrow{D} U^2 \text{ as } n \to \infty \tag{5.1}$$

for a sequence of positive constants $\{d_n\}$ such that $d_n \to \infty$.

Lemma 5.2.1. *Let N_n, U and d_n satisfy (5.1). Then for any $t \geq 0$*

$$\mathcal{L} \lim_{n\to\infty} \frac{\int_0^t \sum_{i=1}^{N_n} \left(W^{(m)}(u) - \mu_W\right) du}{d_n} = U \int_0^t G(u)\, du \tag{5.2}$$

where $G(t)$ is the mean zero stationary Gaussian process with autocovariance $r(t)$ described after (2.16), and $G(t)$ and U are independent for all t.

Proof. For any natural r, $(t_1, t_2, \ldots, t_r)$ and $(a_1, a_2, \ldots, a_r)$ where $t_i \geq 0$ and $a_i \in \mathbf{R}$, $i = 1, \ldots, r$, by the definition of $\{G(t), t \geq 0\}$,

$$\frac{1}{\sqrt{M}} \sum_{i=1}^{r} a_i \left(\sum_{m=1}^{M} (W^{(m)}(t_i) - \mu_W) \right) \xrightarrow{D} \sum_{i=1}^{r} a_i G(t_i) \text{ as } M \to \infty$$

by the Cramér-Wold Theorem [8, p. 49]

Let $Y_m = \sum_{i=1}^{r} a_i \int_0^{t_i} \left(W^{(m)}(t_i) - \mu_W \right) du$. Clearly, $\{Y_m,\, m = 1, 2, \ldots\}$ is an i.i.d. sequence since $\{W^{(m)}(t),\, m = 1, 2, \ldots;\, t \geq 0\}$ are i.i.d. copies of $\{W(t),\, t \geq 0\}$. Since $\{W^{(m)}(t),\, m = 1, 2, \ldots;\, t \geq 0\}$ are i.i.d. processes taking values between $-\mu_W$ and $1 - \mu_W$ it follows that

$$-\mu_W \sum_{i=1}^{r} a_i t_i \leq Y_m \leq (1 - \mu_W) \sum_{i=1}^{r} a_i t_i$$

and

$$-\mu_W \sum_{i=1}^{r} a_i t_i \leq \mathrm{E}[Y_m] \leq (1 - \mu_W) \sum_{i=1}^{r} a_i t_i.$$

Since $\mathrm{E}[Y_m]$ is finite, Fubini's Theorem [66, p. 294] gives

$$\begin{aligned} \mathrm{E}[Y_m] &= \mathrm{E}\left[\sum_{i=1}^{r} a_i \int_0^{t_i} \left(W^{(m)}(u) - \mu_W \right) du \right] \\ &= \sum_{i=1}^{r} a_i \int_0^{t_i} \mathrm{E}\left[W^{(m)}(u) - \mu_W \right] du \\ &= 0 \end{aligned}$$

Moreover, since $(W^{(m)}(t) - \mu_W)$ and $G(t)$ also have the same autocovariance $r(t)$,

$$\begin{aligned}\mathrm{Var}[Y_m] &= \mathrm{Var}\left[\sum_{i=1}^{r} a_i \int_0^{t_i} \left(W^{(m)}(t_i) - \mu_W\right)\, du\right]\\ &= \mathrm{Var}\left[\sum_{i=1}^{r} a_i \int_0^{t_i} G(u)\, du\right],\end{aligned}$$

and so

$$\frac{1}{\sqrt{M}} \sum_{m=1}^{M} Y_m \xrightarrow{D} Z \text{ as } M \to \infty$$

by the Central Limit Theorem, where $Z \sim N\left(0, \mathrm{Var}\left[\sum_{i=1}^{r} a_i \int_0^{t_i} G(u)\, du,\right]\right)$.

Then

$$\begin{aligned}\frac{1}{d_n} \sum_{i=1}^{r} a_i \int_0^{t_i} \sum_{m=1}^{N_n} \left(W^{(m)}(u) - \mu_W\right)\, du &= \frac{1}{d_n} \sum_{m=1}^{N_n} Y_m\\ &\xrightarrow{D} U \sum_{i=1}^{r} a_i \int_0^{t_i} G(u)\, du \text{ as } n \to \infty\\ &= \sum_{i=1}^{r} a_i \left(U \int_0^{t_i} G(u)\, du\right)\end{aligned}$$

where the first line follows from changing the order of the summation and the limit follows from Korolev [51, p. 290]. Independence of U and $G(t)$ follows from the independence of $\{W^{(m)}(t)\}$ and N_n. By the Cramér-Wold Theorem the result for the finite-dimensional distributions follows. $\square$

The real quantity of interest is the integrated rewards from the random number of sources. This can be imagined as modelling the cumulative volume of traffic (e.g. in bytes or packets). Let

$$W^*_{N_n}(Tt) = \int_0^{Tt} \left(\sum_{m=1}^{N_n} W^{(m)}(u)\right) du.$$

Theorem 5.2.2. *For large T, a random number of sources N_n, a random variable U and a sequence of positive constants d_n as in (5.1) above, the aggregate packet process $\{W^*_{N_n}(Tt), t \geq 0\}$ behaves stochastically like*

$$N_n T \frac{\mu_1}{\mu_1 + \mu_2} t + \sigma_{lim} \sqrt{T^{2H} L(T)} d_n U Z_H(t) \tag{5.3}$$

where $H = (3 - \alpha_{min})/2$, σ_{lim} is given by (2.14) or (2.15), and $Z_H(t)$ is standard Fractional Brownian Motion independent of U. Specifically

$$\mathcal{L} \lim_{T\to\infty} \mathcal{L} \lim_{n\to\infty} \frac{\int_0^{Tt} \sum_{m=1}^{N_n} \left(W^{(m)}(u) - \frac{\mu_1}{\mu_1+\mu_2} \right) du}{d_n \sqrt{T^{2H} L(T)}} = \sigma_{lim} U Z_H(t) \ , \ t \geq 0 \tag{5.4}$$

Proof. By Lemma 5.2.1

$$\mathcal{L} \lim_{n\to\infty} \frac{\int_0^{Tt} \sum_{m=1}^{N_n} \left(W^{(m)}(u) - \frac{\mu_1}{\mu_1+\mu_2} \right) du}{d_n} = U \int_0^{Tt} G(u)\, du.$$

But, by Theorem 1 of Taqqu *et. al.* [88]

$$\mathcal{L} \lim_{T\to\infty} \frac{1}{\sqrt{T^{2H} L(T)}} \int_0^{Tt} G(u)\, du = \sigma_{lim} Z_H(t), \ t \geq 0$$

where $H = (3 - \alpha_{min})/2$. The desired result follows directly. □

Corollary 5.2.3. *If, in addition to (5.1), N_n, d_n and U satisfy*

$$\frac{N_n - U^2 d_n^2}{d_n} \xrightarrow{P} 0 \ as \ n \to \infty, \tag{5.5}$$

$\{W^*_{N_n}(Tt), t \geq 0\}$ behaves stochastically like

$$U^2 d_n^2 T \frac{\mu_1}{\mu_1 + \mu_2} t + \sigma_{lim} \sqrt{T^{2H} L(T)} d_n U Z_H(t). \tag{5.6}$$

Specifically

$$\mathcal{L} \lim_{T\to\infty} \mathcal{L} \lim_{n\to\infty} \frac{W^*_{N_n}(Tt) - U^2 d_n^2 T \frac{\mu_1}{\mu_1+\mu_2} t}{d_n \sqrt{T^{2H} L(T)}} = \sigma_{lim} U Z_H(t) \ , \ t \geq 0. \tag{5.7}$$

Proof. Let

$$X_n(t) = \frac{\int_0^{Tt} \sum_{m=1}^{N_n} \left(W^{(m)}(u) - \frac{\mu_1}{\mu_1+\mu_2} \right) du}{d_n},$$

$$R_n = T \frac{\mu_1}{\mu_1 + \mu_2} \frac{N_n - U^2 d_n^2}{d_n}, \text{ and}$$

$$\begin{aligned} Y_n(t) &= \frac{W^*_{N_n}(Tt) - U^2 d_n^2 T \frac{\mu_1}{\mu_1+\mu_2} t}{d_n} \\ &= X_n(t) + Tt \frac{\mu_1}{\mu_1 + \mu_2} \frac{N_n - U^2 d_n^2}{d_n} \\ &= X_n(t) + tR_n. \end{aligned}$$

The Cramér-Wold Theorem and $R_n \xrightarrow{P} 0$ combine to show that $\mathcal{L} \lim_{n\to\infty} Y_n(t) = \mathcal{L} \lim_{n\to\infty} X_n(t)$. The rest of the proof follows as in Theorem 5.2.2. □

Corollary 5.2.4. *Assume now there are R classes of independent on-off sources. Within each class let the descriptors be $\mu_{1,r}$, $\mu_{2,r}$, $\sigma_{1,r}$, $\sigma_{2,r}$, $\alpha_{1,r}$, $\alpha_{2,r}$ and L_r, $r = 1, \ldots, R$. Also assume random sequences $N_{n,r}$ independent of the on-off sources and independent of $N_{n,s}$ for all $r \neq s$, random variables U_r such that $P(U_r \geq 0) = 1$ and a sequence of positive constants d_n such that*

$$\frac{N_{n,r}}{d_n^2} \xrightarrow{D} U_r^2 \ \textit{and} \ \frac{N_{n,r} - U_r^2 d_n^2}{d_n} \xrightarrow{P} 0 \ \textit{as} \ n \to \infty.$$

Then for large T the aggregate cumulative traffic $\{W^*_{N_n}(Tt),\, t \geq 0\}$ behaves stochastically like

$$\left(\sum_{r=1}^{R} U_r^2 d_n^2 \frac{\mu_{1,r}}{\mu_{1,r} + \mu_{2,r}} \right) Tt + \sum_{r=1}^{R} \sigma_{lim,r} \sqrt{T^{2H_r} L_r(T)} U_r Z_{H_r}(t)$$

Note that one sequence d_n is specified for the convergence of the random sums in all the classes. By the independence of the classes the proof is an easy consequence of Theorem 5.2.2.

5.3 Role of Arbitrary Positive Random Variables

The following two examples illustrate the requirement for U in (5.1) above is not very restrictive and that (5.1) and (5.5) are not equivalent. Let U be an arbitrary continuous random variable with $\mathrm{P}(U \geq 0) = 1$, $N_n = \lfloor U^2 n^\gamma \rfloor$ (where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x) and $d_n^2 = n^\gamma$, $\gamma > 0$. Since

$$\frac{U^2 n^\gamma - 1}{n^\gamma} < \frac{N_n}{d_n^2} \leq \frac{U^2 n^\gamma}{n^\gamma}$$

and both upper and lower limits converge in distribution to U^2 as $n \to \infty$ it is clear that

$$\frac{N_n}{d_n^2} \xrightarrow{D} U^2 \text{ as } n \to \infty.$$

Thus any positive continuous random variable U can satisfy (5.1). A similar calculation can be performed for the condition in (5.5).

Now let $N_n = \lfloor U^2(n^\gamma + n^\delta) \rfloor$ where $\gamma > 0$, $\gamma/2 < \delta < \gamma$ and $d_n^2 = n^\gamma$. Then

$$\frac{U^2(n^\gamma + n^\delta) - 1}{n^\gamma} < \frac{N_n}{d_n^2} \leq \frac{U^2(n^\gamma + n^\delta)}{n^\gamma}$$

and so again

$$\frac{N_n}{d_n^2} \xrightarrow{D} U^2 \text{ as } n \to \infty.$$

However, for the condition in (5.5),

$$\frac{U^2 n^\delta - 1}{n^{\gamma/2}} < \frac{N_n - U^2 d_n^2}{d_n} = \frac{\lfloor U^2(n^\gamma + n^\delta) \rfloor - U^2 n^\gamma}{n^{\gamma/2}} \leq \frac{U^2 n^\delta}{n^{\gamma/2}}$$

and both bounds diverge as $n \to \infty$. So in this case (5.1) is satisfied while (5.5) is not.

For modelling, n is an artificial quantity whose role is to ensure N_n is large with high probability. For finite n, a natural question is how large should we make N_n? Simulation results in Norros and Pruthi [70] suggest that 50-100 sources may be sufficient for Gaussian limit results to be close in queuing applications. Then for fixed n_0, if we take $N_{n_0} = \lfloor a(U^2+b) \rfloor$ and $d_{n_0} = \sqrt{a}$ then n_0 is sufficiently large if $\mathrm{P}(N_{n_0} > 50)$ is close to 1. If not, a should be larger. Really the effect of n_0 is buried within a. But it would be totally arbitrary to replace a by n_0, n_0^p, or $n_0^p + n_0^q$, $p > 0$, $q > 0$. The second example above shows however we set a, set $d_n^2 = a$ to use the results of Corollary 5.2.3.

5.4 Stable Subordinators

Let $G_{\alpha,\theta}$ and $G_{\alpha,\theta}(x)$ denote the distribution and distribution function, respectively, of a strictly stable distribution with exponent α and parameter θ corresponding to the characteristic function

$$g_{\alpha,\theta}(s) = \exp\left(-|s|^\alpha \exp(-i\frac{\pi\alpha\theta}{2} \text{ sign } s)\right)$$

where $s \in \mathbf{R}$, $0 < \alpha \leq 2$, $|\theta| \leq \theta_\alpha = \min(1, -1 + 2/\alpha)$.

Suppose N_n is such that for some $\alpha \in (0, 2)$

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{N_n}{d_n^2} \leq x\right) = G_{\frac{\alpha}{2},1}(x),\ x \in \mathbf{R}.$$

Note that since $\alpha/2 < 1$ and $\theta = 1$ it follows that $G_{\frac{\alpha}{2},1}$ cannot be negative and, by (5.1), U^2 has a $G_{\frac{\alpha}{2},1}$ distribution. A random variable with the distribution $G_{\frac{\alpha}{2},1}$ is sometimes called a stable subordinator.

When U^2 is a stable subordinator the process $\{UZ_H(t),\, t \geq 0\}$ is called sub-Fractional Brownian Motion (sub-FBM) [83, p. 365]. It is a symmetric stable process, with characteristic exponent α, different from Linear Fractional Stable Motion or Lévy Stable Motion [83, p. 365]. Sub-FBM is also H-self-similar with stationary increments although they are not ergodic [13]. Its finite-dimensional characteristic functions are of the form

$$\mathrm{E}\left(\imath \sum_{n=1}^{N} \theta_n X(t_n)\right) = \exp\left[-\left(\frac{1}{2}\sum_{m,n=1}^{N} \theta_m \theta_n R(t_m, t_n)\right)^{\alpha/2}\right]$$

where R is the covariance function of $\{\sigma_{lim} Z_H(t)\}$, namely

$$R(t_m, t_n) = \frac{1}{2}\left(|t_m|^{2H} + |t_n|^{2H} - |t_m - t_n|^{2H}\right).$$

5.5 Weak Convergence

As in Taqqu *et. al* [88], the outer limit in (5.4) (as $T \to \infty$) can be replaced by weak convergence of probability measures in $C[0,\infty)$, denoted by $\Rightarrow$. This result is established by combining the analogous weak convergence result of [88] with a conditioning argument to show tightness. Let

$$Y_T(t) = \frac{\int_0^{Tt} G(u)du}{T^H L^{1/2}(T)} \text{ and } X_T(t) = \frac{U\int_0^{Tt} G(u)du}{T^H L^{1/2}(T)} = UY_T(t),$$

where $\{G(t), t \geq 0\}$ and U are as defined above.

Theorem 5.5.1.

$$X_T(t) \Rightarrow \sigma_{lim} U Z_H(t),\ t \geq 0 \ \text{as}\ T \to \infty.$$

Proof. Weak convergence is proven by showing tightness. By Karatzas and Shreve [42, p. 63] it must be established that

$$\lim_{\lambda\uparrow\infty} \sup_{T\geq 1} \mathrm{P}\left(X_T(0) > \lambda\right) = 0$$

(trivial), and for each $\eta > 0$ there exists a δ_0, $0 < \delta_0 < 1$ such that for $\delta < \delta_0$

$$\sup_{T\geq 1} \mathrm{P}\left(\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq R}} |X_T(t) - X_T(s)| \geq \varepsilon \right) \leq \eta, \quad \forall R > 0,\, \varepsilon > 0.$$

Without loss assume $0 < \eta < 1$. Let $f(u)$ denote the density of U and a be such that $\int_a^\infty f(u)\,du = \eta/2$. Then by tightness of Y_T [88, p. 15] there exists δ_0 such that for $\delta < \delta_0$

$$\sup_{T\geq 1} \mathrm{P}\left(\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq R}} |Y_T(t) - Y_T(s)| \geq \frac{\varepsilon}{a} \right) \leq \eta/2, \quad \forall R > 0,\, \varepsilon > 0.$$

For $0 < u < a$, by monotonicity in u

$$\begin{aligned}\sup_{T\geq 1} \mathrm{P}\left(\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq R}} |Y_T(t) - Y_T(s)| \geq \frac{\varepsilon}{u} \right) &\leq \sup_{T\geq 1} \mathrm{P}\left(\sup_{|s-t|<\delta} |Y_T(t) - Y_T(s)| \geq \frac{\varepsilon}{a} \right) \\ &\leq \eta/2.\end{aligned}$$

Thus for $\delta < \delta_0$

$$\begin{aligned} I &= \sup_{T\geq 1} \mathrm{P}\left(\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq R}} |X_T(t) - X_T(s)| \geq \varepsilon \right) \\ &= \sup_{T\geq 1} \int_0^\infty \mathrm{P}\left(\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq R}} |uY_T(t) - uY_T(s)| \geq \varepsilon \right) f(u)\,du \\ &\leq \frac{\eta}{2} \int_0^a f(u)\,du + \int_a^\infty f(u)\,du \\ &\leq \eta \quad \forall R > 0,\, \varepsilon > 0, \end{aligned}$$

$X_T(t)$ is tight and weak convergence follows from Theorem 5.2.2. □

5.6 Mixture Process for Modelling

Equation (5.6) suggests modelling the cumulative volume of traffic with the process

$$\tilde{W}^*(t) = \mu\tilde{U}^2 t + \sigma\tilde{U} Z_H(t),$$

where $Z_H(t)$ is standard FBM. A more convenient form for modelling is

$$W^*(t) = Ut + \sqrt{\gamma U} Z_H(t).$$

The corresponding increment process is

$$X(k) = W^*(k+1) - W^*(k) \stackrel{D}{=} U + \sqrt{\gamma U} Y_H(k)$$

where $Y_H(t)$ is standard Fractional Gaussian Noise (FGN).

The quantity γ plays a fundamental role. Indeed for any particular realization u of U, sample paths of $X(k)$ are Gaussian,

$$X(k) = u + \sqrt{\gamma u} Y_H(k).$$

and the ratio of variance to mean is

$$\frac{\gamma u}{u} = \gamma.$$

Thus we would expect that when the number of sources is the cause of variability *across* blocks of time (corresponding to different realizations of U), the ratio of variance to mean would remain constant.

The fundamental role of γ should not be too surprising. The underlying independent on-off processes have marginal Bernoulli(p) distributions whose ratio of variance to mean is the constant $p(1-p)/p = 1-p$. The sum S_n of n processes has a Binomial distribution whose ratio of variance to mean is the same constant $np(1-p)/np = 1-p$. For large n, the Gaussian approximation for S_n is $N(np, np(1-p))$, maintaining the same ratio.

A comment about modelling with γ is warranted. Namely, it is sensitive to the scale of measurement. Consider again the increment process

$$X(k) = W^*(k+1) - W^*(k) \stackrel{D}{=} U^2 + \gamma U Y_H(k)$$

Now consider the process $X^{(m)}(k)$ for modelling with time units m times longer (e.g. tenths of seconds vs. seconds). Then

$$\begin{aligned} X^{(m)}(k) &= W^*(mk+m) - W^*(mk) \\ &= mU + \sqrt{\gamma U}\left(Z_H(mk+m) - Z_H(mk)\right) \\ &\stackrel{D}{=} mU + \sqrt{\gamma U} m^H \left(Z_H(k+1) - Z_H(k)\right) \\ &\stackrel{D}{=} mU + m^H \sqrt{\gamma U} Y_H(k) \end{aligned}$$

where the second-from-last line follows from the self-similarity of FBM. Then the ratio of variance to mean is $m^{2H-1}\gamma$. Note this is constant only in the i.i.d. case ($H = 0.5$) and goes to ∞ as $m \to \infty$ in the long-range dependent case ($0.5 < H < 1$). While this causes no problems for the i.i.d. case, for the long-range dependent case it does require consistency in the measurement units for different realizations of U. But this would occur naturally- one is not likely to consider a model in, say, seconds and bytes for one realization of U then switch to milliseconds or bits for another realization as part of the same model.

If $W^*(t)$ is modelling the cumulative amount of traffic (e.g. packets, bytes) then γ plays a role in the index of dispersion of counts up to time t, $IDC(t)$. By definition, for each realization u of U

$$IDC(t) = \frac{\mathrm{Var}(W^*(t))}{\mathrm{E}(W^*(t))} = \frac{\gamma u t^{2H}}{ut} = \gamma t^{2H-1}.$$

So a log-log plot of IDC(t) vs. t would have slope $2H-1$ and intercept γ. If $H = 0.5$ then IDC(t) should be approximately a horizontal line.

5.7 Example

Here is a simple example to show how the mixture model might be used. Assume R independent classes of traffic on a link of capacity C. For class i model the cumulative traffic as

$$W_i^*(t) = U_i t + \sqrt{\gamma_i U_i} B_{i,H_i}(t)$$

and so the increment process is

$$X_i(k) = U_i + \sqrt{\gamma_i U_i} Y_{i,H_i}(k)$$

where $B_{i,H}$ is standard FBM with Hurst parameter H_i, and Y_{i,H_i} is standard FGN with Hurst parameter H_i, $i = 1, \ldots, R$. Each class is imagined to have a reserved portion C_i of the bandwidth, perhaps due some reservation technique, *that depends* on U_i. Clearly, for the available (unreserved) bandwidth, C_{unused}

$$C_{unused} = C - \sum_{i=1}^{R} C_i \geq 0.$$

Across blocks of time variability in the number of sources can change the traffic demands of each class through the random variables U_i. This will change the unreserved bandwidth, which might be provided as a best-effort service. Of course, total requested bandwidth could exceed capacity, and then the unreserved bandwidth is 0. Here unreserved bandwidth is considered in which each class is allocated bandwidth based on its mean rate and some desired utilization (as opposed to its effective bandwidth or some other measure). Specifically this means that

$$C_i = \frac{1}{\rho_i} U_i, \quad 0 < \rho_i < 1$$

so that the utilization in class i is ρ_i. Then for the unreserved bandwidth

$$C_{unused} = C - \sum_{i=1}^{R} \frac{1}{\rho_i} U_i$$

and the probability of insufficient bandwidth on the link is

$$\mathrm{P}\left(\sum_{i=1}^{R} \frac{1}{\rho_i} U_i \right) > C.$$

The distribution of the sum depends on the distribution U_i which is in the same family as the number of sources $N_{i,n}$ in that class. So, for example, if the number of sources were well-modelled with Gaussian distributions then the probability of insufficient bandwidth would be particularly easy to find.

5.8 Simulation Results

To demonstrate how a random number of sources can affect the limiting processes, i.i.d. on-off processes were simulated under a variety of assumptions for the number of sources. In all of the simulations the on-off processes had identical properties. The on and off times had independent Pareto distributions with characteristic exponent $\alpha = 1.4$, so $H = 0.8$. The mean on and off times were 1 and 1.5 seconds respectively, so $\mu_1/(\mu_1 + \mu_2) = 0.4$. The simulation was actually performed at the level of hundredths of seconds, but then aggregated by a factor of 100 to yield seconds. This yielded data at the 1 second level while minimizing granularity effects from changing the continuous Pareto data to integers. Prior to aggregation, the on and off states sent data at rate 1 and 0 per hundredths of a second, respectively. Each run had 10000 observations, so about 3 hours worth.

The number of simulated sources, N_n, in each of the four cases considered is summarized in Table 1: no randomness, exponential, Uniform, and Gaussian. Table 1 also shows associated values of d_n^2 and U which are used in the limiting process. The non-random case was 300 sources. In the three cases with randomness, the distributions were chosen specially to have mean 300 (to match with the non-random case) and standard deviation 15. For each of these four assumptions, 1000 superpositions were simulated. To summarize, for each of four assumptions, 1000 sample paths of 3 hours each were simulated.

The top row of Figure 5.1 shows two typical superpositions for 300 sources (i.e. no randomness). The mean is very close to the theoretical value $100 * 300(0.4) = 12000$

	N_n	d_n^2	U
No randomness	300	n/a	n/a
Exponential	$\lfloor 15V + 285 \rfloor$	15	$V + 19$
Uniform	$\lfloor Y \rfloor$	1	Y
Normal	$\lfloor 15Z + 300 \rfloor$	15	$Z + 20$

Table 5.1: Assumed distribution of the number of sources. V$\sim$exponential(1), Y$\sim$Uniform(274,326), Z$\sim$N(0,1)

in both. (Note that 100 arises because of the time aggregation by a factor of 100.) The bottom row shows two superpositions under the exponential assumption for the number of sources. The differences in the mean rates are particularly noticeable.

By equation (5.6) the means are random across different realizations of N_n and are distributed like U. Figure 5.2 shows the 1000 sample means from each of the four cases simulated while Figure 5.3 shows boxplots for the sample means. The effect of randomness in the number of sources is clear.

For each superposition $X_i = (X_{i,1}, \ldots, X_{i,10000})$, $i = 1, \ldots, 1000$, there is a different value of γ that can be estimated. Figure 5.4 shows the 1000 results for each case using the simplistic estimator for *known* Hurst parameter $H = 0.8$

$$\hat{\gamma} = \frac{S^2_{10000,0.8}}{\hat{\mu}}$$

where $\hat{\mu}$ is the sample mean and $S^2_{n,H} = \frac{n-1}{n-n^{2H-1}} S^2$ is the unbiased sample variance S^2 for Fractional Gaussian Noise of length n and Hurst parameter H [7, p. 9]. Unlike the sample mean, even after unbiasing, the sample variance is not a particularly good estimator for the variance, and isn't even normally distributed for $H > .75$. In Chapter 6 better techniques for estimating γ will be discussed. Nevertheless, the boxplots in Figure 5.5 show that despite randomness in the number of sources, γ is quite stable.

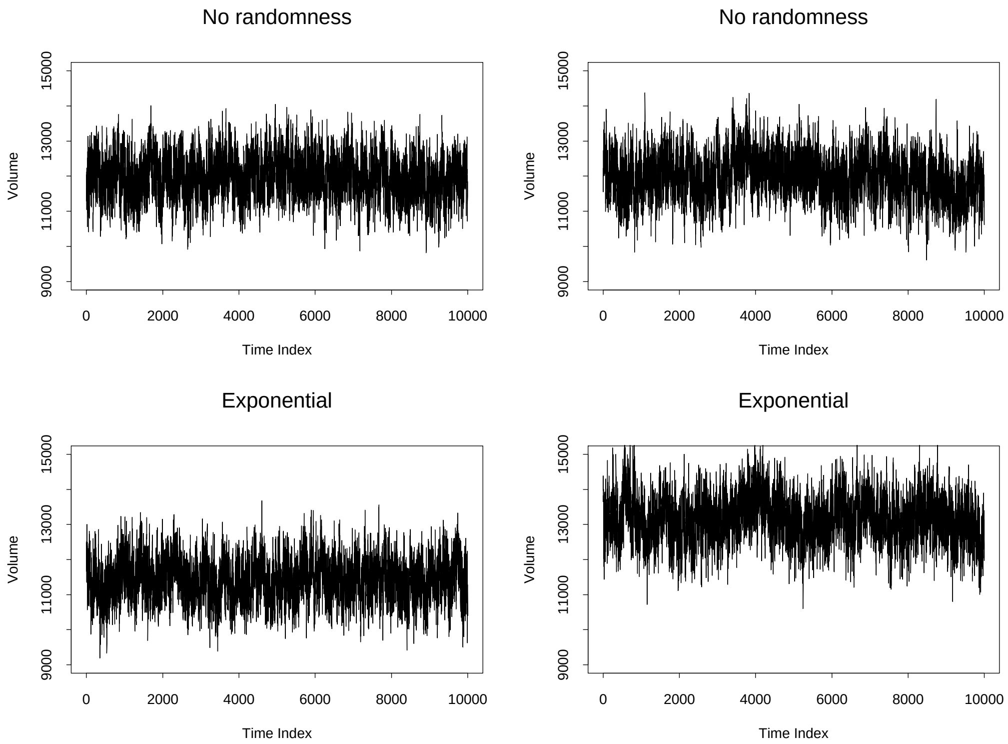

Figure 5.1: Superpositions of on-off sources

Two sample paths each when the number of sources is non-random (top row) and exponentially distributed (bottom row).

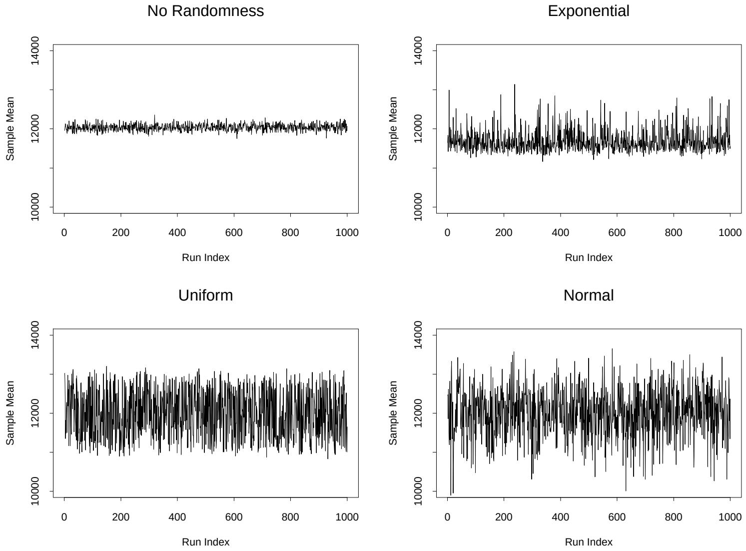

Figure 5.2: 1000 sample means from the superposition of on-off sources

The number of on-off sources is not random (top-left), Exponentially distributed (top-right), Gaussian (bottom-left), and Uniform (bottom-right).

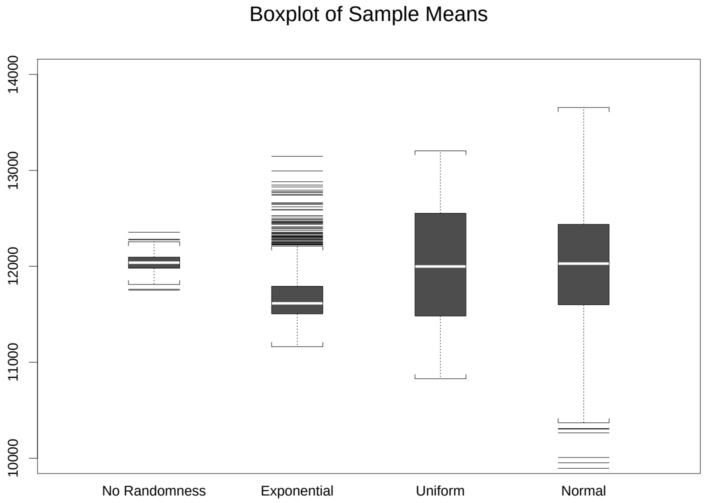

Figure 5.3: Boxplots for the 1000 sample means for the four examples simulated.

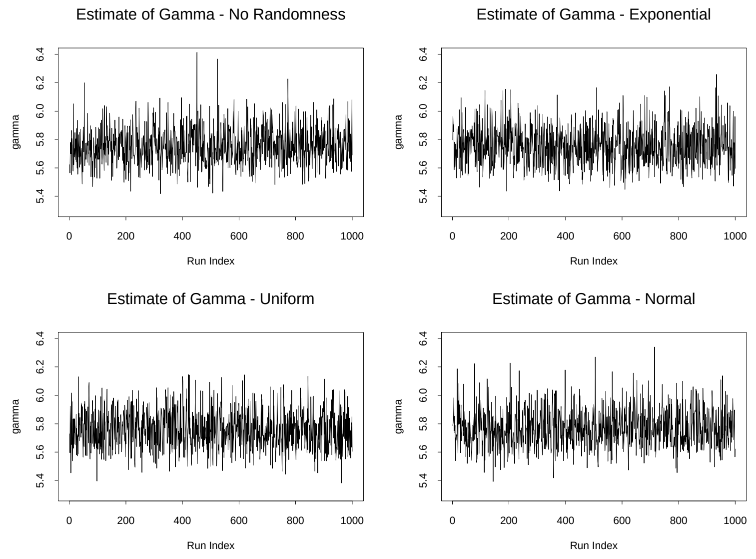

Figure 5.4: 1000 Estimates of $\gamma = \frac{\text{unbiased sample variance}}{\text{sample mean}}$

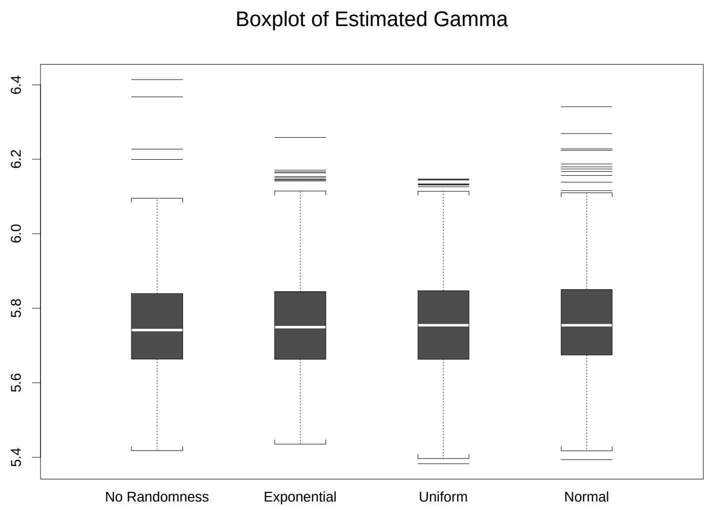

Figure 5.5: Boxplots for estimates of γ in the four examples simulated. Randomness in the number of sources appears to have little effect.

Chapter 6

Fractional Gaussian Noise Scale Mixture Models

The results of Chapter 5 concern the limiting behaviour of the random superposition of alternating on-off processes. While one could model individual on-off sources, the real benefit of those results is in the theoretical justification for modelling the aggregate cumulative traffic with

$$W^*(t) = Ut + \sqrt{\gamma U} Z_H(t),\ t \geq 0 \tag{6.1}$$

where $\gamma > 0$, U is a non-negative random variable capturing the additional randomness from the number of sources, and $Z_H(t)$ is FBM with Hurst parameter H [68]. Then the traffic in unit time increments can be modelled with

$$X^*(k) \stackrel{D}{=} U + \sqrt{\gamma U} Y_H(k),\ k \in \mathcal{Z}^+ \tag{6.2}$$

where $Y_H(k) = Z_H(k+1) - Z_H(k)$ is Fractional Gaussian Noise (FGN) with Hurst parameter H [56]. The parameter γ plays a fundamental role. It is the ratio of the variance to the mean of $Y_H(k)$ and does not depend on the value or distribution of U. This chapter proposes statistical techniques to estimate γ.

Two models based on (6.1) and (6.2), combining Fractional Gaussian Noise (FGN) with random scaling are considered. What distinguishes them is the number of Fractional Gaussian Noise Processes within the model. Each motivates its own estimators for γ.

Section 6.1 presents the "Many-Process Scale Mixture" model in which there are a collection of FGN processes. Associated with each is another random variable capturing randomness due to the number of sources. In Section 6.2, the "One-Process Scale Mixture" model is presented. The key difference is that there is one Fractional Gaussian Noise process and random variance that changes with time. Section 6.3 shows simulation results using γ estimators, intended for the "Many-Process" model, applied to the superposition of a random number of on-off sources with Pareto-distributed on and off times.

6.1 Many-Process Scale Mixture Model

6.1.1 The Model

Imagine R time periods (e.g. R afternoons), each described by the interval $[\tau_i, \tau_i + \Delta)$, $i = 1, \ldots, R$ with starting time τ_i and length $\Delta > 0$. The intervals are assumed to be non-overlapping. The number of sources within each interval may be well-modelled as a superposition of a large, fixed number of identically distributed on-off sources, independent of all other time periods. Between one time period and the next, the number of sources are independent and random according to some probability distribution. Then, for each time period, the cumulative and incremental traffic at time $\tau_i + t$ could be modelled as

$$W_i^*(t+\tau_i) = U_i(t+\tau_i) + \sqrt{\gamma U_i}\left(Z_{H,i}(t+\tau_i) - Z_{H,i}(\tau_i)\right),\ t \in [0,\Delta), i = 1,\ldots,R \tag{6.3}$$

and

$$X_i^*(k) = U_i + \sqrt{\gamma U_i} Y_{H,i}(k),\ k \in \{0, 1\ldots, \lfloor\Delta\rfloor - 1\}, i = 1,\ldots,R \tag{6.4}$$

respectively, where $\{U_i,\ i = 1,\ldots,R\}$ is a sequence of independent, identically distributed (i.i.d.) non-negative random variables, $\{Z_{H,i}(t),\ i = 1,\ldots,R\}$ and $\{Y_{H,i}(k),\ i = 1,\ldots,R\}$ are families of independent Fractional Brownian Motion and Fractional Gaussian Noise, respectively, assumed independent of the sequence $\{U_i,\ i = 1,\ldots,R\}$.

When the time periods are assumed independent, these processes follow simply from the random superposition result of Chapter 5. In this case it will be preferable to think of the processes $W_i^*(t)$ and $X_i^*(k)$ as mutually independent copies (over $i = 1,\ldots,R$) of

$$W_i^*(t) = U_i t + \sqrt{\gamma U_i} Z_{H,i}(t),\ t \in [0,\Delta), i = 1,\ldots,R \tag{6.5}$$

and

$$X_i^*(k) = U_i + \sqrt{\gamma U_i} Y_{H,i}(k),\ k \in \mathcal{Z}^+, i = 1,\ldots,R \tag{6.6}$$

where the role of τ_i is supressed. This model is referred to as the “Many-Process Scale-Mixture Model”.

6.1.2 Estimation of γ

Note that the variance in Equation (6.3) includes γ. For parameter estimation it is easier to start with the modelling equation

$$W_i(t) = \frac{1}{\gamma}U_i t + \sqrt{U_i}Z_{H,i}(t).$$

For modelling, $W_i(t)$ and $W_i^*(t)$ are equivalent (but not equal). For a particular realization u_i of U_i $(i = 1, \ldots, R)$ it will be convenient to speak of

$$W_{i|u_i}(t) = \frac{1}{\gamma}u_i t + \sqrt{u_i}Z_{H,i}(t),$$

$$X_{i|u_i}(k) = \frac{1}{\gamma}u_i + \sqrt{u_i}Y_{H,i}(k). \tag{6.7}$$

and the corresponding estimators $\bar{\mu}_{n,i|\frac{u_i}{\gamma}}$, $\hat{\sigma}^2_{n,i|u_i}$ and $S^2_{n,\hat{H}_{n,i},i|u_i}$.

The problem of estimating γ is essentially that of estimating the slope of the line $y = \gamma x$ from a collection of ordered pairs $(x_i, y_i) = (\bar{\mu}_{n,i}, \hat{\sigma}^2_{n,i})$, $i = 1, \ldots, R$ where $\bar{\mu}_{n,i}$ and $\hat{\sigma}^2_{n,i}$ are the sample mean and Whittle estimate of variance, respectively, for realizations of $X_i(k)$. Alternatively one might use $(x_i, y_i) = (\bar{\mu}_{n,i}, S^2_{n,\hat{H}_{n,i},i})$, $i = 1, \ldots, R$, where $S^2_{n,\hat{H}_{n,i},i}$ is the sample variance of a realization of $X_i(k)$ unbiased for Fractional Gaussian Noise with estimated Hurst parameter $\hat{H}_{n,i}$.

One could try to adapt techniques for measurement-error models, but the random scaling means generally neither x_i nor y_i are Gaussian. Instead, possible estimators are

$$\hat{\gamma}^{(W)}_{n,R} = \frac{\sum_{i=1}^{R}\hat{\sigma}^2_{n,i}}{\sum_{i=1}^{R}\bar{\mu}_{n,i}} \text{ and } \hat{\gamma}^{(S)}_{n,R} = \frac{\sum_{i=1}^{R}S^2_{n,\hat{H}_i,i}}{\sum_{i=1}^{R}\bar{\mu}_{n,i}},$$

where W and S signify use of either the Whittle or sample variance-based estimator of variance, respectively. These are recognizable as method-of-moment estimators.

Note that there are two quantities of interest, the *length*, of the processes, n, and the number of processes, R. The following theorem shows the consistency (as $n \to \infty$) of the proposed estimators for any finite R.

Theorem 6.1.1. *1) If $H_i = H \in (0.5, 1)$ for all $1, \ldots, R$,*

$$\operatorname*{plim}_{n\to\infty} \hat{\gamma}^{(W)}_{n,R} = \gamma \text{ and} \tag{6.8}$$

2) if $H_i = H \in [0.5, 1)$, for all $1, \ldots, R$,

$$\operatorname*{plim}_{n\to\infty} \hat{\gamma}^{(S)}_{n,R} = \gamma. \tag{6.9}$$

Before proving this theorem the following results are needed:

Lemma 6.1.2.

$$\left(\hat{\sigma}^2_{n,1|u_1}, \ldots, \hat{\sigma}^2_{n,R|u_R}\right) \xrightarrow{P} (u_1, \ldots, u_R) \text{ as } n \to \infty \text{ and} \tag{6.10}$$

$$\left(S^2_{n,\hat{H}_{n,1},1|u_1}, \ldots, S^2_{n,\hat{H}_{n,R},R|u_R}\right) \xrightarrow{P} (u_1, \ldots, u_R) \text{ as } n \to \infty. \tag{6.11}$$

Proof. This follows from consistency and for any $\varepsilon > 0$

$$\begin{aligned}
\mathrm{P}\big(\sup_{1\le i\le R} |X_{n,i} - u_i| > \varepsilon\big) &= \mathrm{P}\left(\bigcup_{i=1}^{R} |X_{n,i} - u_i| > \varepsilon\right) \\
&\le \sum_{i=1}^{R} \mathrm{P}(|X_{n,i} - u_i| > \varepsilon) \\
&\to 0 \text{ as } n \to \infty
\end{aligned}$$

where $\{X_{n,i}\} = \{\hat{\sigma}^2_{n,i|u_i}\}$ or $\{X_{n,i}\} = \{S^2_{n,\hat{H}_{n,i},i|u_i}\}$ to prove (6.10) or (6.11), respectively. $\square$

Proof. First, (6.8) is proven with a conditioning argument. For any $\varepsilon > 0$,

$$\begin{aligned}
T &= \lim_{n\to\infty} \mathrm{P}\left(\left|\hat{\gamma}^{(W)}_{n,R} - \gamma\right| > \varepsilon\right) \\
&= \lim_{n\to\infty} \int_0^\infty \cdots \int_0^\infty \mathrm{P}\left(\left|\hat{\gamma}^{(W)}_{n,R} - \gamma\right| > \varepsilon | (U_1, \ldots, U_R) = (u_1, \ldots, u_R)\right) \times \\
&\qquad\qquad f_{U_1,\ldots,U_R}(u_1, \ldots, u_r) du_1 \cdots du_R \\
&= \lim_{n\to\infty} \int_0^\infty \cdots \int_0^\infty \mathrm{P}\left(\left|\frac{\frac{1}{R}\sum_{i=1}^R \hat{\sigma}^2_{n,i|u_i}}{\frac{1}{R}\sum_{i=1}^R \bar{\mu}_{n,i|\frac{u_i}{\gamma}}} - \gamma\right| > \varepsilon\right) f_{U_1,\ldots,U_R}(u_1, \ldots, u_r) du_1 \cdots du_R
\end{aligned}$$

Therefore,

$$\begin{aligned}
T &= \int_0^\infty \cdots \int_0^\infty \lim_{n\to\infty} \mathrm{P}\left(\left|\frac{\frac{1}{R}\sum_{i=1}^R \hat{\sigma}^2_{n,i|u_i}}{\frac{1}{R}\sum_{i=1}^R \bar{\mu}_{n,i|\frac{u_i}{\gamma}}} - \gamma\right| > \varepsilon\right) f_{U_1,\ldots,U_R}(u_1, \ldots, u_r) du_1 \cdots du_R \\
&= \int_0^\infty \cdots \int_0^\infty \mathrm{P}\left(\left|\frac{\sum_{i=1}^R u_i}{\frac{1}{\gamma}\sum_{i=1}^R u_i} - \gamma\right| > \varepsilon\right) f_{U_1,\ldots,U_R}(u_1, \ldots, u_r) du_1 \cdots du_R \\
&= 0.
\end{aligned}$$

The fourth line follows from the Dominated Convergence Theorem since the integrand is dominated by the density. The fifth line follows from the consistency of both the Whittle estimator of variance and the sample mean.

The proof of (6.9) is similar but replaces consistency of the Whittle estimator of variance with consistency of the 'unbiased' sample variance. □

Theorem 6.1.3. *1) If $H_i = H \in (0.5, 1)$ for all $1, \ldots, R$,*

$$n^{1-H}(\hat{\gamma}^{(W)}_{n,R} - \gamma) \xrightarrow{D} \gamma^2 \left(\sum_{i=1}^R U_i\right)^{-1/2} Z \text{ and} \tag{6.12}$$

2) if $H_i = H \in [0.5, 1)$, $H \neq 0.75$, for all $1, \ldots, R$,

$$n^{1-H}\left(\hat{\gamma}^{(S)}_{n,R} - \gamma\right) \xrightarrow{D} \gamma^2 \left(\sum_{i=1}^R U_i\right)^{-1/2} Z, \tag{6.13}$$

where Z is a $N(0, 1)$ random variable independent of $(U_1, U_2, \ldots, U_R)$.

Proof. To prove (6.12) let δ_i and η_i, $i = 1, \ldots, R$, be i.i.d sequences of $N(0,1)$ random variables. Condition on $(U_1, \ldots, U_R) = (u_1, \ldots, u_r)$. Thus

$$
\begin{aligned}
A_{n,(u_1,\ldots,u_R)} &= n^{1-H}\left(\frac{\frac{1}{R}\sum_{i=1}^{R}\hat{\sigma}^2_{n,i|u_i}}{\frac{1}{R}\sum_{i=1}^{R}\bar{\mu}_{n,i|\frac{u_i}{\gamma}}} - \gamma\right)\\
&= \frac{\left[\frac{n^{1-H}}{\sqrt{n}}\sqrt{n}\sum_{i=1}^{R}(\hat{\sigma}^2_{n,i|u_i} - u_i)\right] + \left[-\gamma n^{1-H}\sum_{i=1}^{R}(\bar{\mu}_{n,i|\frac{u_i}{\gamma}} - \frac{u_i}{\gamma})\right]}{\sum_{i=1}^{R}\bar{\mu}_{n,i|\frac{u_i}{\gamma}}}\\
&= \frac{V_1 + V_2}{\sum_{i=1}^{R}\bar{\mu}_{n,i|\frac{u_i}{\gamma}}}.
\end{aligned}
$$

We know $\sqrt{n}(\hat{\sigma}^2_{n,i} - u_i) \xrightarrow{D} X_i \sim \sqrt{2}u_i\delta_i$ and $n^{1-H}(\bar{\mu}_{n,i} - \frac{u_i}{\gamma}) \xrightarrow{D} \sqrt{u_i}\eta_i$ as $n \to \infty$ by the corresponding results for convergence of the Whittle estimator of variance and the sample mean of Fractional Gaussian Noise. This means $V_1 \xrightarrow{P} 0$ and $V_2 \xrightarrow{D} \gamma\sum_{i=1}^{n}\sqrt{u_i}\eta_1$. So, by Slutsky's Theorem and consistency of the sample mean

$$
\begin{aligned}
A_{n,(u_1,\ldots,u_R)} &\xrightarrow{D} \frac{\gamma\sum_{i=1}^{R}\left(\sqrt{u_i}\eta_i\right)}{\sum_{i=1}^{R}\frac{u_i}{\gamma}} \text{ as } n \to \infty\\
&\stackrel{D}{=} \gamma^2\left(\sum_{i=1}^{R}u_i\right)^{-1/2}\eta_1. \qquad (6.14)
\end{aligned}
$$

For the unconditional result we have, for any $x \in \mathbf{R}$

$$
\begin{aligned}
&\mathrm{P}\left(n^{1-H}(\hat{\gamma}^{(W)}_{n,R} - \gamma) \le x\right)\\
&= \lim_{n\to\infty}\int_0^\infty \cdots \int_0^\infty \mathrm{P}\left(A_{n,(u_1,\ldots,u_R)} \le x\right) f_{U_1,\ldots,U_R}(u_1,\ldots,u_r)du_1\cdots du_R\\
&= \int_0^\infty \cdots \int_0^\infty \lim_{n\to\infty}\mathrm{P}\left(A_{n,(u_1,\ldots,u_R)} \le x\right) f_{U_1,\ldots,U_R}(u_1,\ldots,u_r)du_1\cdots du_R\\
&= \int_0^\infty \cdots \int_0^\infty \mathrm{P}\left(\gamma^2\left(\sum_{i=1}^{R}u_i\right)^{-1/2}\eta_1 \le x\right) f_{U_1,\ldots,U_R}(u_1,\ldots,u_r)du_1\cdots du_R\\
&= \mathrm{P}\left(\gamma^2\left(\sum_{i=1}^{R}U_i\right)^{-1/2}\eta_1 \le x\right)
\end{aligned}
$$

where the second line follows from the Dominated Convergence Theorem since the integrand is dominated by the density, and the third line follows from the convergence in (6.14). The fourth line uses the independence of η_1 and $(U_1, \ldots, U_R)$ which is a consequence of the independence between the FGN sequences $\{Y_{H,i}(k),\ i = 1, \ldots, R\}$ and $(U_1, \ldots, U_R)$. This proves the desired result with $Z = \eta_1$.

To prove (6.13), first note that to obtain non-degenerate distributions in the limit $(n-1)S_n^2$ must be scaled by $n^{-1/2}$ for $H \in (0.5, 0.75)$, and n^{1-2H} for $H \in (0.75, 1)$ [7, p. 156]. Again condition on $(U_1, \ldots, U_R) = (u_1, \ldots, u_r)$. Then

$$\begin{aligned} B_{n,(u_1,\ldots,u_R)} &= n^{1-H}\left(\frac{\sum_{i=1}^R S^2_{n,\hat{H}_{n,i},i|u_i}}{\sum_{i=1}^R \bar{\mu}_{n,i|\frac{u_i}{\gamma}}} - \gamma\right) \\ &= n^{1-H}\frac{V_1 + V_2 + V_3}{\sum_{i=1}^R \bar{\mu}_{n,i|\frac{u_i}{\gamma}}}, \end{aligned}$$

where

$$\begin{aligned} V_1 &= \sum_{i=1}^R \frac{n-1}{n - n^{2\hat{H}_{n,i}-1}}(S_n^2 - u_i), \\ V_2 &= \sum_{i=1}^R u_i\left(\frac{n-1}{n - n^{2\hat{H}_{n,i}-1}} - 1\right) \quad \text{and} \\ V_3 &= -\gamma\sum_{i=1}^R \left(\bar{\mu}_{n,i|\frac{u_i}{\gamma}} - \frac{u_i}{\gamma}\right). \end{aligned}$$

Now, if $H \in (0.5, 0.75)$,

$$n^{1-H}V_1 = \sum_{i=1}^R \frac{n^{1-H}\sqrt{n}}{n - n^{2\hat{H}_{n,i}-1}}\left(\frac{n-1}{\sqrt{n}}(S_n^2 - 1)\right) \xrightarrow{P} 0 \text{ as } n \to \infty,$$

while if $H \in (0.75, 1)$,

$$n^{1-H}V_1 = \sum_{i=1}^R \frac{n^H}{n - n^{2\hat{H}_{n,i}-1}}\left(\frac{n-1}{n^{2H-1}}(S_n^2 - 1)\right) \xrightarrow{P} 0 \text{ as } n \to \infty.$$

For the second term,

$$n^{1-H}V_2 = \sum_{i=1}^{R} u_i n^{1-H} \frac{n^{2\hat{H}_{n,i}-1}}{n - n^{2\hat{H}_{n,i}-1}} \xrightarrow{P} 0 \text{ as } n \to \infty.$$

For the third term,

$$n^{1-H}V_3 = -\gamma \sum_{i=1}^{R} n^{1-H} \left(\bar{\mu}_{n,i|\frac{u_i}{\gamma}} - \frac{u_i}{\gamma} \right) \xrightarrow{D} \gamma \sum_{i=1}^{R} \sqrt{u_i} Y_i \stackrel{D}{=} \gamma \left(\sum_{i=1}^{R} u_i \right)^{1/2} \eta_1.$$

and so by Slutsky's theorem

$$B_{n,(u_1,\ldots,u_R)} \xrightarrow{D} \frac{\gamma \left(\sum_{i=1}^{R} u_i \right)^{1/2} Z_2}{\sum_{i=1}^{R} \frac{u_i}{\gamma}} = \gamma^2 \left(\sum_{i=1}^{R} u_i \right)^{-1/2} \eta_1.$$

The result follows from unconditioning, which proceeds exactly as for $A_{n,(u_1,\ldots,u_R)}$ above. $\square$

6.2 One-Process Model

6.2.1 The Model

Alternatively, consider a possible model for the cumulative and incremental traffic for one time interval (e.g. one afternoon). Let $K_0 = 0 < K_1 < K_2 < \cdots$ be an increasing sequence of known constants, partitioning the time axis, and let $\{U_i, i = 0, 1, \ldots\}$ be an identically distributed sequence of non-negative random variables. Define the continuous-time process $\{U(t) = U_i,\ t \in [K_i, K_{i+1})\}$ so the sample paths of $U(t)$ are functions on $[0, \infty)$ that are right-continuous with left-hand limits. Let the cumulative and incremental traffic be modelled as

$$W^*(t) = U(t)\, t + \sqrt{\gamma U(t)} Z_H(t),\ t \geq 0$$

and $X^*(k) =$

$$\begin{cases} U(K_i) + \sqrt{\gamma U(K_i)} Y_H(k), & K_i \leq k < K_{i+1} - 1 \\ \Big(U(K_{i+1})(k+1) - U(K_i)k\Big) + \Big(\sqrt{\gamma U(K_{i+1})} B_H(k+1) - \sqrt{\gamma U(K_i)} B_H(k)\Big), & \\ & K_i \leq k < K_{i+1} \leq k+1 < K_{i+2}, \end{cases}$$

so that between consecutive epochs the variability in the increment process is Fractional Gaussian Noise. In this model there is one underlying Fractional Brownian Motion or Fractional Gaussian Noise, thus the name "Single-Process Scale Mixture" model is appropriate. Clearly this is a special case of the Many-Process model with $R = 1$, $K_1 = \infty$ and $U_i \equiv U$ where U is some non-negative random variable. The desire to estimate γ remains the same. Estimators of γ are left for future work.

6.3 Simulation Results of $\hat{\gamma}$ for Random Superpositions of On-Off Sources

In this section the proposed estimators $\hat{\gamma}^{(W)}_{n,R}$ and $\hat{\gamma}^{(S)}_{n,R}$ are applied to superpositions of simulated on-off data with Pareto distributed on and off times. To generate the data, i.i.d stationary on-off processes $\{W^{(m)}(k), k = 1, \ldots, 10^6, m = 1, \ldots, N\}$ were simulated and aggregated to obtain

$$W^{(m)}(k;T) = \sum_{i=T(k-1)+1}^{Tk} W^{(m)}(i), \text{ and}$$

$$X(k;T,N) = \sum_{m=1}^{N} W^{(m)}(k;T) = \sum_{m=1}^{N} \sum_{i=T(k-1)+1}^{Tk} W^{(m)}(i) = \sum_{i=T(k-1)+1}^{Tk} \sum_{m=1}^{m=N} W^{(m)}(i)$$

where $W^{(m)}(k;T)$ are the time-aggregated processes with block size T, and $X(k;T,N)$ is the random superposition of N time-aggregated processes. The processes $X(k;T,N)$ were simulated under a variety of assumptions for the distribution of the *random* number of sources, N. For the data in Rolls and Takahara [80], $T = 100$ was chosen, yielding $\{X(k;100,N), k = 1\ldots,10^4$.

For each of four assumptions for N, 1000 independent aggregate processes were simulated with corresponding independent values N_i, $i = 1\ldots,1000$, yielding $\{X(k;100,N_i), k = 1\ldots,10^4, i = 1\ldots,1000\}$. It is these processes which we imagine modelling as a collection of Fractional Gaussian Noises as in (6.7).

In all of the simulations the on-off processes have identical properties. The on times have independent Pareto distributions with characteristic exponent $\alpha_1 = 1.4$, and mean $\mu_1 = 1$, while the off times have independent Pareto distributions with characteristic exponent $\alpha_2 = 1.4$, and mean $\mu_2 = 1.5$. The distribution functions and densities are

$$F_j(x) = 1 - \left(\frac{b_j}{x}\right)^{\alpha_j}, \; x \geq b_j \text{ and } f_j(x) = \frac{\alpha_j}{b_j}^{\alpha_j} x^{\alpha_j+1}, \; j = 1,2$$

respectively, where $b_j = \frac{\alpha_j - 1}{\alpha_j}\mu$.

The distribution of the number of simulated processes, N, in each of the four cases considered is summarized in Table 6.1: no randomness, exponential, Uniform, and Gaussian. The non-random case has 300 sources. In the three cases with randomness, the distributions were chosen specially to have mean 300 (to match with the non-random case) and standard deviation 15.

	N
No randomness	300
Normal	$\lfloor 15Z + 300 \rfloor$
Exponential	$\lfloor 15V + 285 \rfloor$
Uniform	$\lfloor Y \rfloor$

Table 6.1: Distribution for the number of sources. V$\sim$ exponential(1), Y$\sim$ Uniform (274,326), Z$\sim$N(0,1)

6.3.1 Calculating the Mean and Variance of Superpositions

For each $i = 1, \ldots, 1000$ the mean and variance of $X_i(k; T, N)$ and the Hurst parameter for its Fractional Gaussian Noise approximation can be considered. The exact Hurst parameter is simply $H = (3 - \min(\alpha_1, \alpha_2))/2 = 0.8$. By properties of the Binomial distribution, $\mathrm{E}[X_i(k; T, N)|N = u] = \frac{\mu_1}{\mu_1+\mu_2} T u$, and so $\mathrm{E}[X_i(k; T, N)] = \frac{\mu_1}{\mu_1+\mu_2} T \mathrm{E}[N_i]$ if $\mathrm{E}[N_i]$ exists (although the existence of $\mathrm{E}[N_i]$ is not needed).

Obtaining the variance of the processes $X_i(k; T, N)$ is more difficult. We have $\mathrm{Var}[X_i(k; T, N)|N = u] = u\mathrm{Var}[X_i(k; T)]$, but how does one calculate $\mathrm{Var}[X_i(k; T)]$. Asymptotic results [88] exist for $T \to \infty$ which suggest

$$\mathrm{Var}[X_i(k; T)] \approx T^{2H} \mathrm{Var}[W^{(i)}(k)]$$

but this approximation appears unsatisfactory for the finite values of T considered here. For small block sizes, obtaining numerical results that agree very well with simulation has proven successful. As in Chapter 2, let $r(v) = \mathrm{E}W(v)W(0) - (\mathrm{W}(0))^2$ denote the autocovariance where $\{W(k), k = 1, \ldots, 10^6\}$ is a single unaggregated on-off process. Then the variance of a single *time-aggregated* process is given by

$$\begin{aligned}
\mathrm{Var}\left[W(k;T)\right] &= \mathrm{E}\left[\sum_{i=(k-1)T+1}^{kT} W(i)\right]^2 - \left(\mathrm{E}\left[\sum_{i=(k-1)T+1}^{kT} W(i)\right]\right)^2 \\
&= T\left(\mathrm{E}[W^2(0)] - (\mathrm{E}[W(0)])^2\right) + 2\sum_{\substack{i,j=1\\ i<j}}^{T}\left(\mathrm{E}[W(i)W(j)] - \left(\mathrm{E}[W(0)]^2\right)\right) \\
&= Tr(0) + 2\sum_{i=1}^{T-1}(T-i)r(i) \qquad (6.15)
\end{aligned}$$

Using results from renewal theory, the Laplace transform $r^e(s)$ of $r(v)$ (defined to be $r^e(s) = \int_0^\infty e^{-st} r(t)\,dt$) is known [88] to be

$$r^e(s) = \frac{\mu_1\mu_2}{(\mu_1+\mu_2)^2 s} - \frac{(1-f_1^e(s))\,(1-f_2^e(s))}{(\mu_1+\mu_2)s^2\,(1-f_1^e(s)f_2^e(s))} \qquad (6.16)$$

where $f_1^e(s)$ and $f_2^e(s)$ are the Laplace transforms for the densities $f_1(s)$ and $f_2(s)$. For the Pareto distributions described above

$$f_j^e(s) = \alpha_j^{1-\alpha_j}(\alpha_j-1)^{\alpha_j}\mu^{\alpha_j}\Gamma\left(-\alpha_j, \frac{\mu(\alpha_j-1)s}{\alpha_j}\right),\ j=1,2,$$

(where $\Gamma(a,z) = \int_z^\infty t^{a-1}\mathrm{e}^{-t}dt$ for $a > 0$ and can be defined in terms of the confluent hypergeometric function of the first kind in the general case).

Numerical inversion of $r^e(s)$ requires an implementation of $\Gamma(a,z)$ that allows $a < 0$, and involves numerically inverting a Laplace transform. The *Mathematica* implementation of $\Gamma(a,z)$ allows $a < 0$, and the package `NLaplaceInversion.m` by Cheng, Sidauruk and Abousleiman [14] provides five inversion algorithms. The results of two algorithms are shown here. The Stehfest method approximates the desired function with a series having N_{max} terms. A value of $N_{max} = 20$, which is associated with the smallest relative error [14, p. 77], was chosen. Figure 6.1 shows $r(v)$,

$v = 0, 1, \ldots, 1000$ and the sample autocovariance $\hat{r}(v)$ for a simulated on-off process with the Pareto on and off times described above. The agreement is remarkably close.

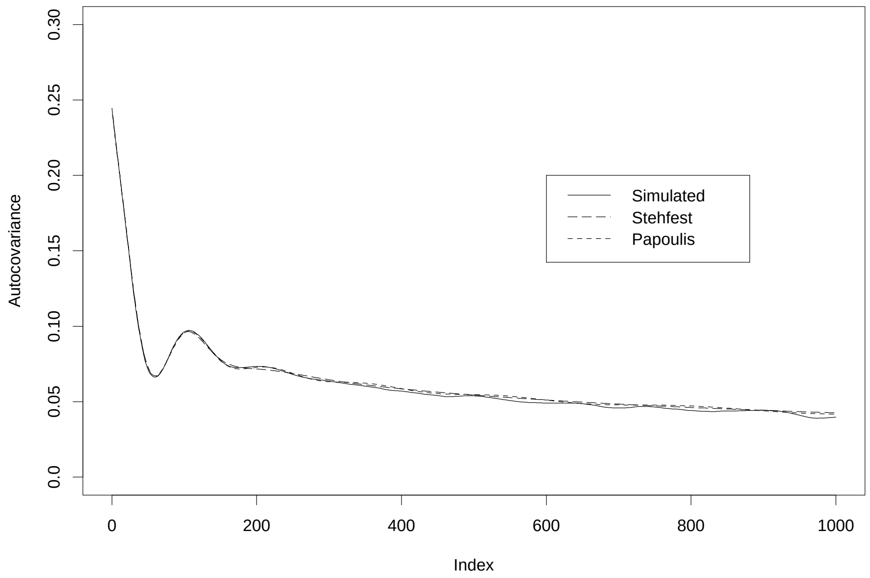

Figure 6.1: Autocovariance of unaggregated data from simulation and numerical inversion.

A second method, Papoulis Legendre, approximates the desired function with a different series of N_{max} terms. Along with N_{max}, there is an arbitrary parameter δ that can be set by the user. Again, $N_{max} = 20$ was chosen, and $\delta = 0.001$ was set based on the excellent qualitative agreement it provides with $\hat{r}(v)$. The other methods give a variety of results for $r(v)$ which seem inappropriate, including spikes, or poor agreement near $r(0) = \text{Var}[W(k)] = 0.24$.

With values for $r(v)$ one can use (6.15) to calculate variances such as those shown in Table 6.2. Figure 6.2 plots the sample mean versus Whittle estimate of variance $(\bar{\mu}_i, \hat{\sigma}^2_{10^4,i})$, $i = 1, \ldots, 1000$ for the four distribution assumptions for N_i. The distribution of N_i is reflected in the shape of the locus of points. But, using sample means of $\bar{\mu}_i$ and $\hat{\sigma}^2_{10^4,i}$ to roughly locate the centers, they are generally very close to (12000,420000). The one exception is from the exponential distribution, with a center around (11690,408150). Since the mean is also small, the smaller variances are likely a consequence of the skewed distribution.

	Mean	Stehfest Variance	Papoulis-Legendre Variance	Asymptotic Variance Approximation
W(k;100)	40	1342.3	1342.1	1584.9
X(k;100,300)	12000	402686	402622	475468

Table 6.2: Calculated variances using numerical inversion of $r^e(s)$ for constant $N = 1$ and $N = 300$

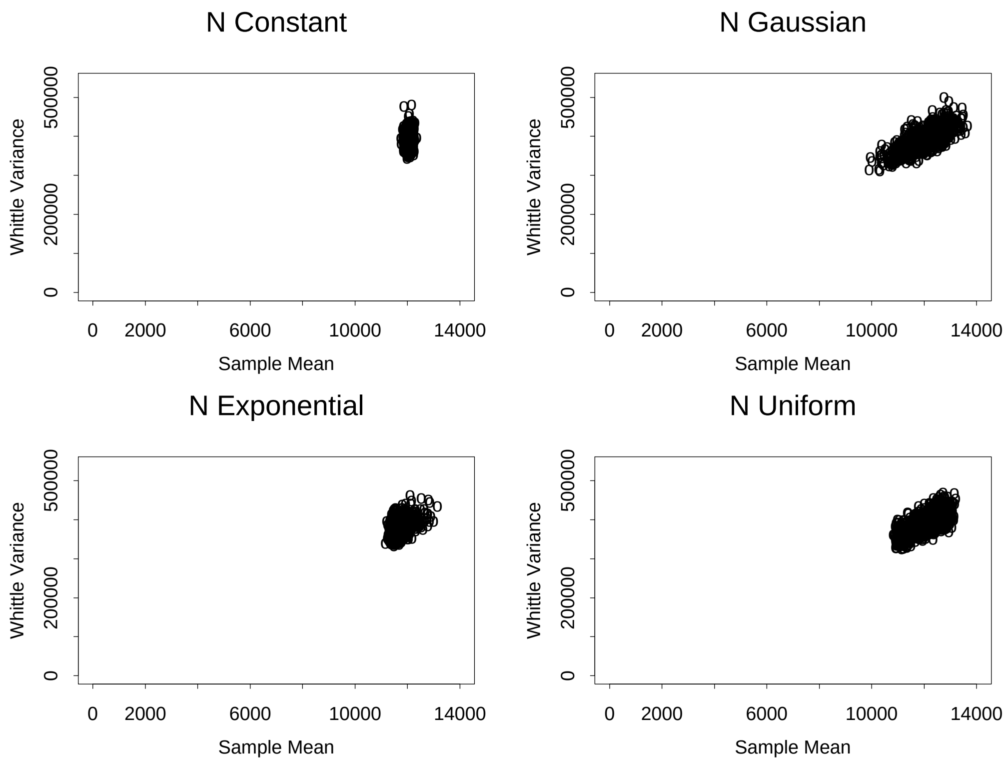

Figure 6.2: Sample mean vs. estimated Whittle variance from N_i simulated superpositions of on-off processes, $i = 1, \ldots, 1000$, where N_i is constant, Gaussian, exponentially or uniformly distributed.

	Whittle Variance Estimator $\hat{\sigma}^2_{10^4,i}$		Unbiased Sample Variance $S^2_{10^4,\hat{H}_{10^4,i},i}$	
	Laplace Inversion	Asymptotic Approximation	Laplace Inversion	Asymptotic Approximation
No randomness	0.062	0.122	0.079	0.117
Normal	0.081	0.130	0.097	0.124
Exponential	0.051	0.148	0.066	0.142
Uniform	0.082	0.133	0.096	0.127

Table 6.3: Root mean square relative error of variance estimates of $\{X_i(k; 100, 300),\ i = 1, \ldots, 1000\}$ relative to "true" value using Laplace inversion and asymptotic approximation.

Root mean square relative error, defined as $\left(\frac{1}{R}\sum_{i=1}^{R}\frac{(x_i-x_0)^2}{x_0^2}\right)^{1/2}$, for values x_i, $i = 1, \ldots, R$, and "true" value x_0, are shown in Table 6.3 for both Whittle scale estimates $\hat{\sigma}^2_{10^4,i}$ and unbiased sample variance estimates $S^2_{10^4,\hat{H}_{10^4,i},i}$. The rows describe the four assumptions for the distribution of the number of processes N_i, $i = 1, \ldots, 1000$. The errors are calculated relative to two "true" variances, 402 686 calculated using Laplace (Stehfest) inversion, and 475 468 calculated using the asymptotic approximation. Values from the Papoulis-Legendre algorithm are identical to those from the Stehfest algorithm and not shown.

Several observations can be made about Table 6.3. Errors in columns 1 and 3 are smaller than errors in columns 2 and 4. In other words, variances of simulated on-off data with known on and off distributions agree more closely with the variance calculated with Laplace inversion than with the asymptotic approximation. Also, errors in column 1 are smaller than their counterparts in column 3, with a difference of about 1.5%. But it would appear that both estimators show errors in the range of 6-10 %, which can be attributed mainly to errors in estimating the variance. The calculated variance is around 402 500, while the variance estimators are closer to 420 000, so about 4.3% too large. It was demonstrated above that the variance estimators have small bias for simulated Fractional Gaussian Noise data. The additional bias for the on-off sources would seem to arise from lack of agreement with the Fractional

Gaussian Noise model.

6.3.2 Estimation of γ

Simulation results for estimating γ are presented here. For the true value of γ, using Laplace (Stehfest) inversion for the variance, $\gamma = 402685/12000 = 33.6$, and using the asymptotic approximation for the variance, $\gamma = 475468/1200 = 39.6$. Table 6.4 shows values of the proposed estimators from 1000 runs of time-aggregated on-off superpositions $X_i(k; 100, N_i)$, $i = 1, \ldots, 1000$. As suggested by theory, an effect from the distribution of N_i is not apparent. Some bias is evident, due to bias in the variance estimate.

	$\hat{\gamma}^{(W)}$	$\hat{\gamma}^{(S)}$
No randomness	34.9	35.3
Normal	35.0	35.5
Exponential	34.9	35.3
Uniform	34.9	35.4

Table 6.4: Estimates of γ from simulated on-off sources.

By dividing the 1000 sample paths into twenty-five blocks of size 40, twenty-five i.i.d estimators $\hat{\gamma}_j^{(W)}$ and twenty-five i.i.d estimators $\hat{\gamma}_j^{(S)}$, $j = 1, \ldots, 25$ can be calculated. Figure 6.3 shows the results for $\hat{\gamma}_j^{(W)}$ under the four assumptions for the underlying distribution of N_i, $i = 1, \ldots, 1000$. (Plots for $\hat{\gamma}_j^{(S)}$ are similar although the values are slightly larger.) The values appear to be closely clustered. QQ-plots (not shown) confirm the notion that $\hat{\gamma}_j^{(W)}$ has a normal distribution.

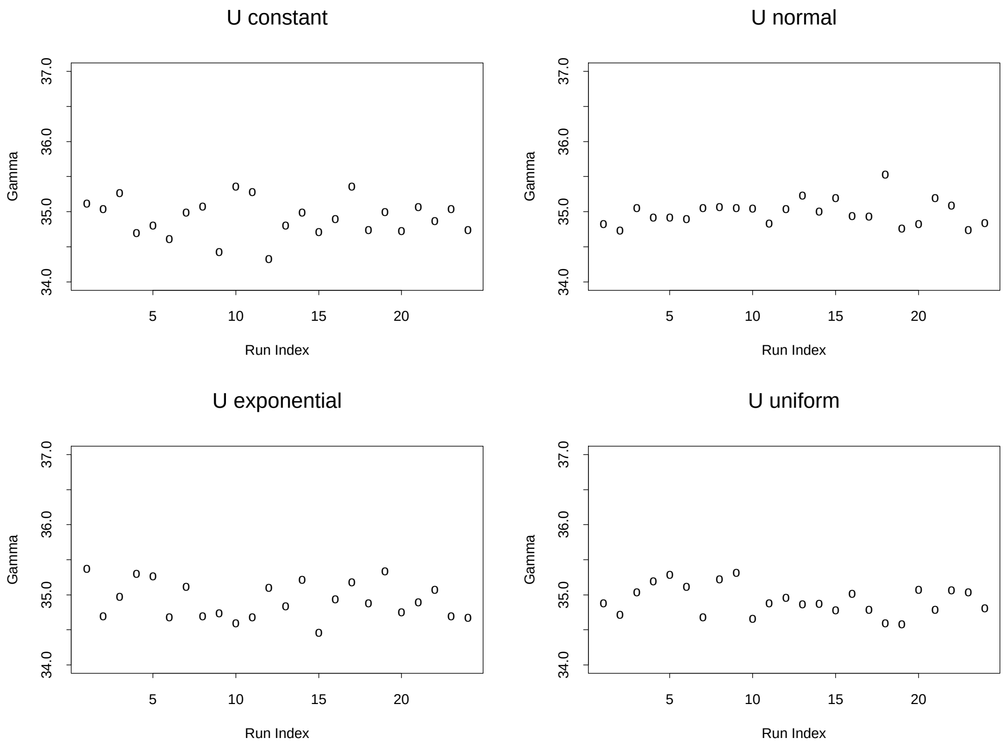

Figure 6.3: Values of $\hat{\gamma}_j^{(W)}$, $j = 1, \ldots, 25$, each from forty superpositions of on-off data.

Chapter 7

Queueing Results

7.1 Introduction

Consider the mixture process described in Section 5.6 for modelling data network traffic. The cumulative traffic in $[0, t)$ is

$$W^*(t) = Ut + \sqrt{\gamma U} Z_H(t)$$

where $Z_H(t)$ is Fractional Brownian Motion with Hurst parameter H and U is a non-negative random variable with density function $f_U(u)$. The incremental traffic (the traffic in $[k, k+1)$) is

$$X(k) = U + \sqrt{\gamma U} Y_H(k)$$

where $Y_H(k)$ is Fractional Gaussian Noise, with Hurst parameter H. Now, imagine the traffic is fed into a queue with server-rate C. As in (2.3), let the steady-state queue length be described by

$$Q^{(U,\gamma U,C)} = \sup_{t \geq 0} \{ \sqrt{\gamma U} Z_H(t) + (U - C)t \}$$

where now the mean rate m is replaced by the "random mean" U, and the variance is replaced by the "random variance" γU. The distribution of $Q^{(U,\gamma U,C)}$ may depend not only on properties of $Z_H(t)$, but also on the distribution of U. There are cases in which the distribution of U makes no difference when $H = 0.5$ but does change the distribution of $Q^{(U,\gamma U y,C)}$ when $H \neq 0.5$.

The server rate C may be constant. But, if bandwidth is *allocated* to the traffic stream, it may be the case that C is a function of U. The relation $U = \rho C$, $0 < \rho < 1$ can be used to allocate bandwidth to a stream of random mean rate, while keeping the utilization constant. Depending on other assumptions, this may or may not allow queue instability.

7.2 Connection to the Many-Process Scale Mixture Model

The Many-Process Scale Mixture model (Section 6.1.1) described a model for cumulative data network traffic in which the variance was a priori random. It is intended as a model to capture additional randomness from an unknown, random number of simultaneous users. In that model the cumulative traffic at time $\tau_i + t$ is

$$W_i^*(t) = U_i t + \sqrt{\gamma U_i} Z_{H,i}(t),\ t \in [0, \Delta), i = 1, \ldots, R \tag{7.1}$$

where Δ is the length of a time interval, R is the number of time intervals modelled, $\{Z_{H,i}(t),\ i = 1, \ldots, R\}$ are mutually independent, standard Fractional Brownian Motions, and $\{U_i,\ i = 1, \ldots, R\}$ is an i.i.d. sequence of non-negative random variables with distribution F_U, independent of the Fractional Brownian Motions. Let U be a generic random variable with distribution F_U. Now, let $Q^{(U_i,\gamma U_i,C_i)}$ be the steady-state queue length of a queue with server-rate C_i fed traffic $W_i^*(t)$ and let $Q^{(U,\gamma U,C)}$ be the queue length corresponding to

$$W^*(t) = Ut + \sqrt{\gamma U} Z_H(t)$$

with server-rate C. Here C_i may be a constant, or $C_i = C(U_i)$ may be a function of U_i. Then the random variables

$$\{Q^{(U_i,\gamma U_i,C_i)},\ i = 1, \ldots, R\}$$

are an i.i.d. collection of random variables with the same distribution as $Q^{(U,\gamma U,C)}$ and describe the marginal steady-state queue length when the value of U is unknown and random.

Each time-period i sees a particular value u_i for U_i according to F_U, with a corresponding steady-state queue length $Q^{(u_i,\gamma u_i,C_i)}$. Then the conditional steady-state queue lengths are

$$\{Q^{(u_i,\gamma u_i,C_{i,u_i})},\ i = 1, \ldots, R\}$$

where C_{i,u_i} is the value of C_i conditional $U_i = u_i$. These random variables are mutually independent, but not identically distributed. In particular, they need not be distributed like $Q^{(U,\gamma U,C)}$.

7.3 Basic Relations

If the server rate C is constant, say $C = C_0$, the relationship between stability of the queue and values for U is readily apparent, and is depicted in Figure 7.1a.

Lemma 7.3.1. *If $C = C_0$ for some constant $C_0 > 0$*

$$P(Q^{(U,\gamma U,C_0)} < \infty) = P(U < C_0) \text{ and } P(Q = \infty) = P(U \geq C_0).$$

Proof. Condition on $U = u$. Then $Q^{(u,\gamma u,C_0)}$ is a.s. finite if and only if $u < C_0$. □

Lemma 7.3.2. *If $C = C_0$ for some constant $C_0 > 0$*

$$P(Q^{(U,\gamma U,C_0)} > x) = E\left[\bar{F}_Q\left(\frac{(C_0 - U)^{H/(1-H)}}{(u\gamma)^{1/2(1-H)}}\, x\right) I_{[U<C_0]}\right] + P(U \geq C_0)$$

where $\bar{F}_Q(x) = P(Q^{(0,1,1)} > x)$.

Proof. By conditioning on U and using the scaling law in Lemma 2.1.1,

$$\begin{aligned}
\mathrm{P}(Q^{(U,\gamma U,C_0)} > x) &= \int_0^\infty \mathrm{P}\left(Q^{(u,\gamma u,C_0)} > x\right) f_U(u)\,du \\
&= \int_0^{C_0} \mathrm{P}\left(Q^{(0,1,1)} > \frac{(C_0 - u)^{H/(1-H)}}{(u\gamma)^{1/2(1-H)}}\, x\right) f_U(u)\,du + \mathrm{P}(U \geq C_0) \\
&= \int_0^{C_0} \bar{F}_Q\left(\frac{(C_0 - u)^{H/(1-H)}}{(u\gamma)^{1/2(1-H)}}\, x\right) f_U(u)\,du + \mathrm{P}(U \geq C_0).
\end{aligned}$$

□

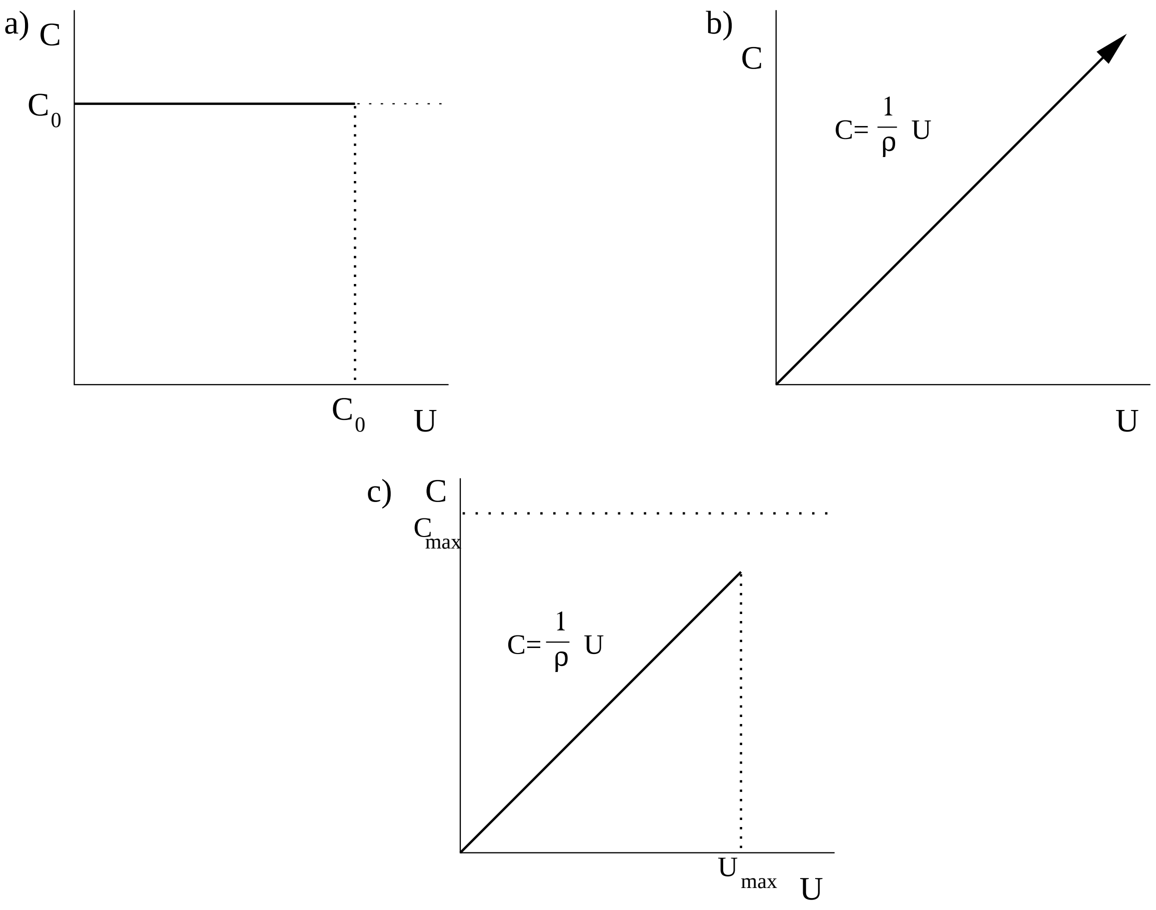

Figure 7.1: Three cases for server rate C versus U: a) $C = C_0$, b) $C = \frac{1}{\rho}U$, and c) $C = \frac{1}{\rho}U$, $U_{max} < C_{max}$. Values of U corresponding to queue stability are shown by a solid line.

If $C = C_0$ is constant it will be assumed that $\mathrm{P}(U \geq C_0) = 0$. This amounts to assuming the mean traffic rate cannot exceed the server rate. The alternative would lead to an unstable queue. Since U may have an unknown distribution, this is a non-trivial assumption. However, physical restrictions may impose a maximum rate on the traffic and then the server rate may be engineered larger. Or, admission control may restrict the traffic so that the mean rate cannot exceed C_0.

If the server rate is allocated to the traffic stream in proportion to its mean rate, so that $C = U/\rho$ for some constant $0 < \rho < 1$, then a variety of results can be obtained. The use of ρ for the constant of proportionality is well-justified since, by definition,

$$\text{server utilization} = \frac{U}{C} = \frac{U}{\frac{U}{\rho}} = \rho.$$

Differences between the case $H = 0.5$ and the long-range dependent case $0.5 < H < 1$ appear. If $H = 0.5$, the distribution of U plays no role.

Corollary 7.3.3. *a) If $C = \frac{1}{\rho}U$,*

$$P(Q^{(U,\gamma U,C)} > x) = E\left[\bar{F}_Q\left(\frac{(\frac{1}{\rho}-1)^{H/(1-H)}U^{(2H-1)/(2-2H)}}{\gamma^{1/2(1-H)}}x\right)\right], \; x > 0$$

where $\bar{F}_Q(x) = P(Q^{(0,1,1)} > x)$.
b) If $C = \frac{1}{\rho}U$ and, in addition, $H = 0.5$

$$P(Q^{(U,\gamma U,C)} > x) = e^{-\frac{(1/\rho)-1}{\gamma}x}, \; x > 0$$

Proof. For (a)

$$\begin{aligned} \mathrm{P}(Q^{(U,\gamma U,C)} > x) &= \int_0^\infty \mathrm{P}(Q^{(u,\gamma u,\frac{u}{\rho})} > x) f_U(u)\,du \\ &= \int_0^\infty \mathrm{P}\left(Q^{(0,1,1)} > \frac{(\frac{u}{\rho}-u)^{H/(1-H)}}{(\gamma u)^{1/2(1-H)}}x\right) f_U(u)\,du \\ &= \int_0^\infty \mathrm{P}\left(Q^{(0,1,1)} > \frac{(\frac{1}{\rho}-1)^{H/(1-H)}u^{H/(1-H)}}{(\gamma u)^{1/2(1-H)}}x\right) f_U(u)\,du \\ &= \int_0^\infty \bar{F}_Q\left(\frac{(\frac{1}{\rho}-1)^{H/(1-H)}u^{(2H-1)/(2-2H)}}{\gamma^{1/2(1-H)}}x\right) f_U(u)\,du. \end{aligned}$$

For (b), note that the exponent of u in the last line is zero, and combine with (2.5). □

Notice that under the assumption $C = \frac{1}{\rho}U$, necessarily $U < C$ and so the queue is stable. This includes the case in which both U and C are assumed to have infinite support (Figure 7.1b) and the case in which $U < U_{max}$, $C < C_{max}$ (so both U and C are bounded) and $C_{max} > \frac{1}{\rho}U_{max}$ (Figure 7.1c).

If U has infinite support (e.g. U has an exponential or Pareto distribution) say $U_{max} = \infty$ otherwise let U_{max} denote the upper-bound for the support of U. The case $U_{max} \geq C_{max}$ again allows an unstable queue, with probability $\mathrm{P}(U \geq C_{max})$.

7.4 Bounds for the Queue Length

The following bound extends a similar bound from Norros [67] to the situation with random mean rate.

Theorem 7.4.1.

$$P(Q^{(U,\gamma U,C)} > x) \geq P\left(Z \geq \frac{x^{1-H}}{\kappa(H)} \frac{(C-U)^H}{\sqrt{\gamma U}}\right)$$

where $Z \sim N(0,1)$ is independent of U and $\kappa(H) = H^H(1-H)^{1-H}$.

Proof. By conditioning on U

$$\begin{aligned}
\mathrm{P}(Q^{(U,\gamma U,C)} > x) &= \int_0^\infty \mathrm{P}(Q^{(u,\gamma u,C)} > x) f_U(u)\, du \\
&\geq \int_0^\infty \max_{t\geq 0}\{\mathrm{P}(\sqrt{\gamma U} Z_H(t) > (C-u)t + x)\} f_U(u)\, du.
\end{aligned}$$

As in Theorem 4.1 of Norros [67, p. 394]

$$\max_{t\geq 0}\{\mathrm{P}(\sqrt{\gamma u} Z_H(t) > (C-u)t + x)\} = \mathrm{P}\left(Z \geq \frac{x^{1-H}}{\kappa(H)} \frac{(C-u)^H}{\sqrt{\gamma u}}\right)$$

where $Z \sim \mathrm{N}(0,1)$. Therefore

$$\mathrm{P}(Q^{(U,\gamma U,C)} > x) \geq \int_0^\infty \mathrm{P}\left(Z \geq \frac{x^{1-H}}{\kappa(H)} \frac{(C-u)^H}{\sqrt{\gamma u}}\right) f_U(u)\, du$$

and the results follows by the Law of Total Probability. Independence of Z and U follows from the independence of $Z_H(t)$ and U. □

Corollary 7.4.2. *a) If $C = \frac{1}{\rho}U$*

$$P(Q^{(U,\gamma U,\frac{1}{\rho}U)} > x) \geq P\left(Z \geq \frac{x^{1-H}(1-\frac{1}{\rho})^H}{\kappa(H)\sqrt{\gamma}} U^{H-1/2}\right).$$

b) If $C = \frac{1}{\rho}U$ and, in addition, $H = 0.5$

$$P(Q^{(U,\gamma U,\frac{1}{\rho}U)} > x) \geq P\left(Z \geq \frac{2\sqrt{1-\frac{1}{\rho}}}{\sqrt{\gamma}}\sqrt{x}\right).$$

We can ask about the relationship between $\mathrm{P}(Q^{(U,\gamma U,\frac{1}{\rho}U)} > x)$ and $\mathrm{P}(Q^{(m,\gamma m,\frac{1}{\rho}m)} > x)$. The latter represents the traditional view in which the mean input rate, $m > 0$ is specifically non-random. Such a comparison seems most appropriate under the assumption $\mathrm{E}[U] = m$ so that the unconditional mean rates are equal. The following theorem addresses this situation. Because of its length, the proof is deferred to Appendix B. It uses the exact expression for the queue length in (2.13), Fubini's Theorem, and Jensen's inequality.

Theorem 7.4.3. *For $0.5 < H < 1$ if $E[U] = m$*

$$P(Q^{(U,\gamma U,\frac{1}{\rho}U)} > x) \geq P(Q^{(m,\gamma m,\frac{1}{\rho}m)} > x)$$

with equality only if $P(U = m) = 1$.

Proof. See Appendix B. □

7.5 Asymptotic Results for the Queue Length

This section describes the *unconditional* asymptotic behaviour for $P(Q^{(U,\gamma U,C)} > x)$. The main result expresses the asymptotics of $P(Q^{(U,\gamma U,C)} > x)$ in terms of the expected value of a function of U. The following simple lemma is needed to deal with lower-order terms.

Lemma 7.5.1. *Let $q(x) > 0$ and $R(x)$ be functions on $(0, \infty)$ such that $\lim_{x\to\infty} \frac{R(x)}{q(x)} = 0$. Let $g(u)$ be an increasing function on $(0, \infty)$, and let U be a random variable on $[l_1, \infty]$ or $[l_1, l_2]$ for some positive constants $l_1 > 0$, $l_2 > 0$, $l_1 < l_2$. Then*

$$\lim_{x\to\infty} \frac{E[R(x\, g(U))]}{E[q(x\, g(U))]} = 0.$$

Proof. By assumption, given $\varepsilon > 0$ $\exists x_0 > 0$ such that $\left|\frac{R(x)}{q(x)}\right| < \varepsilon$ for $x > x_0$. Let x_1 be such that $x_1 = \frac{x_0}{g(l_1)}$, so that $x\, g(l_1) > x_0$ for $x > x_1$ and $x\, g(u) > x_0$ for $u \geq l_1, x > x_1$ since g is increasing. Then, for $u \geq l_1, x > x_1$

$$0 \leq |R(x\, g(u))| < \varepsilon q(x\, g(u))$$

and since $U > l_1$ it follows that for $x > x_1$

$$0 \leq \mathrm{E}[|R(x\, g(U))|] < \varepsilon \mathrm{E}[q(x\, g(U))]$$

and

$$0 \leq \frac{\mathrm{E}[|R(x\, g(U))|]}{\mathrm{E}[q(x\, g(U))]} < \varepsilon.$$

Since $0 \leq R(x\, g(U)) \leq |R(x\, g(U))|$ and ε is arbitrary, the result follows. $\square$

Theorem 7.5.2. *Let U be a random variable on $[l_1, \infty]$ or $[l_1, l_2]$ for some positive constants $l_1 > 0$, $l_2 > 0$, $l_1 < l_2$. For $0.5 < H < 1$*

$$P(Q^{(U, \gamma U, \frac{U}{\rho})} > x) \sim E\left[q(c^{\frac{1}{2-2H}} U^{\frac{2H-1}{2-2H}} x)\right] \tag{7.2}$$

where $c = \frac{(\rho^{-1}-1)^{2H}}{\gamma}$ and

$$q(x) = K_1 x^{-\frac{(1-H)(2H-1)}{H}} e^{-\frac{x^{2-2H}}{2\kappa^2(H)}}$$

and K_1 is a constant depending on H.

Proof. From (2.12)

$$\bar{F}_Q(x) = q(x) + R(x) \tag{7.3}$$

where $\bar{F}_Q(x) = \mathrm{P}(Q^{(0,1,1)} > x)$ and $R(x)$ is a remainder term such that $\lim_{x\to\infty} \frac{R(x)}{q(x)} = 0$ and K_1 is a known constant depending on H. Now combine Corollary 7.3.3 with Lemma 7.5.1 using $g(x) = c^{\frac{1}{2-2H}} u^{(2H-1)/(2-2H)}$ which is increasing for $0.5 < H < 1$ since the exponent is positive. $\square$

Now it is possible to give the asymptotic behaviour of the queue length in several examples. The approach in all cases is the same and relies on the following lemma concerning the incomplete Gamma function $\Gamma(z, a) = \int_a^\infty \mathrm{e}^{-t} t^{z-1}\, dt$.

Lemma 7.5.3. *For $0.5 < H < 1$*

$$\int_k^\infty e^{-\frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}} u^{(2H-1)(z-1)+(2H-2)}\, du$$
$$= \frac{1}{(2H-1)} \left(\frac{c}{2\kappa^2(H)}\right)^{-z} x^{-(2-2H)z} \Gamma\left(z, \frac{c}{2\kappa^2(H)} k^{2H-1} x^{2-2H}\right)$$

Proof. Using the substitution $t = \frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}$ one can see

$$
\begin{aligned}
&\Gamma\left(z, \frac{c}{2\kappa^2(H)} k^{2H-1} x^{2-2H}\right) \\
&= \int_{\frac{c}{2\kappa^2(H)} k^{2H-1} x^{2-2H}}^{\infty} \mathrm{e}^{-t} t^{z-1}\, dt \\
&= (2H-1)\left(\frac{c}{2\kappa^2(H)} k^{2H-1}\right)^z \int_k^\infty \mathrm{e}^{-\frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}} (u^{2H-1} x^{2-2H})^{z-1} u^{2H-2} x^{2-2H}\, du \\
&= (2H-1)\left(\frac{c}{2\kappa^2(H)} k^{2H-1}\right)^z x^{(2-2H)z} \int_k^\infty \mathrm{e}^{-\frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}} u^{(2H-1)(z-1)+(2H-2)}\, du,
\end{aligned}
$$

and the result follows. □

Example - Pareto Distribution

Suppose U has a Pareto distribution on (k, ∞) with characteristic exponent $\alpha > 1$ (thus finite mean μ_U), distribution function $F(x) = 1 - \left(\frac{k}{x}\right)^\alpha$ and density $f_U(x) = \alpha k^\alpha x^{-(\alpha+1)}$. Then

$$
\begin{aligned}
&\mathrm{E}\left[q(c^{\frac{1}{2-2H}} U^{\frac{2H-1}{2-2H}} x)\right] \\
&= \int_0^\infty K_1 c^{\frac{1-2H}{H}} u^{-\frac{(1-2H)^2}{2H}} x^{-\frac{(1-H)(2H-1)}{H}} \mathrm{e}^{-\frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}} f_U(u)\, du \\
&= \alpha k^\alpha K_1 c^{\frac{1-2H}{H}} x^{-\frac{(1-H)(2H-1)}{H}} \int_k^\infty u^{-\frac{(1-2H)^2}{2H}-(\alpha+1)} \mathrm{e}^{-\frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}}\, du \\
&= \alpha k^\alpha K_1 c^{\frac{1-2H}{H}} x^{-\frac{(1-H)(2H-1)}{H}} \int_k^\infty u^{-\frac{(1-2H)^2}{2H}-(\alpha+1)} \mathrm{e}^{-\frac{c}{2\kappa^2(H)} u^{2H-1} x^{2-2H}}\, du \\
&= \frac{\alpha k^\alpha K_1 c^{\frac{1-2H}{2H}}}{(2H-1)} \left(\frac{c}{2\kappa^2(H)}\right)^{-z_1} x^{-\frac{(1-H)(2H-1)}{H}-(2-2H)z_1} \Gamma\left(z_1, \frac{c}{2\kappa^2(H)} k^{2H-1} x^{2-2H}\right)
\end{aligned}
$$

where

$$z_1 = -\frac{4H^2 + 2\alpha H - 4H + 1}{2H(2H-1)}$$

is the solution of

$$-\frac{(1-2H)^2}{2H} - (\alpha+1) = (2H-1)(z_1 - 1) + (2H-2),$$

$$c = \frac{(\rho^{-1}-1)^{2H}}{\gamma},$$

and the last line follows by Lemma 7.5.3.

In the traditional case, without the additional randomness from U,

$$P(Q^{(m,m\gamma,\frac{m}{\rho})} > x) \sim K_1 m^{-\frac{(1-2H)^2}{2H}} c^{\frac{1-2H}{H}} x^{-\frac{(1-H)(2H-1)}{H}} \mathrm{e}^{-\frac{c}{2\kappa^2(H)} m^{2H-1} x^{2-2H}}.$$

The right-side can be compared with the expression above for $\mathrm{E}\left[q(c^{\frac{1}{2-2H}} U^{\frac{2H-1}{2-2H}} x)\right]$ to compare their asymptotic forms. For the Pareto distribution, $\mu_U = \frac{k\alpha}{\alpha-1}$, so an expression for k in terms of the mean would be $k = \frac{\mu_U(\alpha-1)}{\alpha}$. A reasonable comparison would assume $\mu_U = m$ so that $k = \frac{m(\alpha-1)}{\alpha}$. That assumption is made here. Form the ratio

$$\begin{aligned} r(x) &= \left(K_1 m^{-\frac{(1-2H)^2}{2H}} c^{\frac{1-2H}{H}} x^{-\frac{(1-H)(2H-1)}{H}} \mathrm{e}^{-\frac{c}{2\kappa^2(H)} m^{2H-1} x^{2-2H}}\right) \times \\ &\times \left[\frac{\alpha \frac{m(\alpha-1)}{\alpha}^{\alpha} K_1 c^{\frac{1-2H}{2H}}}{(2H-1)} \left(\frac{c}{2\kappa^2(H)}\right)^{-z_1} x^{-\frac{(1-H)(2H-1)}{H}-(2-2H)z_1} \times \right. \\ &\left. \times \Gamma\left(z_1, \frac{c}{2\kappa^2(H)} k^{2H-1} x^{2-2H}\right)\right]^{-1} \\ &= \mathrm{K}_2 \times \frac{\mathrm{e}^{-\eta m^{2H-1} x^{2-2H}}}{x^{-(2-2H)z_1} \Gamma\left(z_1, \eta \left(\frac{m(\alpha-1)}{\alpha}\right)^{2H-1} x^{2-2H}\right)} \end{aligned}$$

where K_2 is a positive constant, and $\eta = \frac{c}{2\kappa^2(H)}$. Figures 7.2–7.5 show log-log plots for the ratio $r(x)$ for various values of H, α, m and η, with $K_2 = 1$ for simplicity (the basic shape of the graph is unaffected). In all cases $r(x) \to 0$, demonstrating that the tail of $\mathrm{E}\left[q(c^{\frac{1}{2-2H}} U^{\frac{2H-1}{2-2H}} x)\right]$ goes to zero more slowly than $P(Q^{(m,m\gamma,\frac{m}{\rho})} > x)$ as $x \to \infty$, when $\mu_U = m$.

Example - "Truncated Pareto" Distribution

Suppose X has a Pareto distribution on $[l_1, \infty)$ with characteristic exponent $\alpha > 0$ and $U = X I_{[X \in [l_1, l_2]]}$ has a "truncated Pareto" distribution Then U has density $f_U(u) = \frac{l_1^\alpha l_2^\alpha}{l_2^\alpha - l_1^\alpha} u^{-(\alpha+1)} I_{[l_1,l_2]}(u)$, mean $\mu_U = \frac{\alpha}{\alpha+1} \frac{l_1^\alpha l_2^\alpha}{l_2^\alpha - l_1^\alpha} \left(l_2^{1-\alpha} - l_1^{1-\alpha}\right)$ and

$$\begin{aligned} &\mathrm{E}\left[q(c^{\frac{1}{2-2H}} U^{\frac{2H-1}{2-2H}} x)\right] \\ =& \frac{\alpha\, l_1^\alpha l_2^\alpha}{l_2^\alpha - l_1^\alpha} \frac{K_1 c^{\frac{1-2H}{2H}}}{(2H-1)} \left(\frac{c}{2\kappa^2(H)}\right)^{-z_1} x^{-\frac{(1-H)(2H-1)}{H}-(2-2H)z_1} \\ &\times \left(\Gamma\left(z_1, 0.5c\kappa^{-2}(H) l_2^{2H-1} x^{2-2H}\right) - \Gamma\left(z_1, 0.5c\kappa^{-2}(H) l_1^{2H-1} x^{2-2H}\right)\right) \end{aligned}$$

and again

$$z_1 = -\frac{4H^2 + 2\alpha H - 4H + 1}{2H(2H-1)}.$$

Example - Uniform Distribution

Suppose U is uniformly distributed on $[l_1, l_2]$. Then

$$\mathrm{E}\left[q(c^{\frac{1}{2-2H}}U^{\frac{2H-1}{2-2H}}x)\right] = \left[\Gamma(z_2, 0.5\, c\kappa^{-2}(H)l_2^{2H-1}x^{2-2H}) - \Gamma(z_2, 0.5c\kappa^{-2}(H)l_1^{2H-1}x^{2-2H})\right]$$
$$\times x^{\frac{-2(4H-1)(H-1)^2}{H(2H-1)}}\left(\frac{c}{2\kappa^2(H)}\right)^{-z_2}\frac{K_1 c^{\frac{1-2H}{2H}}}{(l_2-l_1)(2H-1)}$$

where

$$z_2 = -\frac{4H^2 - 6H + 1}{2H(2H-1)}$$

is the solution of

$$-\frac{(1-2H)^2}{2H} - (\alpha+1) = (2H-1)(z_2-1) + (2H-2).$$

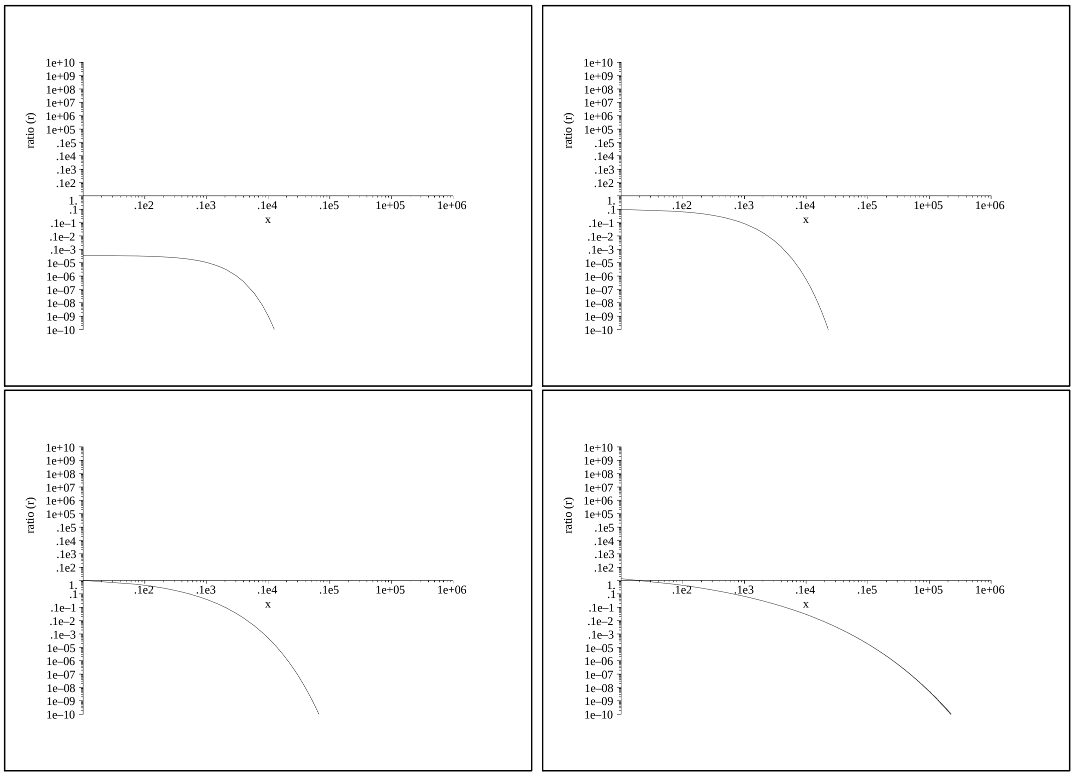

Figure 7.2: Effect of changing H on the ratio $r(x)$ of the asymptotic tails of the queue length distributions: $H = 0.6$ (top left), $H = 0.7$ (top right), $H = 0.8$ (bottom left), $H = 0.9$ (bottom right). In all cases $\alpha = 1.5$, $m = 100$, and $\eta = 0.1$.

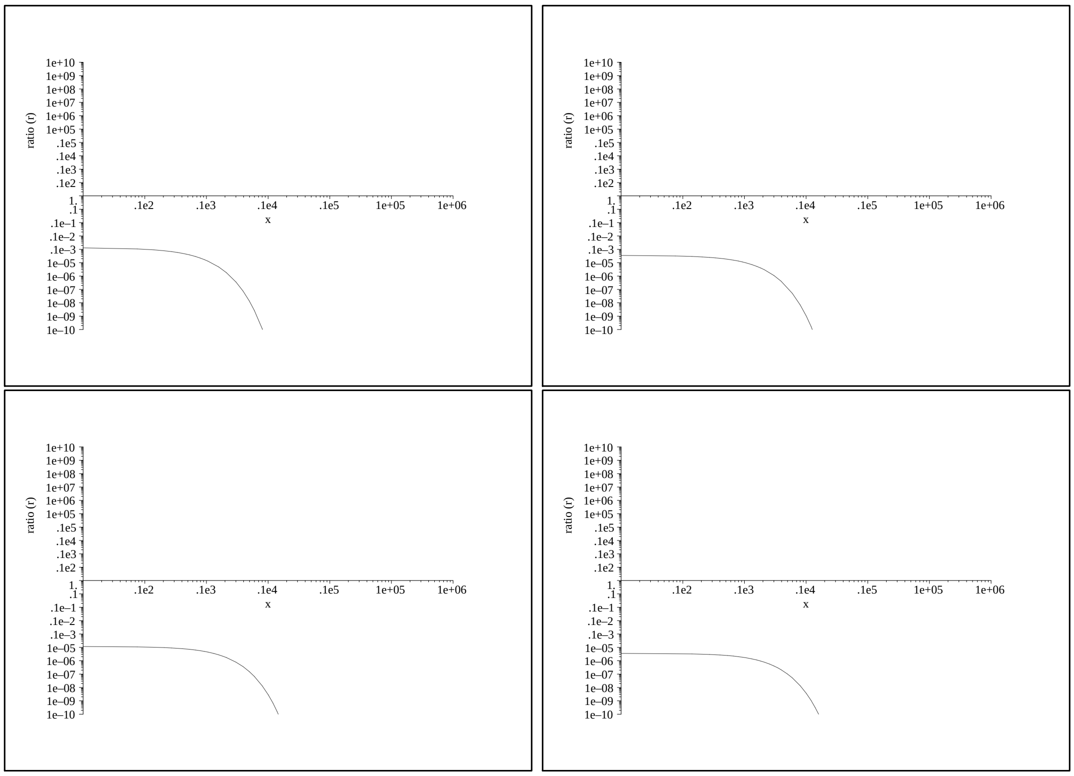

Figure 7.3: Effect of changing α on the ratio $r(x)$ of the asymptotic tails of the queue length distributions: $\alpha = 1.2$ (top left), $\alpha = 1.5$ (top right), $\alpha = 1.7$ (bottom left), and $\alpha = 1.9$ (top right). In all cases, $H = 0.6$, $m = 100$, $\eta = 0.1$.

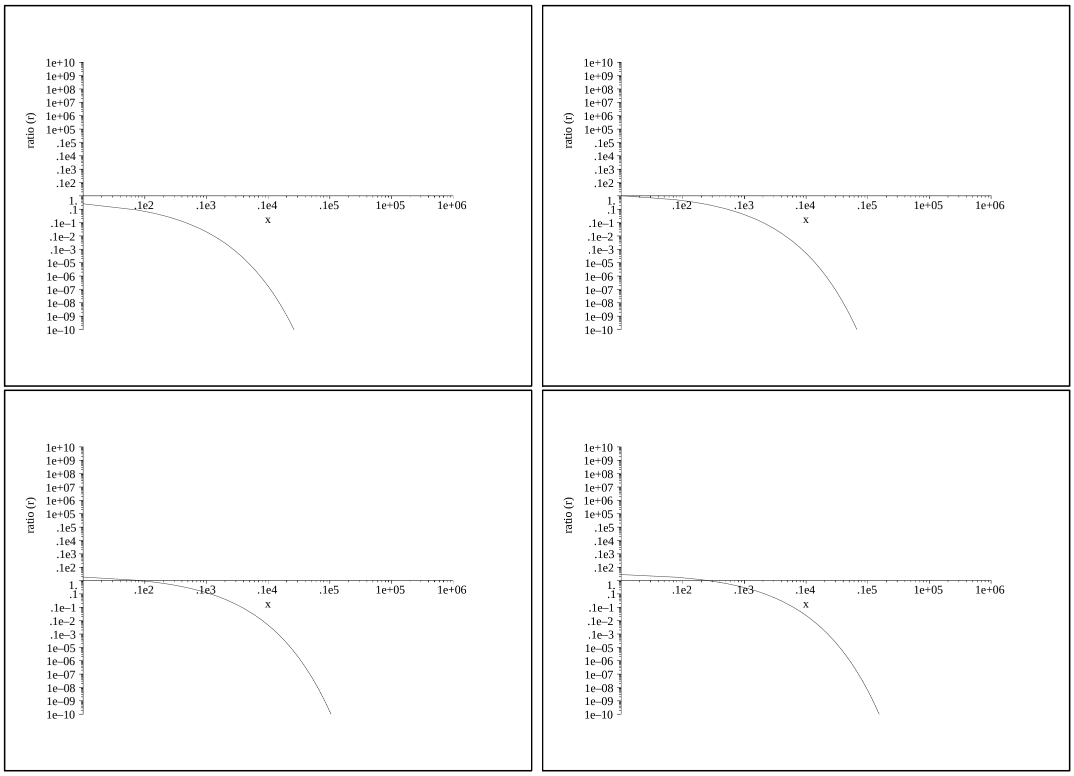

Figure 7.4: Effect of changing α on the ratio $r(x)$ of the asymptotic tails of the queue length distributions: $\alpha = 1.2$ (top left), $\alpha = 1.5$ (top right), $\alpha = 1.7$ (bottom left), and $\alpha = 1.9$ (top right). In all cases, $H = 0.8$, $m = 100$, $\eta = 0.1$.

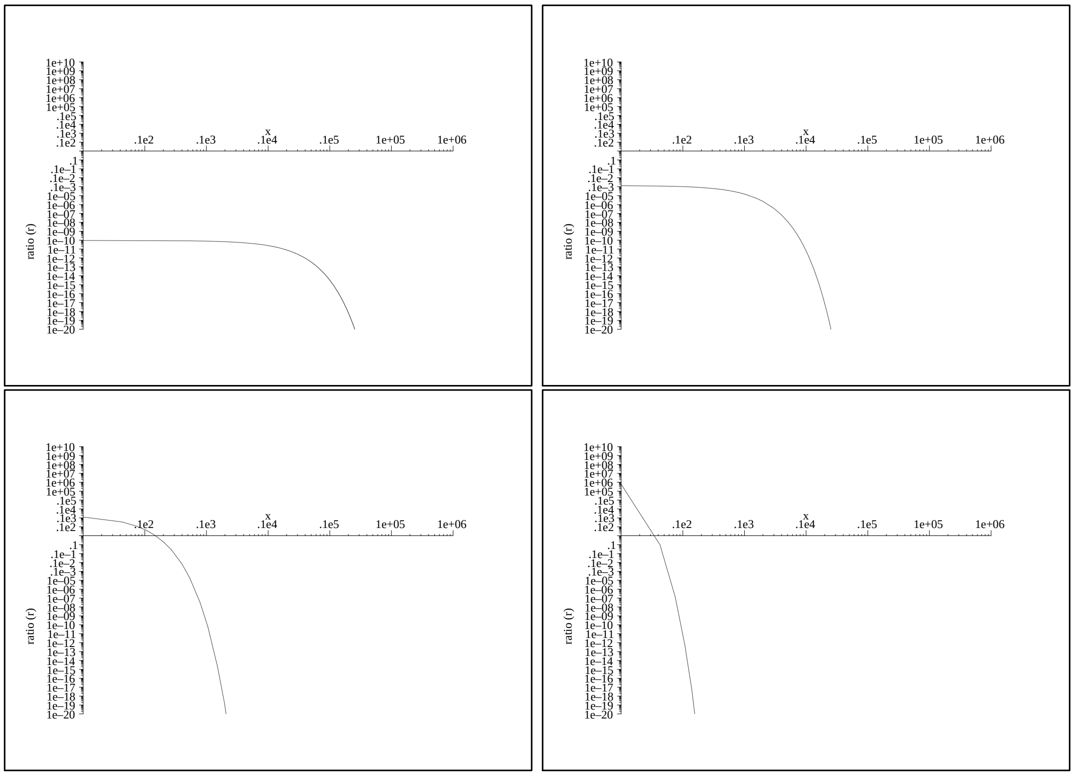

Figure 7.5: Effect of changing η on the ratio $r(x)$ of the asymptotic tails of the queue length distributions: $\eta = 0.01$ (top left), $\eta = .1$ (top right), $\eta = 1$ (bottom left), and $\eta = 10$ (top right). In all cases, $H = 0.6$, $m = 100$, $\alpha = 1.2$.

Chapter 8

Superposition of Alternating On-Off Processes with Lifetimes

Chapter 5 extends results of Taqqu, Willinger and Sherman [88] by using results for growing random sums to introduce additional randomness arising from a large but random number of on-off sources. The process representing the cumulative aggregate work, suitably scaled and centered, is a conditionally Gaussian scale mixture process. An assumption of that approach is that the number of sources, while a priori random, is constant over periods of interest. A natural extension would allow the number of on-off processes to be stochastic and varying with time. Such an approach immediately begs questions about which processes leave and when new processes arrive.

This chapter presents limit results for an aggregate of alternating on-off processes which arrive according to a Poisson process and have random holding times (also called "lifetimes"). The number of on-off processes in their holding periods can be viewed as the number of busy servers in an $M/G/\infty$ queuing system. Thus, the model can be considered as a hybrid of those in Cox and Isham [17] and Taqqu, Willinger and Sherman [88], although the technique owes more to the latter.

As an example, in a cable modem network the download (or upload) activity of each user could be modelled with an on-off process. But each user's session would not exist indefinitely. Rather, it would have a 'lifetime' of random length. Looking at the entire network, user sessions would be coming and going throughout the day. The maximum numbers of users may be known from subscription information, but the number at any given moment in time is random. The results of this chapter describe a model for the total bandwidth for the aggregate traffic when the number of users is large and random. This quantity is of interest because it represents approximately the aggregate bandwidth demand of an Internet Service Provider to provide Internet service, for example.

An assumption of the approach here will be that the lifetimes are independent of the on-off processes. This might be the case, for example, with users browsing web pages at leisure. Then the On times would correspond to the download activity for individual web pages, and the Off times would correspond to the time between requests for new pages. The 'lifetime' would be the length of the entire session, which is influenced by fatigue, meal time, the T.V. schedule, phone calls and many other possible interruptions that are independent of the web pages.

The rest of this chapter is organized as follows. Section 8.1 introduces necessary notation. Section 8.2 presents limit results for the incremental process. That is, the desired result follows after two limits, and the incremental process is the limiting process after just one limit. Section 8.3 describes the asymptotic behaviour of the variance of the incremental process. Section 8.4 describes the main result concerning the aggregate of alternating on-off processes with lifetimes.

8.1 Introduction

Let λ_n be a sequence of positive constants such that $\lambda_n \to \infty$ as $n \to \infty$, and $\{N_n(t), t > 0\}$ be a Poisson arrival process with rate λ_n that describes the number of sources that have joined the system in the time interval $[0, t)$. Associated with $\{N_n(t), t > 0\}$ is a sequence of arrival times $\{T_{\lambda_n,i}\}$, where $T_{\lambda_n,i}$ is the arrival time for source i.

Let $\{V_i\}$ be an i.i.d. sequence with distribution function $H(x)$, so that V_i is the holding time (or lifetime) for source i. Let $\{W_i(t), t \geq 0\}$, $i = 1, 2, \ldots$ be independent copies of the on-off processes $\{W(t), t \geq 0\}$ described in Section 2.1.2. The processes $\{T_{\lambda_n,i}\}$, $\{V_i\}$ and $\{W_i(t)\}$ are all assumed independent. Then we can form

$$X_i(t) = [W_i(t) - \mu_W] \, \mathbf{1}_{[T_{\lambda_n,i}, T_{\lambda_n,i}+V_i)}(t), \tag{8.1}$$

where $\mathbf{1}_A(x)$ is the characteristic function on the set A. The process $\{X_i(t), t \geq 0\}$ is a centered version of $\{W_i(t), t \geq 0\}$, given a 'lifetime' by confining it to the time period $[T_{\lambda_n,i}, T_{\lambda_n,i} + V_i)$ of random length V_i. Finally, let $\{G(t), t \geq 0\}$ be the mean-zero Gaussian process with autocovariance $r(t)$ in [88] arising as

$$\mathcal{L} \lim_{n\to\infty} \frac{\sum_{i=1}^{n} (W_i(t) - \mu_W)}{\sqrt{n}} = G(t) \tag{8.2}$$

where $\mathcal{L}\lim$ denotes convergence of the finite-dimensional distributions.

8.2 Limit Results for the Incremental Process

Let $Z(t)$ be a Gaussian process such that for $s < t$:

$$\begin{aligned} &\text{i) } E[Z(t)] = 0, \\ &\text{ii) } \operatorname{Var}[Z(t)] = \sigma_W^2 \int_0^t \bar{H}(z)\, dz, \text{ and} \\ &\text{iii) } \operatorname{Cov}[Z(s), Z(t)] = r(t-s) \int_{t-s}^{t} \bar{H}(z)\, dz, \end{aligned} \tag{8.3}$$

where $\bar{H}(z) = 1 - H(z)$. It will be shown that to obtain limiting results in this setting $Z(t)$ takes the place of the incremental process $G(t)$ in the limiting integrated process. It is not quite as nice to work with, however. Although for fixed s, $\operatorname{Cov}[Z(s), Z(t)] \to 0$ as $t \to \infty$, $Z(t)$ is not covariance stationary.

The Gaussian limiting result obtained here is a consequence of the simple result for the Poisson process

$$\frac{N_n(t)}{\lambda_n t} \xrightarrow{P} 1 \text{ as } n \to \infty \tag{8.4}$$

where λ_n is an increasing sequence of constants and $N_n(t)$ has rate $\lambda_n t$. The upper limit of the summation is a Poisson process. Because of the convergence to a constant, Korolev's result for growing random sums leads to a constant times a Gaussian random variable. Other increasing processes could be considered. In particular, ones for which the limit in (8.4) is a non-degenerate random variable could prove interesting. These could lead to non-Gaussian limiting processes. This is left for future work.

Theorem 8.2.1. *For the processes $X_i(t)$ of (8.1) and $Z(t)$ of (8.3)*

$$\mathcal{L} \lim_{n\to\infty} \frac{\sum_{i=1}^{N_n(t)} X_i(t)}{\sqrt{\lambda_n}} = Z(t). \tag{8.5}$$

Proof. For any $d \in \mathcal{Z}^+$, let $(c_1, \ldots, c_d)$ be arbitrary and let $t_1 < t_2 < \cdots < t_d$ be a partition of the time axis. To simplify the exposition, let $t_0 = 0$, $I_k = [t_k, t_{k+1})$, $k = 0, 1, \ldots, d-1$, $I_d = [t_d, \infty)$ and let

$$Z_n = \frac{c_1 \sum_{k=1}^{N_n(t_1)} X_k(t_1) + \cdots c_d \sum_{k=1}^{N_n(t_d)} X_k(t_d)}{\sqrt{\lambda_n}}. \tag{8.6}$$

The goal is to find the limiting distribution of Z_n as $n \to \infty$ and then use the Cramér-Wold Theorem [8, p. 29] to conclude that the left- and right-hand sides of (8.6) have the same finite-dimensional distributions.

Classify an arrival as type (i,j), $i = 0, \ldots, d$, $j = i, \ldots, d$, if it arrives in I_i and ends in I_j. Let $N_{n,i,j}(t)$ be the number of type (i,j) arrivals by time t, which depends on n through $N_n(t)$. Conditional on an arrival at time y, the probability an arrival is of type (i,j), $P_{(i,j)}(y)$, is

$$P_{(i,j)}(y) = \begin{cases} 0 & , y \notin I_i \\ H(t_{i+1} - y) & , y \in I_i,\ i = j \neq d \\ H(t_{j+1} - y) - H(t_j - y) & , y \in I_i,\ i < j,\ j \neq d \\ \bar{H}(t_d - y) & , y \in I_i,\ i < j,\ j = d \\ 1 & , y \in I_i,\ i = j = d. \end{cases}$$

By Proposition 5.3 of Ross [81, p. 273], for fixed t, the random variables $N_{n,i,j}(t)$, $i = 0, \ldots, d$, $j = i, \ldots, d$ are mutually independent and Poisson distributed with mean

(and variance)

$$\begin{aligned}\mathrm{E}[N_{n,i,j}(t)] &= \mathrm{Var}[N_{n,i,j}(t)] \\ &= \lambda_n \int_0^t P_{(i,j)}(y)dy = \lambda_n p_{(i,j)}(t),\ i = 0, \dots, d\,, j = i, \dots, d \end{aligned} \tag{8.7}$$

where $p_{(i,j)}(t) = \int_0^t P_{(i,j)}(y)dy$.

Note that for $t > t_{k+1}$ there can be no more arrivals starting in $I_0 \cup \cdots \cup I_k$ and so $N_{n,i,j}(t) = N_{n,i,j}(t_k-)$ for $t > t_{k+1}$, $i = 0, \dots, k$, $j = i, \dots, d$. As a consequence, we will restrict our attention of $N_{n,i,j}(t)$ to $t = t_d$ for all $i \le j < d$. The random variables $N_{n,i,j}(t_d)$ count the total number of arrivals starting in I_i and ending in I_j, $i = 0, \dots, d-1$, $j = i, \dots, d$. The related values $p_{i,j}(t_d)$ are shown in Table 8.1. For

$p_{i,j}(t_d)$	$j=0$	$j=1$	$\cdots$	$j=d-1$	$j=d$
$i{=}0$	$\int_0^{t_1} H(t_1-s)\,ds$	$\int_0^{t_1} H(t_2-s)-H(t_1-s)\,ds$	$\cdots$	$\int_0^{t_1} H(t_d-s)-H(t_{d-1}-s)\,ds$	$\int_0^{t_1} 1-H(t_d-s)\,ds$
$i{=}1$	0	$\int_{t_1}^{t_2} H(t_2-s)\,ds$	$\cdots$	$\int_{t_1}^{t_2} H(t_d-s)-H(t_{d-1}-s)\,ds$	$\int_{t_1}^{t_2} 1-H(t_d-s)\,ds$
$i{=}d{-}1$	0	0	$\ddots$	$\int_{t_{d-1}}^{t_d} H(t_d-s)\,ds$	$\int_{t_{d-1}}^{t_d} 1-H(t_d-s)\,ds$
$i{=}d$	0	0	$\cdots$	0	1

Table 8.1: Values of $p_{i,j}(t_d)$

$k = 1, \dots, d$ there is a convenient decomposition

$$N_n(t_k) = [N_n(t_k) - N_n(t_k-)] + \sum_{i=0}^{k-1}\sum_{j=i}^{d} N_{n,i,j}(t_d) \tag{8.8}$$

where the term in square brackets is added to include a possible arrival exactly at time t_k. Such an arrival is 'atypical' because it is the only possible way for a source to have a starting time outside $I_0 \cup \cdots \cup I_{k-1}$ and still be counted at t_k. Arrivals after t_d will be of no interest to us since they do not contribute to the behavior at the partition times, which are the times of interest.

Now the limiting behavior of Z_n can be considered. In (8.1) the lifetimes were explicitly forced by forcing the process to be zero before its arrival and after its departure. With the additional notation developed above, an equivalent form that is more easily analyzed is possible.

For each process k of type (i,j) it contributes to Z_n during its holding period at $t_{i+1},\dots,t_j$. Let $\left(X_k^{(i,j)}(t_{i+1}),\dots,X_k^{(i,j)}(t_j)\right)$ be copies of $\Big(W(t_{i+1})-\mu_W,\dots,W(t_j)-\mu_W\Big)$, mutually independent in i, j, and k. Notice that for fixed i and j the sequence of random vectors is i.i.d. Then the contribution of the k-th type (i,j) process is

$$c_{i+1}X_k^{(i,j)}(t_{i+1})+\cdots+c_jX_k^{(i,j)}(t_j)$$

and there are $N_{n,i,j}(t_d)$ such processes. Thus

$$Z_n \overset{D}{=} \sum_{i=0}^{d-1}\sum_{j=i+1}^{d}\left[\frac{\sum_{k=1}^{N_{n,i,j}(t_d)}\left(c_{i+1}X_k^{(i,j)}(t_{i+1})+\cdots+c_jX_k^{(i,j)}(t_j)\right)}{\sqrt{\lambda_n}}\right]. \tag{8.9}$$

(Strictly speaking there is also a term from arrivals occurring exactly at t_k, $k=1,2,\dots,d$. Since these arrivals have probability zero that term can be safely ignored.) Now two lemmas will be used to consider the limiting behavior of the term in square brackets, and then the entire right-hand side as $n\to\infty$. The proofs are deferred to the end of the section.

Lemma 8.2.2.

$$\frac{\sum_{k=1}^{N_{n,i,j}(t_d)}\left(c_{i+1}X_k^{(i,j)}(t_{i+1})+\cdots+c_jX_k^{(i,j)}(t_j)\right)}{\sqrt{\lambda_n}} \overset{D}{\to} Z^{(i,j)} \text{ as } n\to\infty, \tag{8.10}$$

$i=0,\dots,d-1, j=i+1,\dots,d$, where $Z^{(i,j)}$ are mutually independent random variables such that

$$\begin{aligned}Z^{(i,j)} &= \sqrt{p_{(i,j)}(t_d)}\left(c_{i+1}G^{(i,j)}(t_{i+1})+\cdots+c_jG^{(i,j)}(t_j)\right)\\ &\sim N\left(0, p_{(i,j)}(t_d)\left[\sigma_W^2c_{i+1}^2+\cdots+\sigma_W^2c_j^2+2\sum_{u=i+1}^{j}\sum_{v=u+1}^{j}c_uc_vr(|t_v-t_u|)\right]\right)\end{aligned}$$

where $r(u)$ is the autocovariance of $\{G(t),\, t\geq 0\}$.

Lemma 8.2.3.

$$\sum_{i=0}^{d-1}\sum_{j=i+1}^{d}Z^{(i,j)} \overset{D}{=} \sum_{i=1}^{d}c_iZ(t_i). \tag{8.11}$$

By Lemmas 8.2.2 and 8.2.3

$$Z_n \xrightarrow{D} \sum_{i=1}^{d} c_i Z(t_i).$$

Since $(c_1, \cdots, c_d)$ and $(t_1, \cdots, t_d)$ are arbitrary, the result follows from the Cramér-Wold Theorem. $\square$

Proof of Lemma 8.2.2. For fixed i and j let

$$Y_k^{(i,j)} = c_{i+1} X_k^{(i,j)}(t_{i+1}) + \cdots + c_j X_k^{(i,j)}(t_j)$$

so we wish to show

$$\frac{\sum_{k=1}^{N_{n,i,j}(t_d)} Y_k^{(i,j)}}{\sqrt{\lambda_n}} \xrightarrow{D} Z^{(i,j)} \text{ as } n \to \infty. \tag{8.12}$$

The result will be established for the random sum by establishing necessary conditions and using a known convergence result.

By the definition of $\{G(t), t \geq 0\}$ and the Cramér-Wold Theorem [8, p. 29]

$$\frac{\sum_{k=1}^{M} Y_k^{(i,j)}}{\sqrt{M}} \xrightarrow{D} c_{i+1} G^{(i,j)}(t_{i+1}) + \cdots + c_j G^{(i,j)}(t_j) \text{ as } M \to \infty$$

where $\{G^{(i,j)}(t), t \geq 0\}$ are independent copies of $\{G(t), t \geq 0\}$ in i and j. (Since the processes $X_k^{(i,j)}(t)$ are mutually independent in i and j, $Y_k^{(i,j)}$ and $G^{(i,j)}(t)$ share this property.) Let

$$\Delta^2 = \text{Var}\left[c_{i+1} G^{(i,j)}(t_{i+1}) + \cdots + c_j G^{(i,j)}(t_j)\right]$$

so that

$$\frac{\sum_{k=1}^{M} Y_k^{(i,j)}}{\Delta\sqrt{M}} \xrightarrow{D} \Phi \text{ where } \Phi \sim \mathrm{N}(0,1).$$

From (8.7) and Chebyshev's inequality, for any $t > 0$ and any $\epsilon > 0$

$$\mathrm{P}\left(\left|\frac{N_{n,i,j}(t)}{\lambda_n p_{(i,j)}(t)} - 1\right| \geq \epsilon\right) \leq \frac{\text{Var}[N_{n,i,j}(t)]}{\epsilon^2 \lambda_n^2 p_{(i,j)}^2(t)} = \frac{1}{\epsilon^2 \lambda_n p_{(i,j)}(t)} \xrightarrow{P} 0 \text{ as } n \to \infty$$

and so

$$\frac{\sqrt{N_{n,i,j}(t_d)}}{\sqrt{\lambda_n\, p_{(i,j)}(t_d)}} \xrightarrow{P} 1 \text{ and } N_{n,i,j}(t) \xrightarrow{P} \infty \text{ as } n \to \infty.$$

Finally, the sequence $\{Y_k^{(i,j)}\}$ is i.i.d. with finite variance so it is easy to show (e.g. with Chebyshev's inequality) that the sequence is uniformly asymptotically constant. That is, given $\epsilon > 0$

$$\lim_{n\to\infty} \max_{1\le j\le n} \mathrm{P}\left(|Y_j| > \epsilon\sqrt{n}\right) = 0.$$

By Theorem 3.3.3 of Gnedenko and Korolev [30, pp. 64-65]

$$\frac{\sum_{k=1}^{N_{n,i,j}(t_d)} Y_k^{(i,j)}}{\Delta\sqrt{\lambda_n\, p_{(i,j)}(t_d)}} \xrightarrow{D} \Phi \text{ as } n \to \infty$$

and therefore

$$\begin{aligned}
\frac{\sum_{k=1}^{N_{n,i,j}(t_d)} Y_k^{(i,j)}}{\sqrt{\lambda_k}} &\xrightarrow{D} \sqrt{p_{(i,j)}(t_d)}\Delta\Phi \qquad \text{as } n \to \infty \\
&\overset{D}{=} \sqrt{p_{(i,j)}(t_d)}\left(c_{i+1}G^{(i,j)}(t_{i+1}) + \cdots + c_j G^{(i,j)}(t_j)\right) \\
&= Z^{(i,j)}.
\end{aligned}$$

Mutual independence of the $Z^{(i,j)}$ follows from mutual independence of $G^{(i,j)}(t)$.

□

Proof of Lemma 8.2.3. Consider the left-hand side of (8.11): $\sum_{i=0}^{d-1}\sum_{j=i+1}^{d} Z^{(i,j)}$. As the sum of a finite number of independent, mean-zero Gaussian random variables, it shares these properties. The variance is a linear combination of terms of the form c_{i+1}^2, $i = 0,\ldots,d-1$ and $c_{i+1}c_j$, $i = 0,\ldots,d-1$, $j = i+2,\ldots,d$. Coefficients for each term can be found, and are not too difficult because of some helpful simplifications. The c_{i+1}^2 term, which arises from arrivals before t_{i+1} ending after t_{i+1}, is

$$\begin{aligned}
&c_{i+1}^2\sigma_W^2\left[(p_{(0,i+1)}(t_d) + \cdots + p_{(0,d)}(t_d)) + (p_{(1,i+1)}(t_d) + \cdots + p_{(1,d)}(t_d)) + \cdots\right. \\
&\qquad\qquad\qquad\qquad \left.\cdots + (p_{(i,i+1)}(t_d) + \cdots + p_{(i,d)}(t_d))\right] \\
&= c_{i+1}^2\sigma_W^2\left[\int_0^{t_1}\bar{H}(t_{i+1}-z)\,dz + \int_{t_1}^{t_2}\bar{H}(t_{i+1}-z)\,dz + \cdots + \int_{t_i}^{t_{i+1}}\bar{H}(t_{i+1}-z)\,dz\right] \\
&= c_{i+1}^2\sigma_W^2\int_0^{t_{i+1}}\bar{H}(x)\,dx
\end{aligned}$$

where the last line follows from the substitution $x = t_{i+i} - z$ and the additivity of integrals. In a similar way, the $c_{i+1}c_j$ term, $i+1 < j$, which arises from arrivals before

t_{i+1} ending after t_j, is

$$2c_{i+1}c_j r(t_j - t_{i+1})\left[(p_{(0,j)}(t_d) + \cdots + p_{(0,d)}(t_d)) + (p_{(1,j)}(t_d) + \cdots + p_{(1,d)}(t_d)) + \cdots \right.$$
$$\left. \cdots + (p_{(i,j)}(t_d) + \cdots + p_{(i,d)}(t_d))\right]$$
$$= 2c_{i+1}c_j r(t_j - t_{i+1})\left[\int_0^{t_1} \bar{H}(t_j - z)\, dz + \int_{t_1}^{t_2} \bar{H}(t_j - z)\, dz + \cdots + \int_{t_i}^{t_{i+1}} \bar{H}(t_j - z)\, dz\right]$$
$$= 2c_{i+1}c_j r(t_j - t_{i+1})\int_{t_j - t_{i+1}}^{t_j} \bar{H}(x) dx$$

where the last line follows from the substitution $x = t_j - z$ and the additivity of integrals. Thus

$$\mathrm{Var}\left[\sum_{i=0}^{d-1}\sum_{j=i+1}^{d} Z^{(i,j)}\right]$$
$$\overset{D}{=} \sigma_W^2 \sum_{i=0}^{d-1}\left(c_{i+1}^2 \int_0^{t_{i+1}} \bar{H}(x)dx\right) + 2\sum_{i=0}^{d-1}\sum_{j=i+2}^{d}\left(c_{i+1}c_j r(t_j - t_{i+1})\int_{t_j - t_{i+1}}^{t_j} \bar{H}(x)dx\right)$$
$$= \sigma_W^2 \sum_{i=1}^{d}\left(c_i^2 \int_0^{t_i} \bar{H}(x)dx\right) + 2\sum_{i=1}^{d}\sum_{j=i+1}^{d}\left(c_i c_j r(t_j - t_i)\int_{t_j - t_i}^{t_j} \bar{H}(x)dx\right).$$

But, by the definition of $Z(t)$ and the Cramér-Wold Theorem, $\sum_{i=1}^d c_i Z(t_i)$ is a mean-zero Gaussian random variable and

$$\mathrm{Var}\left[\sum_{i=1}^{d} c_i Z(t_i)\right]$$
$$\overset{D}{=} \sigma_W^2 \sum_{i=1}^{d}\left(c_i^2 \int_0^{t_i} \bar{H}(x)dx\right) + 2\sum_{i=1}^{d}\sum_{j=i+1}^{d}\left(c_i c_j r(t_j - t_i)\int_{t_j - t_i}^{t_j} \bar{H}(x)dx\right).$$

Since both the left- and right-hand sides of (8.11) are mean-zero Gaussian random variables with the same variance the result is proven. □

8.3 Variance and Covariance of the Integrated Process

The goal of this section is to establish the asymptotic behaviour as t goes to infinity of the variance $V(t)$ and covariance $C(s,t)$, $s < t$ of the integrated process

$$Y_t = \int_0^t Z(u)\, du \tag{8.13}$$

where

$$V(t) = \mathrm{Var}\left[Y_t\right] \text{ and } C(s,t) = \mathrm{Cov}\left[Y_s, Y_t\right].$$

As in Taqqu *et. al* [88], the asymptotic behaviour of the variance will play an important role in characterizing the limiting process of $\{Y_{Tt}\}$ as $T \to \infty$ under a suitable scaling. Indeed, since $Z(u)$ is a Gaussian process with mean zero, so is $\{Y_{Tt}\}$ for fixed T. However, unlike those results in which the stationary process is integrated, $Z(u)$ is not stationary. This means the automatic deduction that $\{Y_t\}$ has stationary increments cannot be made. Careful consideration of the variance *and* covariance of $\{Y_t\}$ is necessary.

It is assumed that the distribution function of the lifetimes, $H(x)$, has a regularly varying tail so we may write

$$\bar{H}(x) = x^{-p} L_V(x), \ p > 0, p \neq 1$$

where $L_V(x)$ is slowly varying. (Equality may be written by the Characterization Theorem [9, p. 17]). This does not require that $\bar{H}(x)$ be continuous, since integrals may be taken as Lebesgue integrals. It does require that $\bar{H}(x) > 0$. Pareto distributions, and Stable distributions totally skewed to the right with infinite mean would satisfy this, while distributions bounded with probability 1 clearly would not. If $p = 1$ it is further assumed that

$$\int_1^\infty L_V(t)/t \, dt < \infty.$$

Consideration of the integrated tail of $H(x)$ will be necessary, and that depends on the value of p. If $p > 1$ then $-p < -1$ and

$$\bar{H}_I(x) = \int_x^\infty \bar{H}(z)\, dz \sim \frac{x^{-p-1}}{p-1} L_V(x) \text{ as } x \to \infty \tag{8.14}$$

by Proposition 1.4.3. If $0 < p < 1$, then $-1 < -p < 0$ and for sufficiently large X,

$$\int_X^y \bar{H}(z)\, dz \sim \frac{y^{-p+1} L_V(y)}{1-p} \text{ as } y \to \infty \tag{8.15}$$

by Proposition 1.4.2 or Karamata's Theorem 1.4.5. In the boundary case $p = 1$, with the assumption that $\int_1^\infty \frac{L_V(t)}{t}\, dt < \infty$ then

$$\frac{1}{L_V(x)} \int_x^\infty \frac{L_V(t)}{t}\, dt \to \infty \text{ as } x \to \infty. \tag{8.16}$$

As in Chapter 5, for the on and off-time distributions, it is assumed $1 < \alpha_{min} \leq 2$. Regular variation of the autocovariance of the stationary on-off processes $r(t)$ and its integral $\int_0^y r(t)\,dt$ will be needed. In the current setting, the first lemma requires that the tail of the on and off-time distributions, $\bar{F}_1(x)$ and $\bar{F}_2(x)$ have regularly varying tails in which the slowly varying functions have limits in $(0, \infty)$. The second lemma requires $r(u)$ be ultimately monotone as a means to connect $\mathrm{Var}\left[\int_0^t W(u)\,du\right]$ and $r(t)$ via their their Laplace transforms.

Lemma 8.3.1. *If the distributions F_1 and F_2 satisfy Theorem 2.1.5 then for $1 < \alpha_{min} < 2$*

$$r(u) = \frac{\sigma^2_{lim}}{2}(3-\alpha_{min})(2-\alpha_{min})u^{1-\alpha_{min}} L_r(u), \quad \textit{and} \tag{8.17}$$

$$\int_0^y r(u)\,du = \frac{\sigma^2_{lim}}{2}(3-\alpha_{min})y^{2-\alpha_{min}} L_{rI}(y) \tag{8.18}$$

for some slowly varying functions $L_r(y)$ and $L_{rI}(y)$ such that $lim_{y\to\infty}\frac{L_r(y)}{L(y)} = 1$ and $lim_{y\to\infty}\frac{L_{rI}(y)}{L(y)} = 1$.

Proof. For the first equation, Theorem 2.1.5 establishes

$$r(u) \sim \frac{\sigma^2_{lim}}{2}(3-\alpha_{min})(2-\alpha_{min})\bar{L}_{min}u^{1-\alpha_{min}}.$$

By definition $L(t) \equiv L_{min}(t)$ and

$$\lim_{t\to\infty}\frac{L(t)}{\bar{L}_{min}} = 1$$

so

$$r(u) \sim \frac{\sigma^2_{lim}}{2}(3-\alpha_{min})(2-\alpha_{min})L(u)u^{1-\alpha_{min}}.$$

The Characterization Theorem [9, p. 17] establishes the equality with the slowly varying function $L_r(t)$. An application of Karamata's Theorem 1.4.5 and the Characterization theorem establishes the second equation. □

Alternatively, it is known [88, p. 12] that

$$\mathrm{Var}\left[\int_0^t W(u)\,du\right] = 2\int_0^t\left(\int_0^v r(u)\,du\right)dv \sim \sigma^2_{lim}t^{3-\alpha_{min}}L(t) \text{ as } t\to\infty. \tag{8.19}$$

Lemma 8.3.2. *Assume $r(u)$ is ultimately monotone and $1 < \alpha_{min} < 2$. Under relation (8.19), for some slowly varying functions $L_r(y)$ and $L_{rI}(y)$ such that $lim_{y\to\infty}\frac{L_r(y)}{L(y)} = 1$ and $lim_{y\to\infty}\frac{L_{rI}(y)}{L(y)} = 1$,*

$$\int_0^y r(u)\,du = \frac{\sigma_{lim}^2}{2}(3-\alpha_{min})y^{2-\alpha_{min}}L_{rI}(y) \text{ and}$$
$$r(u) = \frac{\sigma_{lim}^2}{2}(3-\alpha_{min})(2-\alpha_{min})u^{1-\alpha_{min}}L_r(u).$$

Proof. Since $r(u)$ is ultimately monotone and $\lim_{u\to\infty} r(u) = 0$, $r(u)$ is ultimately non-negative or ultimately non-positive, and so $\int_0^y r(u)\,du$ is ultimately monotone. Then (8.19), the Monotone Density Theorem [9, p. 39], and positive coefficients yield

$$\int_0^y r(u)\,du \in RV(y^{2-\alpha_{min}}).$$

Use of the equality follows from the Characterization Theorem [9, p. 17].

Since $r(u)$ is ultimately monotone, the same approach yields the second equation. From (8.19) it follows that $\lim_{u\to\infty}\frac{L_r(u)}{L(u)} = 1$ and $\lim_{y\to\infty}\frac{L_{rI}(y)}{L(y)} = 1$ by Karamata's Theorem . □

The case $\alpha_{min} = 2$ is different and must be handled specially. For example, $r(u)$ may decrease exponentially fast, and so a regularly varying tail assumption is not valid although $\alpha_{min} = 2$. For now, only the case $1 < \alpha_{min} < 2$ will be considered. The cases in which $\alpha_{min} = 2$ from a regularly varying tail, or from a tail with faster decay, are left for future work.

To consider $V(t)$ several cases can be distinguished, and these are illustrated in Figure 8.1. Cases 1, 2 and 3 lead to $V(t) \in RV(4-\alpha_{\min}-p)$. Although cases 1 and 2 could be considered together, they are separated because of the limiting processes to which they correspond. Cases 2 and 3 must be distinguished because the integrated tail of $\bar{H}(x)$ must be handled differently with Karamata's Theorem depending on whether $0 < p < 1$ or $p > 1$. Case 4 leads to $V(t) \sim ct$ as $t \to \infty$ for some constant c. Separate cases for $p = 1$, $p > 2$, and $\alpha_{min} = 2$ (discussed above) are left for future work.

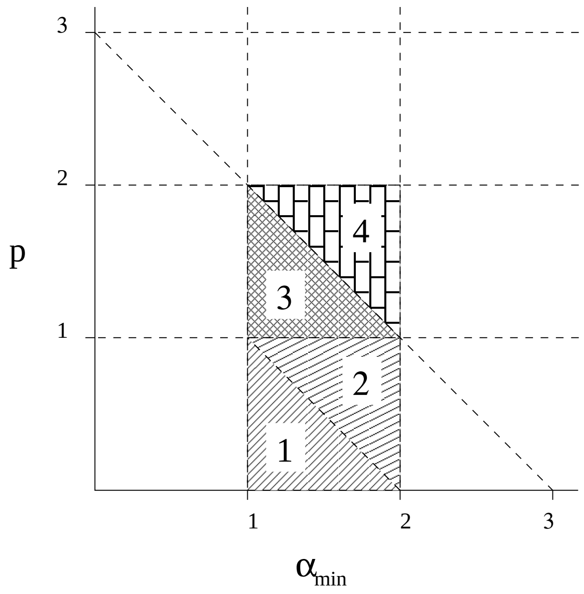

Figure 8.1: Differing cases for the tail behaviour of the variance $V(t)$. Cases 1, 2, and 3 all lead to $V(t) \in RV(4-\alpha_{\min}-p)$. Case 4 has $V(t) \sim ct$ for some constant c. The long diagonal line corresponds to $3-\alpha_{\min}-p=0$.

It is helpful to relate $V(t)$ exactly to the autocovariance $r(u)$. This is the content of the following lemma.

Lemma 8.3.3. *For the variance $V(t)$ and any constant $X \geq 0$*

$$V(t) = I_1 + I_2 + I_3 \text{ where}$$

$$\begin{aligned}
I_1 &= 2\int_0^X \int_0^y r(x) \left(\int_x^y \bar{H}(z)\,dz \right) dx\,dy, \\
I_2 &= 2\int_X^t \int_0^X r(x) \left(\int_x^y \bar{H}(z)\,dz \right) dx\,dy, \text{ and} \\
I_3 &= 2\int_X^t \int_X^y r(x) \left(\int_x^y \bar{H}(z)\,dz \right) dx\,dy.
\end{aligned}$$

Proof. Since $Z(u)$ is a mean zero process

$$\begin{aligned} V(t) &= \mathrm{E}\left[\int_0^t Z(u)\,du \int_0^t Z(v)\,dv\right] \\ &= 2\int_0^t \int_0^y \mathrm{E}\left[Z(u)Z(v)\right]\,dudv \\ &= 2\int_0^t \int_0^y r(|y-u|)\left(\int_{y-u}^y \bar{H}(z)\,dz\right)\,du\,dy \\ &= 2\int_0^t \int_0^y r(x)\left(\int_x^y \bar{H}(z)\,dz\right)\,dx\,dy \end{aligned}$$

where the last line follows from the substitution $x = y - u$. Figure 8.2 shows the various regions of integration that lead to the statement of the lemma.

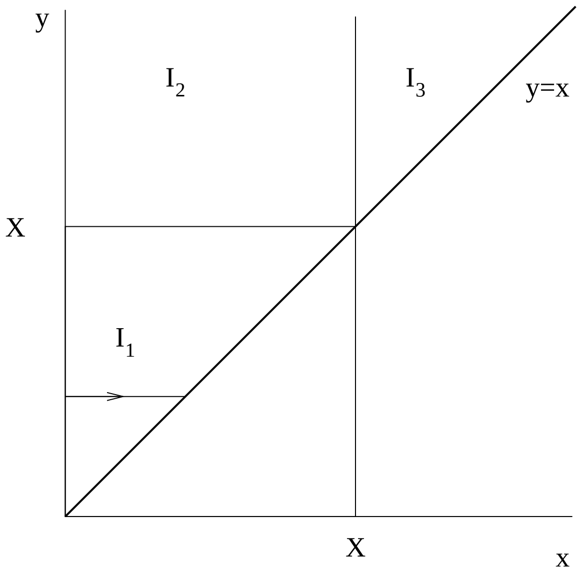

Figure 8.2: Regions of Integration: $V(t) = I_1 + I_2 + I_3$**.**

$\square$

Notice in particular that since $0 < \bar{H}(s) < 1$, I_1 is a finite constant for any fixed X- it does not depend on t.

The covariances will also be necessary. It is helpful to relate them to the variances as much as possible. For $s < t$

Lemma 8.3.4.

$$C(s,t) = \frac{1}{2}\left(V(s) + V(t) - I(s,t)\right) \tag{8.20}$$

where

$$I(s,t) = 2\int_s^t \int_0^{y-s} r(x) \left(\int_x^y \bar{H}(z)\,dz\right)\,dx\,dy.$$

Proof. From the simple relation

$$\mathrm{E}[(Y_t - Y_s)^2] = \mathrm{E}[Y_s^2] + \mathrm{E}[Y_t^2] - 2C(s,t)$$

follows that

$$C(s,t) = \frac{1}{2}\left(V(s) + V(t) + \mathrm{E}[(Y_t - Y_s)^2]\right). \tag{8.21}$$

By the definition of Y_t and the properties of $Z(t)$

$$\begin{aligned}
\mathrm{E}[(Y_t - Y_s)^2] &= \int_s^t \int_s^t \mathrm{E}[Z(u)Z(v)]\,du\,dv \\
&= 2\int_s^t \int_s^y r(y-u)\left(\int_{y-u}^y \bar{H}(z)\,dz\right)\,du\,dy \\
&= 2\int_s^t \int_0^{y-s} r(x)\left(\int_x^y \bar{H}(z)\,dz\right)\,dx\,dy \\
&= I(s,t)
\end{aligned}$$

where the third line follows by the change of variables $x = y - u$. □

This approach to the covariance is common in the literature. Often it is used with processes that have stationary increments, so that $I(s,t)$ is simply $V(t-s)$. Without stationary increments $I(s,t)$ must be considered separately from the variance. The convenient result that will be proven is that $I(Ts, Tt)$ will behave *asymptotically* as a function of $T(t-s)$ as $T \to \infty$, and that will be good enough to develop interesting results.

Theorem 8.3.5. *Under the previous assumptions for* $\{Z(t),\, t \geq 0\}$ *and* $\{Y(t)\, t \geq 0\}$, *and with*

$$\sigma^2 = \frac{\sigma_{lim}^2(3 - \alpha_{min})}{(3 - \alpha_{min} - p)(4 - \alpha_{min} - p)},$$

Cases 1, 2 and 3:

For $1 < \alpha_{min} < 2$ *and* $2 - \alpha_{min} - p > -1$

$$V(t) \sim \sigma^2 t^{4-\alpha_{min}-p} L_r(t) L_V(t) \text{ as } t \to \infty \text{ and}$$
$$I(s,t) \sim \sigma^2 (t-s)^{4-\alpha_{min}-p} L_r(t-s) L_V(t-s) \text{ as } (t-s) \to \infty.$$

Case 4:

For $1 < \alpha_{min} < 2$, $-2 < 2 - \alpha_{min} < -1$, $p < 2$

$$V(t) \sim 2ct \text{ where } c = \int_0^\infty r(x) \bar{H}_I(x)\, dx \text{ and}$$
$$I(s,t) \sim 2c(t-s) \;\text{ as } (t-s) \to \infty$$

Proof. By Proposition 1.4.1 there exists $X \geq 0$ such that $L_V(x)$ and $L_V(x)L_r(x)$ are locally bounded and locally integrable on $[X, \infty)$. In the following sections the behaviour of $V(Tt)$ as $T \to \infty$ will be important. Since only the tail of $V(t)$ is important for this, it is assumed in this proof that $s > X$, $t > X$ and $t - s > X$. In particular, this means the integrals I_1 and I_2 are non-zero.

Cases 1 and 2:

Consider integral I_2. Let $J_2(y)$ be the inner integral

$$J_2(y) = \int_0^X r(x) \left(\int_x^y \bar{H}(s) \right) dx \tag{8.22}$$

so that $I_2 = 2\int_X^t J_2(y)\, dy$. Then

$$\begin{aligned} J_2(y) &= \int_0^X r(x) \left(\int_0^y \bar{H}(z)\, dz - \int_0^x \bar{H}(z)\, dz \right) dx \\ &= c_1 \int_0^y \bar{H}(z)\, dz - c_2 \\ &= c_1 \int_X^y \bar{H}(z)\, dz + c_3 - c_2 \end{aligned}$$

where

$$c_1 = \int_0^X r(x)\, dx,\; c_2 = \int_0^X r(x) \left(\int_0^x \bar{H}(z)\, dz \right) dx, \text{ and } c_3 = c_1 \int_0^X \bar{H}(z)\, dz.$$

Therefore

$$\begin{aligned}
I_2 &= 2\int_X^t J_2(y)\,dy \\
&= 2c_1 \int_X^t \int_X^y \bar{H}(z)\,dz\,dy + 2\int_X^t (c_3 - c_2)dy \\
&= 2c_1 \int_X^t \int_X^y \bar{H}(z)\,dz\,dy + 2(c_3 - c_2)t - 2(c_3 - c_2)X
\end{aligned}$$

Using (8.15), since both $-p+1 > -1$ and $-p+2 > -1$, two applications of Proposition 1.4.2 yield

$$2c_1 \int_X^t \int_X^y \bar{H}(z)\,dz\,dy \sim \frac{t^{-p+2}}{(1-p)(2-p)} L_V(t) \text{ as } t \to \infty.$$

Since $-p+2 > 1$ the first term of I_2 is regularly varying with higher order than the linear term and so

$$I_2 \sim \frac{2c_1}{(1-p)(2-p)} t^{2-p} L_V(t) \text{ as } t \to \infty.$$

For integral I_3, let $J_3(y)$ be the inner integral

$$\begin{aligned}
J_3(y) &= \int_X^y r(x) \left(\int_x^y \bar{H}(z)\,dz \right) dx \\
&= \int_X^y \bar{H}(z)\,dz \int_X^y r(x)\,dx - \int_X^y r(x) \left(\int_X^x \bar{H}(z)\,dz \right) dx
\end{aligned}$$

so that $I_3 = 2\int_X^t J_3(y)\,dy$. For the first term, by (8.15) and Karamata's Theorem 1.4.5, since $1 - \alpha_{min} > -1$

$$\lim_{y\to\infty} \frac{\int_X^y \bar{H}(z)\,dz \int_X^y r(x)\,dx}{y^{3-\alpha_{min}-p} L_V(y) L_r(y)} = \frac{\sigma_{lim}^2 (3 - \alpha_{min})}{2(1-p)}.$$

For the second term, by Karamata's theorem 1.4.5, since $2 - \alpha_{min} - p > -1$

$$\lim_{y\to\infty} \frac{\int_X^y r(x) \left(\int_X^x \bar{H}(z)\,dz \right) dx}{y^{3-\alpha_{min}-p} L_V(y) L_r(y)} = \frac{\sigma_{lim}^2 (3 - \alpha_{min})(2 - \alpha_{min})}{2(1-p)(3 - \alpha_{min} - p)}.$$

Taking the difference yields

$$\lim_{y\to\infty} \frac{J_3(y)}{y^{3-\alpha_{min}-p} L_V(y) L_r(y)} = \frac{\sigma_{lim}^2 (3 - \alpha_{min})}{2(3 - \alpha_{min} - p)}.$$

Since $L_V(y)L_r(y)$ is slowly varying, and X was chosen so that $L_V(y)L_r(y)$ is locally bounded and integrable on $[X,\infty)$, $J_3(y)$ is regularly varying by Karamata's Theorem 1.4.6. By Karamata's Theorem 1.4.5, since $3-\alpha_{min}-p>0$

$$\begin{aligned} I_3 &= 2\int_X^t J_3(y)\,dy \\ &\sim \frac{\sigma^2_{lim}(3-\alpha_{min})}{(3-\alpha_{min}-p)(4-\alpha_{min}-p)} t^{4-\alpha_{min}-p} L_V(y)L_r(y) \text{ as } t\to\infty. \end{aligned}$$

By assumption $p<1<4-\alpha_{min}-p$ and so I_3 is regularly varying with higher order than I_1 and I_2. More specifically,

$$\frac{I_1}{t^{4-\alpha_{min}-p}L_V(y)L_r(y)} \in RV(-(4-\alpha_{min}-p))$$

and

$$\frac{I_2}{t^{4-\alpha_{min}-p}L_V(y)L_r(y)} \in RV(\alpha_{min}-2)$$

and both go to zero as $t\to\infty$ since both $-(4-\alpha_{min}-p)<0$ and $\alpha_{min}-2<0$ [9, p. 22]. Therefore

$$V(t) \sim \frac{\sigma^2_{lim}(3-\alpha_{min})}{(3-\alpha_{min}-p)(4-\alpha_{min}-p)} t^{4-\alpha_{min}-p} L_V(y)L_r(y).$$

To consider $I(s,t)$, let $M(y,s)$ be the inner integral

$$M(y,s) = \int_0^{y-s} r(x)\left(\int_x^y \bar{H}(z)\,dz\right)dx$$

so that

$$I(s,t) = 2\int_s^t M(y,s)dy.$$

Then by (8.22)

$$M(y,s) = J_2(y) + \int_X^{y-s} r(x)\left(\int_x^y \bar{H}(z)\,dz\right)dx$$

and $J_2(y)\sim \frac{1}{1-p}y^{1-p}L_V(y)$. In the second term of $M(y,s)$, the inner integral can be split so that the lower bound is X. Then by Karamata's Theorem 1.4.5 applied to the two parts

$$\frac{\int_X^{y-s} r(x)\left(\int_X^y \bar{H}(z)\,dz\right)dx}{y^{1-p}(y-s)^{2-\alpha_{min}}L_V(y)L_r(y-s)} = \frac{\sigma^2_{lim}(3-\alpha_{min})}{2(1-p)}$$

$$\frac{\int_X^{y-s} r(x)\left(\int_X^x \bar{H}(z)\,dz\right)\,dx}{(y-s)^{3-\alpha_{min}-p}L_V(y-s)L_r(y-s)} = \frac{\sigma_{lim}^2(3-\alpha_{min})(2-\alpha_{min})}{2(1-p)(3-\alpha_{min}-p)}.$$

Since $L_V(y)$ is slowly varying

$$\lim_{y\to\infty}\frac{L_V(y)}{L_V(y-s)} = \lim_{y\to\infty}\left(\frac{L_V(y)}{y}\frac{y}{y-s}\frac{y-s}{L_V(y-s)}\right) = 1$$

and $3-\alpha_{min}-p > 1-p$

$$\begin{aligned}\lim_{y\to\infty}\frac{M(y,s)}{(y-s)^{3-\alpha_{min}-p}L_V(y-s)L_r(y-s)} &= \frac{\sigma_{lim}^2}{2(1-p)}\left(1-\frac{2-\alpha_{min}}{3-\alpha_{min}-p}\right)\\ &= \frac{\sigma_{lim}^2(3-\alpha_{min})}{2(3-\alpha_{min}-p)}\end{aligned}$$

and $M(y,s)$ is regularly varying. This means for $I(s,t)$,

$$\begin{aligned}&I(s,t)\\ &= 2\int_s^t M(y,s)\,dy\\ &= 2\int_0^{t-s} M(w+s,s)\,dw\\ &= 2\left(\int_0^X M(w+s,s)\,dw - \int_X^{t-s} M(w+s,s)\,dw\right)\\ &\sim \frac{\sigma_{lim}^2(3-\alpha_{min})}{(3-\alpha_{min}-p)(4-\alpha_{min}-p)}(t-s)^{4-\alpha_{min}-p}L_V(t-s)L_r(t-s) \text{ as } (t-s)\to\infty\\ &= \sigma^2(t-s)^{4-\alpha_{min}-p}L_V(t-s)L_r(t-s).\end{aligned}$$

Case 3:

By assumption $1 < p < 2$ and $p < 3-\alpha_{min}$. Since $p > 1$ we must proceed differently from above. First consider I_2. Let $J_4(y)$ be the inner integral

$$\begin{aligned}J_4(y) &= \int_0^X r(x)\int_x^y \bar{H}(z)\,dz\,dx\\ &= \int_0^X r(x)\left(\bar{H}_I(x)-\bar{H}_I(y)\right)\,dx\\ &= \int_0^X r(x)\bar{H}_I(x)\,dx - \bar{H}_I(y)\int_0^X r(x)\,dx\\ &= c_5 - c_6\bar{H}_I(y)\end{aligned}$$

where

$$c_5 = \int_0^X r(x)\bar{H}_I(x)\,dx, c_6 = \int_0^X r(x)\,dx$$

and therefore

$$\begin{aligned} I_2 &= 2\int_x^t J_4(y)\,dy \\ &= 2c_5 t - 2c_6 \int_X^t \bar{H}_I(y)\,dy - 2c_5 X. \end{aligned}$$

By (8.14) and Karamata's Theorem 1.4.5

$$\lim_{y\to\infty} \frac{\int_X^t \bar{H}_I(y)\,dy}{t^{2-p}L_V(t)} = \frac{1}{(p-1)(2-p)}.$$

Since $2-p<1$, the linear term is regularly varying with higher order, and so

$$I_2 \sim 2c_5 t \text{ as } t\to\infty.$$

For I_3, let $J_5(y)$ be the inner integral

$$\begin{aligned} J_5(y) &= \int_X^y r(x)\left(\int_x^y \bar{H}(z)\,dz\right)dx \\ &= \int_X^y r(x)\left(\bar{H}_I(x) - \bar{H}_I(y)\right)dx \\ &= \int_X^y r(x)\bar{H}_I(x)\,dx - \bar{H}_I(y)\int_X^y r(x)\,dx \end{aligned}$$

so that $I_3 = 2\int_X^t J_5(y)\,dy$.

By (8.14) and Karamata's Theorem 1.4.5, since $(1-\alpha_{min})+(1-p) > -1$

$$\lim_{y\to\infty} \frac{\int_X^y r(x)\bar{H}_I(x)\,dx}{y^{3-\alpha_{min}-p}L_r(y)L_V(y)} = \frac{\sigma^2_{lim}(3-\alpha_{min})(2-\alpha_{min})}{2(p-1)(3-\alpha_{min}-p)}.$$

Similarly, by (8.14) and Karamata's Theorem 1.4.5, since $1-\alpha_{min} > -1$

$$\lim_{y\to\infty} \frac{\bar{H}_I(y)\int_X^y r(x)\,dx}{y^{3-\alpha_{min}-p}L_r(y)L_V(y)} = \frac{\sigma^2_{lim}(3-\alpha_{min})}{2(p-1)}.$$

As above for $J_3(y)$, $J_5(y)$ is regularly varying and

$$\begin{aligned} I_3 &= 2\int_X^t J_5(y)\,dy \\ &\sim \frac{\sigma^2_{lim}(3-\alpha_{min})}{(3-\alpha_{min}-p)(4-\alpha_{min}-p)} t^{4-\alpha_{min}-p} L_V(y)L_r(y) \text{ as } t\to\infty. \end{aligned}$$

Since $4-\alpha_{min}-p=(2-p)+(2-\alpha_{min})$ and $0<2-\alpha_{min}<1$, $4-\alpha_{min}-p>2-p$. Therefore I_3 is regularly varying with higher order than I_1 and I_2, and

$$V(t)\sim\frac{\sigma^2_{lim}(3-\alpha_{min})}{(3-\alpha_{min}-p)(4-\alpha_{min}-p)}t^{4-\alpha_{min}-p}L_V(y)L_r(y) \text{ as } t\to\infty.$$

The result for $I(s,t)$ can be shown in a manner similar to that for cases 1 and 2. But, since $p>1$ one must use

$$\int_x^y \bar{H}(z)\,dz=\bar{H}_I(x)-\bar{H}_I(y)$$

and adapt as necessary.

Case 4:

Since $2-\alpha_{min}-p<-1$ it is possible to use the second part of Karamata's theorem 1.4.5 which does not involve X. Thus X plays a less noticeable role in the proof.

$$\begin{aligned}
I_2+I_3&=2\int_0^t\int_0^y r(x)\left(\bar{H}_I(x)-\bar{H}_I(y)\right)\,dx\,dy\\
&=2\int_0^t\left(\int_0^y r(x)\bar{H}_I(x)\,dx-\bar{H}_I(y)\int_0^y r(x)\,dx\right)\,dy\\
&=2\int_0^t\left(\int_0^\infty r(x)\bar{H}_I(x)\,dx-\int_y^\infty r(x)\bar{H}_I(x)\,dx-\bar{H}_I(y)\int_0^y r(x)\,dx\right)\,dy\\
&=2\int_0^t c\,dx-2\int_0^t\left(\int_y^\infty r(x)\bar{H}_I(x)\,dx+\bar{H}_I(y)\int_0^y r(x)\,dx\right)\,dy\\
&=2ct-2\int_0^t\left(\int_y^\infty r(x)\bar{H}_I(x)\,dx+\bar{H}_I(y)\int_0^y r(x)\,dx\right)\,dy
\end{aligned}$$

where

$$c=\int_0^\infty r(x)\bar{H}_I(x)\,dx \tag{8.23}$$

which is finite by Karamata's Theorem 1.4.5 since for the sum of their tail indices $(1-\alpha_{min})+(1-p)<-1$.

By Karamata's Theorem 1.4.5, using techniques similar to those used above, as $y\to\infty$

$$\begin{aligned}
\int_y^\infty r(x)\bar{H}_I(x)\,dx&\sim\frac{\sigma^2_{lim}(3-\alpha_{min})(2-\alpha_{min})}{-2(p-1)(3-\alpha_{min}-p)}y^{4-\alpha_{min}}L_V(u)L_r(y),\\
\bar{H}_I(y)\int_0^y r(x)\,dx&\sim\frac{\sigma^2_{lim}(3-\alpha_{min})}{-2(p-1)}y^{4-\alpha_{min}}L_V(u)L_r(y)
\end{aligned}$$

and so

$$\lim_{y\to\infty} \frac{\int_0^t \left(\int_y^\infty r(x)\bar{H}_I(x)\,dx + \bar{H}_I(y)\int_0^y r(x)\,dx\right)\,dy}{y^{4-\alpha_{min}}L_V(u)L_r(y)} = 2\frac{\sigma^2_{lim}(3-\alpha_{min})}{(3-\alpha_{min}-p)(4-\alpha_{min}-p)}.$$

Since $4-\alpha_{min}-p<1$ the linear term is regularly varying with higher order. Thus

$$\lim_{t\to\infty}\frac{V(t)}{t} = 2c$$

while

$$\lim_{t\to\infty}\frac{V(t)}{t\,L_V(t)\,L_r(t)} = \lim_{t\to\infty}\frac{2c}{L_V(t)\,L_r(t)}.$$

The second limit requires more information about $L_V(t)$ and $L_r(t)$. Recall that both positive constants and the logarithm function are examples of slowly varying functions, and they would give quite different limits.

□

To consider $I(s,t)$, begin as for case 1 by letting $M_2(y,s)$ be the inner integral

$$\begin{aligned} M_2(y,s) &= \int_0^{y-s} r(x)\left(\int_x^y \bar{H}(z)\,dz\right)\,dx \\ &= c - \int_{y-s}^\infty r(x)\bar{H}_I(x)\,dx - \bar{H}_I(y)\int_0^X r(x)\,dx - \bar{H}_I(y)\int_X^{y-s} r(x)\,dx \end{aligned}$$

where c is defined in (8.23). Using Karamata's Theorem 1.4.5 the second, third and fourth terms are asymptotically regularly varying with negative index as $(y-s)\to\infty$, so the constant term has higher order.

This means for $I(s,t)$,

$$\begin{aligned} I(s,t) &= 2\int_s^t M_2(y,s)\,dy \\ &= 2\int_0^{t-s} M_2(w+s,s)\,dw \\ &\sim 2c(t-s) \text{ as } (t-s)\to\infty \end{aligned}$$

8.4 The Limiting Processes

Now it is possible the state main conclusions of this chapter. The first theorem is the limit result corresponding to cases 2 and 3, while the second theorem is for case 4.

Theorem 8.4.1. *For $1 < \alpha_{min} < 2$ and $1 < 4 - \alpha_{min} - p < 2$*

$$\mathcal{L}\lim_{T\to\infty}\mathcal{L}\lim_{n\to\infty}\frac{\int_0^{Tt}\sum_{i=1}^{N_n(u)}X_i(u)\,du}{\sqrt{\lambda_n L_V(t)L_r(T)T^H}} = \sqrt{\frac{\sigma_{lim}^2(2H-1+p)}{2H(2H-1)}}Z_H(t) \tag{8.24}$$

where $H = (4-\alpha_{min}-p)/2$ and $Z_H(t)$ is standard Fractional Brownian Motion with Hurst parameter H.

Proof. By Theorem 8.2.1,

$$\mathcal{L}\lim_{n\to\infty}\frac{\sum_{i=1}^{N_n(t)}X_i(t)}{\sqrt{\lambda_n}} = Z(t)$$

is Gaussian, mean zero and has covariance $\mathrm{Cov}[Z(s),Z(t)] = r(t-s)\int_{t-s}^{t}\bar{H}(z)\,dz$, where $\bar{H}(z) = 1 - H(z)$. Now, for any $t \geq 0$ and $T > 0$ let Y_{Tt} be defined by

$$Y_{Tt} = \frac{\int_0^{Tt}Z(u)\,du}{\sqrt{L_V(T)L_r(T)T^H}}$$

Then for the characteristic function

$$\begin{aligned}\mathrm{E}[\exp(isY_{Tt})] &= \exp\left(-\frac{V(Tt)}{2L_V(t)L_r(T)T^{2H}}s^2\right)\\ &\to \exp\left(-\frac{\sigma_{lim}^2(2H-1+p)}{2H(2H-1)}s^2\right) \text{ as } T\to\infty\end{aligned}$$

since

$$\lim_{T\to\infty}\frac{V(Tt)}{L_V(t)L_r(T)T^{2H}} = \frac{\sigma_{lim}^2(2H-1+p)}{2H(2H-1)}t^{2H}$$

by Theorem 8.3.5. Thus, for fixed $t \geq 0$

$$Y_{Tt} \xrightarrow{D} Y \text{ as } T\to\infty$$

where

$$Y \sim \mathrm{N}\left(0, \frac{\sigma_{lim}^2(2H-1+p)}{2H(2H-1)}t^{2H}\right). \tag{8.25}$$

For the covariances, by (8.21) and Theorem 8.3.5

$$\begin{aligned}\mathrm{Cov}[Y_{Ts},Y_{Tt}] &= \frac{1}{2}\frac{(V(Ts)+V(Ts)-C(Ts,Tt))}{L_V(T)\,L_r(T)\,T^{4-\alpha_{min}-p}}\\ &\to \sigma^2\left(s^{4-\alpha_{min}-p}+t^{4-\alpha_{min}-p}-(t-s)^{4-\alpha_{min}-p}\right) \text{ as } T\to\infty\\ &= \sigma^2\left(s^{2H}+t^{2H}-(t-s)^{2H}\right)\end{aligned}$$

Thus, by the definition of Fractional Brownian Motion [83] the left- and right-hand sides of (8.24) have the same finite-dimensional distributions. □

Theorem 8.4.2. *For* $1 < \alpha_{min} < 2$ *and* $0 < 4 - \alpha_{min} - p < 1$

$$\mathcal{L}\lim_{T\to\infty}\mathcal{L}\lim_{n\to\infty}\frac{\int_0^{Tt}\sum_{i=1}^{N_n(u)} X_i(u)\,du}{\sqrt{\lambda_n L_V(t) L_r(T)}T^H} = 2cZ_{0.5}(t) \tag{8.26}$$

where

$$c = \int_0^\infty r(x)\bar{H}_I(x)\,dx$$

and $Z_{0.5}(t)$ *is standard Brownian Motion.*

Proof. The proof is similar to that for Theorem 8.24 but shows, for fixed $t \geq 0$

$$Y_{Tt} \xrightarrow{D} Y \text{ as } T \to \infty \text{ where } Y \sim \mathrm{N}(0, 2ct) \tag{8.27}$$

and

$$\begin{aligned}\mathrm{Cov}[Y_{Ts}, Y_{Tt}] &= \frac{1}{2}\frac{(V(Ts) + V(Ts) - C(Ts, Tt))}{T}\\ &\to \sigma^2\left(s + t - (t - s)\right) \text{ as } T \to \infty\\ &= 0\end{aligned}$$

and so the left- and right-hand sides of (8.26) have the same finite-dimensional distributions. □

One could also consider a similar result for case 1 in which $2 < 4 - \alpha_{min} - p < 3$. This would appear to correspond to a Hurst parameter H where $1 < H < 1.5$. However these values for H are impossible for a Gaussian self-similar process with stationary increments [83, p. 316]. Exact characterization of the limiting process in this case is left for future work.

Chapter 9

Conclusions and Future Work

This thesis presents results inspired by current research activity in self-similar modelling of packet or byte level data network traffic. The observed presence of long-range dependence and power-tails has led to a need for simulation techniques and traffic models that incorporate one or both of these properties, and statistical estimators with known properties in this setting.

Fractional Gaussian Noise is a canonical long-range dependent Gaussian sequence for which simulation and statistical estimation techniques are known. A fast, approximate simulation technique is improved and shown to produce data with have properties more like Fractional Gaussian Noise. The improvement is demonstrated with numerous simulations using several known statistical estimators.

Two estimators for the variance of Fractional Gaussian Noise are proposed. One is based on Whittle's approximate Maximum Likelihood Estimator technique which provides an estimator of scale, but not the variance. The estimator is shown to be consistent and its limiting distribution is found. Another consistent estimator, in which the sample variance is rescaled, is also proposed. This estimator requires an estimate of the Hurst parameter H of the Fractional Gaussian Noise, and Whittle's estimate is used for that. Small bias and variance of both estimators are demonstrated in numerous simulations.

An extension of a structural model for the superposition of processes with strictly alternating on and off times (possibly heavy-tailed) is considered. In the extended model, the number of processes is allowed to be random. Both convergence of the finite-dimensional distributions and weak convergence of probability measures to a scale mixture Fractional Brownian Motion are proven for the centered and scaled process. In a special case, the limiting process is shown to be sub-Fractional Brownian Motion. Simulation results are included which demonstrate the effect of a random number of sources on the aggregate traffic stream.

A "Many-Process Scale Mixture" traffic model for the aggregate of data network traffic based on a collection of processes

$$W_i(t) = U_i t + \sqrt{\gamma U_i} Z_{H,i}(t)$$

is proposed for disjoint time intervals. Here $\{U_i,\, i = 1, 2 \ldots, R\}$ is an i.i.d. sequence of positive random variables that capture the randomness in the number of sources, independent of $\{Z_{H,i}(t),\, t \geq 0\}$ which are independent copies of Fractional Brownian Motion. The model is intended to capture the additional randomness in the aggregate of identical on-off sources whose number is large, fixed within a time interval but random from one interval to the next. The independence assumptions would be valid if the times between intervals were sufficiently large, for example.

In the "Many-Process Scale Mixture" traffic model, realizations from each time interval will have different realizations of U_i, and so the mean and variance of the traffic stream will change with the number of sources. However, the ratio of variance to mean, γ, is shown to be a quantity of interest because it is unaffected by the number of sources. Two estimators for γ are proposed. Both estimators are shown to be consistent and their rates of convergence are found.

The proposed estimators assume that the centered data is from Fractional Gaussian Noise. With real world data, robustness of the estimators of γ will depend on robustness of the variance estimators used. Using the estimators on real network data is left for future consideration. Testing whether observed data from a network exhibits the property of constant γ over a number of disjoint time intervals is also left for future work. The "One Process Scale Mixture Model" was also proposed. Estimators for γ in this case will take a different form and are left for future work.

A variety of queueing results for the queue length of a first-in first-out queue with input traffic

$$W(t) = Ut + \sqrt{\gamma U} Z_H(t)$$

are presented. Particular attention is paid to the situation in which the server rate scales with U so the utilization is constant. Results are shown in which the distribution of U plays no role if $H = 0.5$ but does play a role in the case of long-range dependence $(0.5 < H < 1)$.

A lower bound is found for the complementary distribution function of the queue length when $\mathrm{P}(U = m) \neq 1$, $\mathrm{E}[U] = m$. The bound compares the function under this assumption with the traditional setting in which U is the degenerate constant m. With this result it would now be interesting to know how much larger the tail probability is in this case. This is left for future work.

Asymptotics for the queue length are also considered, and an asymptotic form for the tail of the queue length distribution function is found in terms of an expected value of a function of U. While examples are given for three distributions of U, more examples from common distributions would be interesting.

A structural traffic model with strictly alternating on-off processes that incorporates random lifetimes whose distributions have regularly varying tails, and Poisson arrivals for the on-off processes, is introduced. With both arrivals and lifetimes, the number of processes varies randomly with time. Limiting results for the finite-dimensional distributions are found and are similar to those obtained when the on-off processes do not have lifetimes. The key difference is that the degree of self-similarity depends not only on the distribution of on and off times, but also on the distribution of lifetimes.

The limit results for this model do not include all possible values for the distribution parameters. Developing limit results for boundary and exceptional cases is a priority. It would be interesting to know if the limiting process in all outstanding cases is Fractional Brownian Motion with some Hurst parameter that can be determined.

Aggregate sample paths of alternating on-off processes with lifetimes have not been simulated here. With the aggregate of simulated on-off processes, standard techniques can be used to estimate the Hurst parameter. Those same techniques can be applied to estimate the Hurst parameter when the individual on-off processes have lifetimes. Demonstrating both a qualitative and a quantitative effect on the Hurst parameter by the lifetime distribution would be interesting.

A further extension of the on-off model with lifetimes may be possible by replacing the Poisson process by a more general increasing process. In this case, a non-Gaussian limiting process, like the scale mixture models proposed here, may be possible. This is also left for future work.

Appendix A

Sample Variance of Fractional Gaussian Noise

Let X_i be identically distributed $N(0,1)$ random variables. For $X_1, \ldots, X_n$ the sample variance is defined to be

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{A.1}$$

$$= \frac{1}{n-1}\left[\sum_{1}^{n} X_i^2 - \frac{1}{n}(\sum_{1}^{n} X_i)^2\right] \tag{A.2}$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean.

If the X_i are independent then S^2 is an unbiased estimator of σ^2. In other words $\mathrm{E}S^2 = \sigma^2$. On the other hand, if the X_i are from Fractional Gaussian Noise the sample variance is biased.

Theorem A.0.3. *Let $X_1, \ldots, X_n$ be from a Fractional Gaussian Noise with Hurst parameter H, so that they have autocovariance function*

$$f(k) = \frac{\sigma^2}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right).$$

Then

$$ES^2 = \sigma^2\left(1 - \frac{n^{2H-1}-1}{n-1}\right)$$

Proof.

$$\begin{aligned}
\mathrm{E}S^2 &= \mathrm{E}\left\{\frac{1}{n-1}\left[\sum_1^n X_i^2 - \frac{1}{n}(\sum_1^n X_i)^2\right]\right\} \\
&= \frac{1}{n-1}\left[n\sigma^2 - \frac{1}{n}\mathrm{E}\left(\sum_1^n X_i\right)^2\right]
\end{aligned}$$

But

$$\begin{aligned}
&\frac{1}{n}\mathrm{E}\left(\sum_1^n X_i\right)^2 \\
&= \frac{1}{n}\left\{n\sigma^2 + 2\left[(n-1)f(1) + (n-2)f(2) + \cdots + (n-k)f(k) + \cdots + f(n-1)\right]\right\} \\
&= \sigma^2 + \sigma^2\left(\frac{n^{2H} - n}{n}\right).
\end{aligned}$$

so that

$$\begin{aligned}
\mathrm{E}S^2 &= \frac{1}{n-1}\left[n\sigma^2 - \sigma^2 - \sigma^2\left(\frac{n^{2H} - n}{n}\right)\right] \\
&= \sigma^2\left(1 - \frac{n^{2H-1} - 1}{n-1}\right)
\end{aligned}$$

□

A quick scan of the table below shows for 100 observations and H=0.75 (moderately strong dependence) the sample variance is only 91% of the true value.

	Number of Observations(n)			
H	100	1000	10 000	100 000
0.5	1.000	1.000	1.000	1.000
0.75	0.909	0.969	0.990	0.997
0.9	0.608	0.750	0.842	0.900

Table A.1: Sample Variance of Standard Fractional Gaussian Noise

Sample Variance for Standard Fractional Gaussian Noise. Note that for the i.i.d. case (H=0.5) the sample variance is always 1.

Appendix B

Proof of Theorem 7.4.3

B.1 Proof

Proof. Let $Z_1^* = \sup_{t\in[0,1]} Z_H(t)$ and $c_2 = \frac{H(2H-1)(2-2H)}{\beta(H-\frac{1}{2},2-2H)}$. As in (2.13), for $H \in (0.5, 1)$

$$\bar{F}_Q(x) = \mathrm{P}(Q^{(0,1,1)} > x) = \mathrm{E}_{Z_1^*}\left[\exp\left(-\left(\frac{x}{Z_1^*}\right)^{(1-H)/H} Z_1^* - \frac{c_2^2}{2}\left(\frac{x}{Z_1^*}\right)^{(2-2H)/H}\right)\right].$$

By Corollary 7.3.3

$$\begin{aligned}
\mathrm{P}(Q^{(U,\gamma U,\frac{U}{\rho})} > \ x) &= \int_0^\infty \bar{F}_Q\left(D_\gamma\, u^{\frac{2H-1}{2-2H}}\, x\right) f_U(u)\, du \\
&= \int_0^\infty \mathrm{E}_{Z_1^*}\left[\exp\left(-D_\gamma^{\frac{1-H}{H}} u^{\frac{2H-1}{2H}} \left(\frac{x}{Z_1^*}\right)^{\frac{1-H}{H}} Z_1^*\right) \times\right. \\
&\qquad \left.\times \exp\left(-\frac{c_2^2}{2} D_\gamma^{\frac{2(1-H)}{H}} u^{\frac{2H-1}{H}} \left(\frac{x}{Z_1^*}\right)^{\frac{2-2H}{H}}\right)\right] f_U(u)\, du
\end{aligned}$$

where

$$D_\gamma = \frac{(\frac{1}{\rho}-1)^{H/(1-H)}}{\gamma^{1/2(1-H)}} \text{ and } \bar{F}_Q(x) = \mathrm{P}(Q^{(0,1,1)} > x).$$

Note that by definition of γ and ρ, $D_\gamma > 0$. Now, let

$$g(z) = D^{\frac{1-H}{H}} \frac{x}{z}^{\frac{1-H}{H}}$$

and

$$h(u,z) = \begin{cases} \exp\left(-zg(z)u^{\frac{2H-1}{2H}} - \frac{c_2^2}{2} g^2(z)\left(u^{\frac{2H-1}{2H}}\right)^2\right), & u \geq 0,\ z > 0 \\ 0, & u \geq 0,\ z = 0. \end{cases}$$

Then, in terms of h and expectations

$$\mathrm{P}(Q^{(U,\gamma U,\frac{U}{\rho})} > x) = \mathrm{E}_U \mathrm{E}_{Z_1^*} h(U, Z_1^*) \tag{B.1}$$

Lemma B.1.1. *For any $z_0 > 0$, $h(u, z_0)$ is strictly convex in u.*

Proof. See the Maple calculations in Appendix B.2 that show $\frac{d^2\, h(u,z_0)}{d\,u^2} > 0$. □

Lemma B.1.2. *For any $x > 0$ and $0 < \rho < 1$*

$$E_U E_{Z_1^*} h(U, Z_1^*) = E_{Z_1^*} E_U h(U, Z_1^*)$$

Proof. Because of the exponential function, by definition $h(u, z)$ is non-negative. Since

$$\int_0^\infty h(u,z) f_{Z_1^*}(z)\, dz = \mathrm{P}\left(Q^{(u,\gamma u,\frac{u}{\rho})} > x\right)$$

we know $0 \leq \int_0^\infty h(u,z) f_{Z_1^*}(z)\, dz \leq 1$ and therefore

$$\int_0^\infty \left(\int_0^\infty h(u,z) f_{Z_1^*}(z)\, dz\right) f_U(u)\, du$$

is finite. By Tonelli's theorem [66, p. 293] and then Fubini's theorem [66, p. 294]

$$\begin{aligned}
\mathrm{E}_U \mathrm{E}_{Z_1^*} h(U, Z_1^*) &= \int_0^\infty \left(\int_0^\infty h(u,z) f_{Z_1^*}(z)\, dz\right) f_U(u)\, du \\
&= \int_0^\infty \left(\int_0^\infty h(u,z) f_{Z_1^*}(z)\right) f_U(u)\, du\, dz \\
&= \mathrm{E}_{Z_1^*} \mathrm{E}_U h(U, Z_1^*)
\end{aligned}$$

□

Since $P(Z_1^* = 0)$ is not necessarily zero we must take some care because $g(0)$ is undefined. Define

$$S_1 = \mathrm{E}_{Z_1^*}[E_U\, h(U, Z_1^*) | Z_1^* > 0] \mathrm{P}(Z_1^* > 0) \text{ and } S_2 = \lim_{b \to 0^+} \mathrm{E}[h(U, b)] \mathrm{P}(Z_1^* = b).$$

Then, by Lemma B.1.2 and conditioning on Z_1^*

$$\begin{aligned}
\mathrm{E}_U \mathrm{E}_{Z_1^*} h(U, Z_1^*) &= \mathrm{E}_{Z_1^*} \mathrm{E}_U h(U, Z_1^*) \\
&= S_1 + S_2
\end{aligned}$$

By Lemma B.1.1 and Jensen's inequality, for any $z_0 > 0$, $\mathrm{E}_U[h(U, z_0)] \geq h(m, z_0)$ since $\mathrm{E}_U[U] = m$. Thus

$$S_1 \geq \mathrm{E}_{Z_1^*} h(m, Z_1^*)|Z_1^* > 0]\mathrm{P}(Z_1^* > 0) \tag{B.2}$$

The second term, S_2, must considered in a different manner. For the long-range dependent case $H \in (0.5, 1)$, $\frac{2H-1}{H} > 0$ and $\frac{2(H-1)}{H} > 0$. Since $D_\gamma > 0$ and $x > 0$, $\lim_{z \to 0^+}(-zg(z)) = -\infty$ and $\lim_{z \to 0^+} -g^2(z) = -\infty$. Thus, for any fixed $u = u_0$,

$$\lim_{z \to 0^+} h(u_0, z) = 0 = h(u_0, 0) \tag{B.3}$$

. By properties of the exponential function, for any fixed $u \geq 0$ and $z \geq 0$, $0 \leq h(u, z) \leq 1$. By the Dominated Convergence Theorem

$$\lim_{b \to 0^+} \mathrm{E}_U[h(U, b)] = \mathrm{E}_U[\lim_{b \to 0^+} h(U, b)] = \mathrm{E}_U[h(U, 0)] = 0$$

and therefore

$$S_2 = \lim_{b \to 0^+} \mathrm{E}[h(U, b)]\mathrm{P}(Z_1^* = b) = \lim_{b \to 0^+} \mathrm{E}[h(U, b)] \lim_{b \to 0^+} \mathrm{P}(Z_1^* = b) = 0. \tag{B.4}$$

Combining (B.1), (B.2), (B.3) and (B.4)

$$\begin{aligned}
\mathrm{P}(Q^{(U, \gamma U, \frac{U}{\rho})} > x) &= S_1 + S_2 \\
&\geq \mathrm{E}_{Z_1^*}[E_U\, h(m, Z_1^*)|Z_1^* > 0]\mathrm{P}(Z_1^* > 0) + \lim_{b \to 0^+} \mathrm{E}[h(m, b)]\mathrm{P}(Z_1^* = b) \\
&= \mathrm{P}(Q^{(m, \gamma m, \frac{m}{\rho})} > x).
\end{aligned}$$

□

B.2 Maple Output- Concavity of $h(u, z)$

Check the concavity of h(u,z)

```
> assume(H>0.5,H<1);
```

In this context a=z*g(z) and b=(c_2*g(z))^2/2, and both are positive for z>0.

```
> assume(a>0,b>0);
> h:=(u,H)->exp(-a*u^(1-(1/(2*H)))-b*(u^(1-1/(2*H)))^2);
```

$$h := (u,\, H) \to e^{(-a\, u^{(1-1/2\, \frac{1}{H})} - b\, (u^{(1-1/2\, \frac{1}{H})})^2)}$$

```
> diff(h(u,H),u);
```

$$\left(-\frac{a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})}\,(1-\frac{1}{2\,H\tilde{}})}{u} - \frac{2\,b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2\,(1-\frac{1}{2\,H\tilde{}})}{u}\right)$$
$$e^{(-a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})} - b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2)}$$

```
> diff(%,u);
```

$$\left(-\frac{a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})}\,(1-\frac{1}{2\,H\tilde{}})^2}{u^2} + \frac{a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})}\,(1-\frac{1}{2\,H\tilde{}})}{u^2}\right.$$
$$\left.-\frac{4\,b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2\,(1-\frac{1}{2\,H\tilde{}})^2}{u^2} + \frac{2\,b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2\,(1-\frac{1}{2\,H\tilde{}})}{u^2}\right)$$
$$e^{(-a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})} - b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2)} +$$
$$\left(-\frac{a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})}\,(1-\frac{1}{2\,H\tilde{}})}{u} - \frac{2\,b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2\,(1-\frac{1}{2\,H\tilde{}})}{u}\right)^2$$
$$e^{(-a\tilde{}\; u^{(1-\frac{1}{2\,H\tilde{}})} - b\tilde{}\;(u^{(1-\frac{1}{2\,H\tilde{}})})^2)}$$

```
> factor(%);
```

$$\frac{1}{4} u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\,(2\,H\tilde{}-1)\, e^{(-u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\,(b\tilde{}\; u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})} + a\tilde{}))} (a\tilde{} - 4\,b\tilde{}\; u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\, H\tilde{}$$
$$+ 4\,b\tilde{}\; u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})} + 2\,u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\, H\tilde{}\; a\tilde{}^2 + 8\,H\tilde{}\; a\tilde{}\; b\tilde{}\;(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^2$$
$$+ 8\,H\tilde{}\; b\tilde{}^2\,(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^3 - u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\, a\tilde{}^2 - 4\,a\tilde{}\; b\tilde{}\;(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^2$$
$$- 4\,b\tilde{}^2\,(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^3) \,/ (H\tilde{}^2\, u^2)$$

All terms except the long one in big brackets are positive. What about the long one?

```
> longterm:=a-4*b*u^(1/2*(2*H-1)/H)*H+4*b*u^(1/2*(2*H-1)/H)+
  2*u^(1/2*(2*H-1)/H)*H*a^2+8*H*a*b*(u^(1/2*(2*H-1)/H))^2+
  8*H*b^2*(u^(1/2*(2*H-1)/H))^3-u^(1/2*(2*H-1)/H)*a^2-
  4*a*b*(u^(1/2*(2*H-1)/H))^2-4*b^2*(u^(1/2*(2*H-1)/H))^3;
```

$$\begin{aligned} longterm := {} & a\tilde{} - 4\,b\tilde{}\,u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\,H\tilde{} + 4\,b\tilde{}\,u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})} + 2\,u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\,H\tilde{}\,a\tilde{}^2 \\ & + 8\,H\tilde{}\,a\tilde{}\,b\tilde{}\,(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^2 + 8\,H\tilde{}\,b\tilde{}^2\,(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^3 - u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})}\,a\tilde{}^2 \\ & - 4\,a\tilde{}\,b\tilde{}\,(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^2 - 4\,b\tilde{}^2\,(u^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})})^3 \end{aligned}$$

Substitute w in place of u^((2*H-1)/H). Both are non-negative.

```
> subs(u=w^(H/(2*H-1)),longterm);
```

$$\begin{aligned} & a\tilde{} - 4\,b\tilde{}\,\%1\,H\tilde{} + 4\,b\tilde{}\,\%1 + 2\,\%1\,H\tilde{}\,a\tilde{}^2 + 8\,H\tilde{}\,a\tilde{}\,b\tilde{}\,\%1^2 \\ & + 8\,H\tilde{}\,b\tilde{}^2\,\%1^3 - \%1\,a\tilde{}^2 - 4\,a\tilde{}\,b\tilde{}\,\%1^2 - 4\,b\tilde{}^2\,\%1^3 \\ & \%1 := (w^{(\frac{H\tilde{}}{2\,H\tilde{}-1})})^{(\frac{2\,H\tilde{}-1}{2\,H\tilde{}})} \end{aligned}$$

```
> simplify(%,symbolic);
```

$$\begin{aligned} & a\tilde{} - 4\,b\tilde{}\,\sqrt{w}\,H\tilde{} + 4\,b\tilde{}\,\sqrt{w} + 2\,\sqrt{w}\,H\tilde{}\,a\tilde{}^2 + 8\,H\tilde{}\,a\tilde{}\,b\tilde{}\,w \\ & + 8\,H\tilde{}\,b\tilde{}^2\,w^{(3/2)} - \sqrt{w}\,a\tilde{}^2 - 4\,a\tilde{}\,b\tilde{}\,w - 4\,b\tilde{}^2\,w^{(3/2)} \end{aligned}$$

```
> simplify(%);
```

$$\begin{aligned} & (8\,H\tilde{} - 4)\,b\tilde{}^2\,w^{(3/2)} + (8\,H\tilde{}\,a\tilde{} - 4\,a\tilde{})\,b\tilde{}\,w \\ & + ((-4\,H\tilde{} + 4)\,b\tilde{} + 2\,H\tilde{}\,a\tilde{}^2 - a\tilde{}^2)\,\sqrt{w} + a\tilde{} \end{aligned}$$

$$\begin{aligned} & a\tilde{} - 4\,b\tilde{}\,\sqrt{w}\,H\tilde{} + 4\,b\tilde{}\,\sqrt{w} + 2\,\sqrt{w}\,H\tilde{}\,a\tilde{}^2 + 8\,H\tilde{}\,a\tilde{}\,b\tilde{}\,w \\ & + 8\,H\tilde{}\,b\tilde{}^2\,w^{(3/2)} - \sqrt{w}\,a\tilde{}^2 - 4\,a\tilde{}\,b\tilde{}\,w - 4\,b\tilde{}^2\,w^{(3/2)} \end{aligned}$$

```
> collect(%,[w,b]);
```

$$\begin{aligned} & (8\,H\tilde{} - 4)\,b\tilde{}^2\,w^{(3/2)} + (8\,a\tilde{}\,H\tilde{} - 4\,a\tilde{})\,b\tilde{}\,w \\ & + ((-4\,H\tilde{} + 4)\,b\tilde{} + 2\,H\tilde{}\,a\tilde{}^2 - a\tilde{}^2)\,\sqrt{w} + a\tilde{} \end{aligned}$$

Substitute d in place of H. With long range dependence,0<d<1

```
> subs(H=(d+1)/2,%);
```

$$\begin{aligned} & 4\,d\,b\tilde{}^2\,w^{(3/2)} + (8\,a\tilde{}\,(\frac{d}{2} + \frac{1}{2}) - 4\,a\tilde{})\,b\tilde{}\,w \\ & + ((-2\,d + 2)\,b\tilde{} + 2\,(\frac{d}{2} + \frac{1}{2})\,a\tilde{}^2 - a\tilde{}^2)\,\sqrt{w} + a\tilde{} \end{aligned}$$

```
> simplify(%);
```

$$4\,d\,b\tilde{~}^2\,w^{(3/2)} + 4\,a\tilde{~}\,d\,b\tilde{~}\,w - 2\,\sqrt{w}\,b\tilde{~}\,d + 2\,b\tilde{~}\,\sqrt{w} + \sqrt{w}\,a\tilde{~}^2\,d + a\tilde{~}$$

```
> collect(%,[w,b]);
```

$$4\,d\,b\tilde{~}^2\,w^{(3/2)} + 4\,a\tilde{~}\,d\,b\tilde{~}\,w + ((-2\,d + 2)\,b\tilde{~} + a\tilde{~}^2\,d)\,\sqrt{w} + a\tilde{~}$$

All terms are positive. Thus the long term (above) is positive. Thus the second derivative of h is positive. Thus h is concave up.

Appendix C

S-plus Routines

C.1 Revised Paxson FGN Generator

Paxson's S-plus code to approximate sample paths of Fractional Gaussian Noise (FGN) involves the main function `ss.gen.fourier` and three other functions that are called within: `FGN.spectrum`, `FGN.B.est.adjust`, and `FGN.B.est`. Only the first function needs to be revised.

It is often convenient to use an optional shell function that properly normalizes the data so that the marginal distribution is $N(0,1)$. That function, `makestdfgn`, is also shown. It is actually quite simple once the proper scaling is known.

C.1.1 Revised function: `revised.fgn.fourier`

```
"revised.fgn.fourier"<-
function(n, H)
{
#
#July 8, 2001: changed z0 from 0 by David Rolls
#
# Returns a Fourier-generated sample path
```

```
# of a "self similar" process, consisting
# of n points and Hurst parameter H
# (n should be even).
n <- n/2 # Approximate ideal power spectrum.
lambda <- ((1:n) * pi)/n # Adjust for estimating power
# spectrum via periodogram.
f <- FGN.spectrum(lambda, H) # Construct corresponding complex
# numbers with random phase.
f <- f * rexp(n) # Last element should have zero phase.
z <- complex(modulus = sqrt(f), argument = 2 * pi * runif(n))
# Expand z to correspond to a Fourier
# transform of a real-valued signal.
z[n] <- abs(z[n]) # Inverse FFT gives sample path.
# Change the next two lines to incorporate z0 DR
z0 <- as.complex(rnorm(1, mean = 0, sd = (2 * n)^(H - 0.5)))
zprime <- c(z0, z, Conj(rev(z)[-1]))
Re(fft(zprime, inv = T))
}
```

Optional Shell Function `makestdfgn`

```
"makestdfgn"<-
function(n, h)
{
# Computes an FGN sequence of length n (even) with
# Hurst parameter h, normed to variance 1.
#
return(revised.fgn.fourier(n, h)/sqrt(n))
}
```

C.2 Whittle-based Variance Estimator

Beran's S-plus implementation of Whittle's Maximum Likelihood Estimator available for download involves the main function `whittle` and eleven other functions that are called as needed: `Qeta`, `CetaFGN`, `CetaARIMA`, `fspecFGN`, `fspecARIMA`, `per`, `Qmin`, `ffourier.FGN.est`, `FGN.spectrum`, `FGN.B.est.adjust`, and `FGN.B.est`. The last four are the same functions used to simulate the spectrum of Fractional Gaussian Noise in the S-Plus implementation of Paxson's technique. To estimate the variance, changes are needed only in the first two functions listed. Necessary changed are indicated with the initials DR.

C.2.1 Revised function: `whittle`

```
"whittle"<-
   function(xinput, nsub = 1, model = "farima", pp = 1, qq = 1, h = 0.5,
ar = c(0.5), ma = c(0.5), out = T, spec = F)

{
assign("out", out, frame = 1)
nmax <- length(xinput)
startend <- c(1, nmax)
istart <- startend[1]
iend <- startend[2]
nloop <- nsub
assign("n", trunc((iend - istart + 1)/nloop), frame = 1)
nhalfm <- trunc((n - 1)/2)
if(model == "farima")
assign("imodel", 2, frame = 1)
if(model == "fgn")
assign("imodel", 1, frame = 1)
assign("p", 0, frame = 1)
assign("q", 0, frame = 1)
if(imodel == 2) {
assign("p", pp, frame = 1)
assign("q", qq, frame = 1)
}
```

```
else {
assign("p", 0, frame = 1)
assign("q", 0, frame = 1)
}
eta <- c(h)
if(p > 0) {
eta[2:(p + 1)] <- ar
}
if(q > 0) {
eta[(p + 2):(p + q + 1)] <- ma
}
M <- length(eta) #loop
thetavector <- c()
i0 <- istart
flax <- vector("list", nloop) #necessary to make nsub/nloop work.  VT.
for(iloop in (1:nloop)) {
h <- max(0.2, min(h, 0.9))
eta[1] <- h
i1 <- i0 + n - 1
y <- xinput[i0:i1]
#standardize data-not performed DR
vary <- var(y)
# y <- (y - mean(y))/sqrt(var(y))
# y <- (y - mean(y))
#periodogram of the data
if(spec == T) {
assign("yper", per(y)[2:(nhalfm + 1)], w = 1)
}
else {
assign("yper", per(y)[2:(nhalfm + 1)], frame = 1)
}
s <- 2 * (1 - h)
etatry <- eta
# Modified to make nlmin not give incorrect result.  VT
if(imodel == 1) {
result <- nlminb(start = etatry, objective = Qmin,
lower = 0, upper = 0.999)
}
else {
```

```
result <- nlminb(start = etatry, objective = Qmin)
}
eta <- result$parameters
sturno <- result$message
theta1 <- Qeta(eta)$theta1
fspectheta1 <- Qeta(eta)$fspectheta1
theta <- c(theta1, eta)
thetavector <- c(thetavector, theta)
#calculate goodness of fit statistic
Qresult <- Qeta(eta) #output
M <- length(eta)
if(imodel == 1) {
SD <- CetaFGN(eta)
SD <- matrix(SD, ncol = M, nrow = M, byrow = T)/n
}
else {
#  Changed to eliminate crashing in solve.qr in CetaARIMA.  VT
cat("M = ", M, "\n")
if(M > 2) {
for(i in 3:M) {
  for(j in 2:(i - 1)) {
    temp <- eta[i] + eta[j]
    if(abs(temp) < 0.0001) {
      cat(
        "Problem with estimating confidence intervals, parameter ",
        i, "and  parameter ", j,
        "are the same, eliminating.\n")
      eta <- eta[ - i]
      eta <- eta[ - j]
      M <- M - 2
      p <- p - 1
      q <- q - 1
    }
  }
}
}
SD <- CetaARIMA(eta, p, q)
SD <- matrix(SD, ncol = M, nrow = M, byrow = T)/n
}
```

```
Hlow <- eta[1] - 1.96 * sqrt(SD[1, 1])
Hup <- eta[1] + 1.96 * sqrt(SD[1, 1])
if(out == T) {
cat("theta=", theta, fill = T)
cat("H=", eta[1], fill = T)
cat("95%-C.I. for H: [", Hlow, ",", Hup, "]", fill = T)
}
#
#Changing of the signs of the moving average parameters
#in order to respect the sign of the Splus convention
#
if(q > 0) {
eta[(p + 2):(p + q + 1)] <-  - eta[(p + 2):(p + q + 1)]
}
etalow <- c()
etaup <- c()
for(i in (1:length(eta))) {
etalow <- c(etalow, eta[i] - 1.96 * sqrt(SD[i, i]))
etaup <- c(etaup, eta[i] + 1.96 * sqrt(SD[i, i]))
}
if(out == T) {
cat("95%-C.I.:", fill = T)
print(cbind(etalow, etaup), fill = T)
}
if(spec == T) {
cat("periodogram is in yper", fill = T)
assign("fest", Qresult$theta1 * Qresult$fspec, w = 1)
cat("spectral density is in fest", fill = T)
}
flax[[iloop]] <- list() # To make nsub work.  VT.
flax[[iloop]]$parameter <- eta
flax[[iloop]]$sigma2 <- bBb # * var(xinput)
flax[[iloop]]$conv.type <- sturno
# DR changes:
# output fspectheta1, the estimated variance, and two values
# from Beran's Goodness of Fit technique that were calculated
# but not output
flax[[iloop]]$fspectheta1 <- fspectheta1
flax[[iloop]]$estvariance <- (theta1)/fspectheta1
```

```
flax[[iloop]]$BeranGoF <- Qresult$Tn
flax[[iloop]]$pval <- Qresult$pval
#
remove("bBb", frame = 0) #next subseries
i0 <- i0 + n #
#Changing of the signs of the moving average parameters
#
if(q > 0) {
eta[(p + 2):(p + q + 1)] <-  - eta[(p + 2):(p + q + 1)]
}
}
return(flax) #
# Clean up
#
remove("n", frame = 1)
remove("p", frame = 1)
remove("q", frame = 1)
remove("imodel", frame = 1)
remove("out", frame = 1)
if(spec == F) {
remove("yper", frame = 1)
}
}
```

C.2.2 Revised function: `Qeta`

```
"Qeta"<-
function(eta)
{
#       cat("in function Qeta", fill=T)
h <- eta[1] #spectrum at fourier frequencies
if(imodel == 1) {
fspec <- fspecFGN(eta, n)
theta1 <- fspec$theta1
fspectheta1 <- theta1
# Retain fspectheta1 for variance estimate DR 18/12/01
fspec <- fspec$fspec
}
else {
```

```
fspec <- fspecARIMA(eta, p, q, n)
theta1 <- fspec$theta1
fspec <- fspec$fspec
}
yf <- yper/fspec
yfyf <- yf^2
A <- 2 * ((2 * pi)/n) * sum(yfyf)
B <- 2 * ((2 * pi)/n) * sum(yf)
Tn <- A/(B^2)
z <- (sqrt(n) * (pi * Tn - 1))/sqrt(2)
pval <- 1 - pnorm(z)
theta1 <- B/(2 * pi)
fspec <- fspec
Qresult <- list(n = n, h = h, eta = eta, A = A, B = B, Tn = Tn, z = z,
pval = pval, theta1 = theta1, fspectheta1 = fspectheta1, fspec
 = fspec)
# fspectheta1 added to list for variance estimate by DR 18/12/01
drop(Qresult)
}
```

Bibliography

[1] R. J. Adler, R. Feldman, and M. S. Taqqu, eds. *A Practical Guide to Heavy-Tails: Statistical Techniques and Applications*, Birkhäuser, Boston, 1998.

[2] A. T. Andersen and B. Friis Nielsen, "A Markovian Approach to Modeling Packet Traffic with Long-Range Dependence," *IEEE J. on Select. Areas Commun.*, vol. 16, no. 5, pp. 719-732, 1998.

[3] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic Theory of a Data–Handling System," *Bell Syst. Tech. Journal*, vol. 61, pp. 1871-1894, Oct. 1982.

[4] S. Asmussen, *Applied Probability and Queues*, Wiley, New York, 1987.

[5] V. Beneš, *General Stochastic Processes in the Theory of Queues*, Addison-Wesley, Reading, Massachusetts, 1963.

[6] J. Beran, "Statistical Methods for Data with Long-Range Dependence," *Statistical Science*, vol. 7, no. 4, pp. 404-416, 1992.

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

[8] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Toronto, 1968.

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

[10] F. Brichet, J. W. Roberts, A. Simonian, and D. Veitch, "Heavy Traffic Analysis of a Storage Model with Long-Range Dependent On/Off Sources," *Queueing Systems*, vol. 23, pp. 197-215, 1996.

[11] F. Brichet, A. Simonian, L. Massoulié, and D. Veitch, "Heavy Load Queueing Analysis with LRD On/Off Sources," In *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, eds., John Wiley & Sons, New York, pp. 115-141, 2000.

[12] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 1987.

[13] S. Cambanis, C. D. Hardin Jr. and A. Weron, "Ergodic Properties of Stationary Stable Processes," *Stochastic Processes and Their Appl.*, vol. 24, pp. 1-18, 1987.

[14] A. H-D. Cheng, P. Sidauruk, Y. Abousleiman, "Approximate Inversion of the Laplace Transform," *The Mathematica Journal*, vol. 4, no. 2, pp. 76-82, 1994.

[15] C-L. Cheng and J. W. Van Ness, *Statistical Regression with Measurement Error*, Arnold, London, 1999.

[16] D. Cox, "Long-Range Dependence: a Review," in *Statistics: An Appraisal*, H. A. David and H. T. David eds., Iowa State Univ. Press, pp. 55-74, 1984.

[17] D. R. Cox and V. Isham, *Point Processes*, Chapman & Hall, New York, 1980.

[18] M. E. Crovella, M. S. Taqqu and A. Bestavros, "Heavy-Tailed Distributions in the World Wide Web," In *A Practical Guide to Heavy-Tails: Statistical Techniques and Applications*, R. J. Adler, R. Feldman, and M. S. Taqqu, eds., Birkhäuser, Boston, pp. 3-25, 1998.

[19] M. Crovella and A. Bestavros, "Self-Similarity in World Wide Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835-846, Dec. 1997.

[20] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications,* Springer, New York, 1998.

[21] M. H. Degroot. *Probability and Statistics*. Addison-Wesley, Reading, Mass., 1975.

[22] R. L. Dobrushin, "A Lemma on the Limit of the Superpositions of Random Functions," *Russian Math Surveys*, vol. 10, pp. 157-159, 1955 (In Russian).

[23] N. G. Duffield and N. O'Connell, "Large Deviations and Overflow Probabilities for the General Single-Server Queue, With Applications," *Math Proc. Camb. Phil. Soc.*, vol. 118, pp. 363-374, 1995.

[24] A. I. Elwalid and D. Mitra, "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," *IEEE Trans. Communication*, vol.1, pp. 329-343, June 1993.

[25] A. Elwalid, D. Heyman, T. V. Lakshman, and D. Mitra, "Fundamental Bounds and Approximations for ATM Multiplexers with Applications to Video Teleconferencing," *IEEE J. Select. Areas. Commun.*, vol. 13, pp. 1004-1016, Aug. 1995.

[26] B. S. Everitt and D. J. Hand, *Finite Mixture Distributions*, Chapman and Hall, New York, 1981.

[27] A. Erramilli, O. Narayan and W. Willinger, "Experimental Queueing Analysis with Long-Range Dependent Traffic," *IEEE/ACM Trans. Networking*, vol. 4, pp. 209-223, April 1996.

[28] R. J. Gibbens and P. J. Hunt "Effective Bandwidths for the Multi-Type UAS Channel," *Queueing Systems*, vol. 9, pp. 17-28, 1991.

[29] P. W. Glynn and W. Whitt, "Logarithmic Asymptotics for Steady-State Tail Probabilities in a Single-Server Queue," *Journal of Applied Probability*, pp. 131-155, 1994.

[30] B. V. Gnedenko and V. Y. Korolev, *Random Summation: Limit Theorems and Applications*, C.R.C. Press, New York, 1996.

[31] C. M. Goldie and C. Klüppelberg, "Subexponential Distributions," in *A Practical Guide to Heavy-Tails: Statistical Techniques and Applications*, R. J. Adler, R. Feldman, and M. S. Taqqu, eds., Birkhäuser, Boston, pp. 435-459, 1998.

[32] C. W. J. Granger and R. Joyeux, "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *J. of Time Series Analysis*, vol. 1, pp. 15-29, 1980.

[33] M. Greiner, M Jobmann and L. Lipsky, "The Importance of Power-Tail Distributions for Modeling Queueing Systems," *Operations Research*, vol. 47., no. 2, pp. 313-326, 1999.

[34] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes* (2nd ed.), Oxford University Press, London, 1992.

[35] R. Guérin, H. Ahmadi and M. Nagshineh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks," *IEEE J. Select. Areas Commun..*, vol. 9, pp. 968-981, Sept. 1991.

[36] C. C. Heyde and Y. Yang, "On Defining Long-Range Dependence," *Journal of Applied Probability*, vol. 34, pp. 949-944, 1997.

[37] B. Hill, "A Simple General Approach to Inference about the Tail of a Distribution," *Annals of Statistics*, vol. 3, no. 5, pp. 1163-1174, 1975.

[38] J. R. M. Hosking, "Fractional Differencing," *Biometrika*, vol. 68, pp. 165-176, 1981.

[39] J. Hüsler and V. Piterbarg, "Extremes of a Certain Class of Gaussian Processes," *Stochastic Processes and their Applications*, vol. 83, pp. 257-271, 1999.

[40] I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Nordhoff, The Netherlands, 1971.

[41] A. Karasaridis and D. Hatzinakos, "Network Heavy Traffic Modeling using Stable Self-Similar Processes," *IEEE Transactions in Communications*, vol. 49. no. 7, pp. 1203-1214, July 2001.

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

[43] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes (2nd Edition)*, Academic Press, New York, 1975.

[44] F. P. Kelly, "Effective Bandwidths at Multi-Class Queues," *Queueing Systems*, vol. 9, pp. 5-16, 1991.

[45] F. P. Kelly, "Notes on Effective Bandwidths," in *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, I. B. Ziedens, eds. Oxford University Press, pp. 141-168, 1996.

[46] G. Kesidis, J. Walrand, and C. Chang, "Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources," *IEEE Trans. Commun.*, vol. 1, pp. 424-428, Aug. 1993.

[47] E. W. Knightly and N. B. Shroff, "Admission Control for Statistical QoS: Theory and Practice," *IEEE Network*, vol. 13, no. 2, pp. 20-29, 1999.

[48] A. N. Kolmogorov, "Weiner's Spiral and Some Other Interesting Curves in Hilbert's Space," *Dokl. Akad. Nauk. USSR*, vol. 26, no. 2, pp. 115-118, 1940 (in Russian).

[49] V. Yu. Korolev, "On Limit Distributions of Randomly Indexed Random Sequences," *Theory Probab. Appl.*, vol. 37, pp. 535-542, 1992.

[50] V. Yu. Korolev, "The Asymptotics of Randomly Indexed Random Sequences: Independent Indices," *J. Soviet Math.*, vol. 59, no. 4, pp. 926-938, 1992.

[51] V. Yu. Korolev, "Convergence of Random Sequences with Independent Random Indices I," *Theory Probab. Appl.*, vol. 39, no. 2, pp. 282-297, 1994.

[52] V. Yu. Korolev, "On the Convergence of Distributions of Random Sums of Independent Random Variables to Stable Laws," *Theory Probab. Appl.*, vol. 42, no. 4, pp. 695-696, 1997.

[53] H. R. Kunsch, "Statistical Aspects of Self-similar Processes," *Proceedings of the First World Congress of the Bernoulli Society*, vol. 1, pp. 67-74, 1987.

[54] S. Ledesma and D. Liu, "Synthesis of Fractional Gaussian Noise Using Linear Approximation for Generating Self-Similar Network Traffic," *Computer Communication Review*, vol. 30, no. 2, pp. 4-17, 2000.

[55] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic," *Proceedings of ACM Sigcomm 1993*, pp. 183-193, 1993.

[56] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," *IEEE/ACM Trans. on Networking*, vol. 2, no. 1, pp. 1-15, 1994.

[57] P. Lévy, "Théorie des erreurs la loi de Gauss et les lois exceptionelles," *Bulletin de la Société Mathématique de France*, vol. 52, pp. 49-85, 1924.

[58] J. B. Levy and M. S. Taqqu, "Renewal Reward Processes with Heavy-Tailed Inter-Renewal Times and Heavy-Tailed Rewards," *Bernoulli*, vol. 6, no. 1, pp. 23-44, 2000.

[59] A. M. Makowski and M. Parulekar, "Buffer Asymptotics for $M/G/\infty$ Input Processes," in *Self-Similar Network Traffic and Performance Evaluation*, K. Park, W. Willinger, eds., John Wiley & Sons, Toronto, pp. 215-248, 2000.

[60] B. B. Mandelbrot, "Long-Run Linearity, Locally Gaussian Processes, H-spectra and Infinite Variances," *International Economic Review*, vol. 10, pp. 82-113, 1969.

[61] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *Siam Review*, vol. 10, pp. 422-437, 1968.

[62] L. Massoulie and A. Simonian, "Large Buffer Asymptotics for the Queue with Fractional Brownian Input," Journal of Applied Probability, vol. 36, pp. 894-906, 1999.

[63] A. Mohammadi, S. Kumar and D. Klymyshyn, "Characterization of Effective Bandwidth as a Metric of Quality of Service for Wired and Wireless ATM Networks," *Proceedings of ICC 1997.*

[64] S. Molnár and A. Vidács, "On Modeling and Shaping Self-Similar ATM Traffic," in *Proceedings of the 15th International Teletraffic Congress*, pp. 1409-1420, 1997.

[65] O. Narayan, "Exact Asymptotic Queue Length Distribution for Fractional Brownian Traffic," *Adv. Perf. Analy.*, vol. 1, pp. 39-63, 1998.

[66] O. A. Nielsen, *An Introducton to Integration and Measure Theory*, John Wiley & Sons, New York, 1997.

[67] I. Norros, "A Storage Model with Self-Similar Input," *Queueing Systems*, vol. 16, pp. 387-396, 1994.

[68] I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," *IEEE J. Select. Areas. Commun.*, vol. 13, pp. 953-962, Aug. 1995.

[69] I. Norros, "Queueing Behavior Under Fractional Brownian Traffic," In *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, eds., John Wiley & Sons, New York, pp. 101-114, 2000.

[70] I. Norros and P. Pruthi, "On the Applicability of Gaussian Traffic Models," In *The 13th Nordic Teletraffic Seminar*, P.J. Emstad, B.E. Helvik, and A.H. Myskja, eds., pp. 37-50, Trondheim, August 1996.

[71] M. Parulekar, *Buffer Engineering for $M/G/\infty$ Traffic Models*, Ph.D. Thesis, Electrical Engineering Department, University of Maryland, College Park, 1999.

[72] V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Trans. on Networking,* vol. 3, no. 3, pp. 226-244, 1995.

[73] V. Paxson. "Fast Approximation of Self-Similar Network Traffic," Technical Report, LBL-36750/UC-405, 1995. See `<http://ita.ee.lbl.gov/html/contrib/fft-fgn.html>`.

[74] V. Paxson. "Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-similar Network Traffic," *Computer Communication Review*, vol. 27, no. 5, pp. 5-18, 1997.

[75] B. L. S. P. Rao, *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Academic Press, San Diego, CA, 1992.

[76] Q. Ren and H. Kobayashi, "Diffusion Approximation Modeling for Markov Modulated Bursty Traffic and Its Applications to Bandwidth Allocation in ATM Networks," *IEEE J. on Select. Areas Commun.*, vol. 16, no. 5, 1998.

[77] H. Robbins, "The Asymptotic Distribution of the Sum of a Random Number of Random Variables," *Bull. Amer. Math. Soc.*, vol 54, pp. 1151-1161, 1948.

[78] D. A. Rolls, *An Investigation of Static and Dynamic Call Admission Control Methods By Simulation*, M.Sc. Project, Queen's University, 1997.

[79] D. A. Rolls, "Improved Fast Approximate Synthesis of Fractional Gaussian Noise," *Proceedings of the Hawaii International Conference on Statistics and Related Fields*, June 2002.

[80] D. A. Rolls and G. K. Takahara, "Random Superposition of On-Off Processes with Heavy-Tailed Inter-renewal Times," *Proceedings of the 2002 Hawaii International Conference on Statistics and Related Fields*, June 2002.

[81] S. Ross, *Probability Models*, 7th ed., Academic Press, New York, 2000.

[82] G. Samorodnitsky, "Extrema of Skewed Stable Processes," *Stochastic Processes and their Applications*, vol. 30, pp. 17-39, 1988.

[83] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, 1994.

[84] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

[85] M. S. Taqqu Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process. *Z. Wahr. verw. Geb.*, vol. 31, pp. 287-302, 1975.

[86] M. S. Taqqu, "A Bibliographical Guide to Self-Similar Processes and Long-Range Dependence," in *Dependence in Probability and Statistics,* E. Eberlein and M. S. Taqqu, eds., pp. 137-165, Birkhäuser, Boston, 1985.

[87] M. S. Taqqu and J. B. Levy, "Using Renewal Processes to Generate Long-Range Dependence and High Variability," *Dependence in Probability and Statistics*, E. Eberlein and M.S. Taqqu, eds., pp. 73-89, Birkhäuser, Boston, 1985.

[88] M. S. Taqqu, W. Willinger and R. Sherman, "Proof of a Fundamental Result in Self-Similar Traffic Modeling," *Computer Communication Review*, vol. 27, pp. 5-23, 1997.

[89] H. Teicher, "On the Mixture of Distributions," *Ann. Math. Statist.*, vol. 31, pp. 55-73, 1960.

[90] V. Teverovsky, S-plus code currently available at URL: `http://math.bu.edu/people/murad/methods/implementations/whittle/whittle.S`.

[91] B. Tsybakov and N. D. Georganas, "Overflow Probability in an ATM Queue with Self-Similar Input Traffic," *Proc. ICC 1997*, pp. 822-826, 1997.

[92] B. Tsybakov and N. D. Georganas, "Self-Similar Processes in Communication Networks," *IEEE Trans. on Information Theory*, vol. 44, no. 5, pp. 1713-1725, 1998.

[93] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, "Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements," *Statistical Science,* vol. 10, no. 1, pp. 67-85, 1995.

[94] W. Willinger, M. S. Taqqu, R. Sherman, and D.V. Wilson, "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," *IEEE/ACM Trans. on Networking* vol. 5, no. 1, pp. 71-86, 1997.

[95] W. Willinger and V. Paxson, "Where Mathematics Meets the Internet," *Notices of the American Mathematical Society,* vol. 45, no. 8, pp. 961-970, September 1998.

[96] Y. Yajima, "A Central Limit Theorem of Fourier Transforms of Strongly Dependent Stationary Processes," *J. Time Series Anal.*, vol. 10, no. 4, pp. 375-383, 1989.

[97] V. M. Zolotarev, *One-Dimensional Stable Distributions*, vol. 65 of "Translations of Mathematical Monographs," American Mathematical Society. Translation of the original 1983 Russian edition, 1986.

Vita

David A. Rolls

EDUCATION

1997 - pres. • Ph.D. (Mathematics),
Queen's University at Kingston, Canada.
Supervisor: Dr. Glen Takahara
In progress.

1997 • Master of Science (Mathematics),
Queen's University at Kingston, Canada.
Supervisor: Dr. Glen Takahara

1993 • Bachelor of Education (Dean's Honour List),
University of Western Ontario, London, Ontario, Canada.

1993 • Bachelor of Mathematics
(Honours Teaching Option – Co-operative With Distinction),
University of Waterloo, Canada.

UNIVERSITY TEACHING EXPERIENCE

Aug. 2002 - pres. Instructor
Department of Mathematics & Statistics
University of North Carolina at Wilmington

Queen's University

Winter 2001 Teaching Fellow: Probability for Elec. and Comp. Engineers
Winter 2000 Teaching Fellow: Calculus 2

Winter 1999 Graduate T.A.: Calculus 2
1997-1998 Graduate T.A.: Integral and Differential Calculus
Jan. - Mar. 1997 Instructor: Calculus 1(J) - Repeat Section

ADDITIONAL TEACHING EXPERIENCE

1994 - 1996 Conversational English Instructor:
ECC Foreign Language Institute, Hiroshima & Osaka, Japan

Sept. - Dec. 1992 High School Mathematics Teacher (Short-term Contract):
Lester B. Pearson High School, Burlington, Ontario, Canada

High School Mathematics Teacher (Intern):
1992 (4 months) York Mills Collegiate Institute, North York, Ontario, Canada
1991 (4 months) West Carleton High School, Dunrobin, Ontario, Canada

- Continued -

CONTRIBUTED TALKS

D. A. Rolls, "Improved Fast Approximate Synthesis of Fractional Gaussian Noise," *Proceedings of the Hawaii International Conference on Statistics and Related Fields*, Honolulu, Hawaii, June 2002.

D. A. Rolls and G. K. Takahara,"Random Superposition of On-Off Processes with Heavy-tailed Inter-renewal times," *Proceedings of the Hawaii International Conference on Statistics and Related Fields*, Honolulu, Hawaii, June 2002.

D. A. Rolls and G. K. Takahara,"Random Superposition of On-Off Processes with Heavy-Tailed Inter-Renewal Times (Preliminary Report)," MAA-AMS Joint Meeting, San Diego, January 2002.

D. A. Rolls and G. K. Takahara,"Random Aggregate of On-Off Processes with Heavy-Tailed Inter-Renewal Times," 11th INFORMS Applied Probability Society Conference, New York, July 2001.

D. A. Rolls,"Random Superposition of On/Off Processes," Mathematics Graduate Student Seminar, Queen's University, March 2001.

D. A. Rolls,"Traffic Models and Queuing," Mathematics Graduate Student Seminar, Queen's University, February 2000.

D. A. Rolls,"Static and Dynamic Call Admission Control," Mathematics Graduate Student Seminar, Queen's University, October 1997.

PUBLICATIONS

Reviewed Conference Papers:

D. A. Rolls, "Improved Fast Approximate Synthesis of Fractional Gaussian Noise," *Proceedings of the Hawaii International Conference on Statistics and Related Fields*, Honolulu, Hawaii, June 2002.

D. A. Rolls and G. K. Takahara,"Random Superposition of On-Off Processes with Heavy-tailed Inter-renewal times," *Proceedings of the Hawaii International Conference on Statistics and Related Fields*, Honolulu, Hawaii, June 2002.

Unpublished Work:

D. A. Rolls, "An Investigation of Static and Dynamic Call Admission Control Methods By Simulation," Unpublished Master's Project, Queens University, 1997.

- Continued -

CONFERENCES ATTENDED

- MITACS 2nd Annual General Meeting, Montreal, Canada, May 2001
- MITACS 1st Annual General Meeting, Toronto, Canada, June 2000
- 1999 IEEE International Conference on Communications, Vancouver, Canada
- 1998 Canadian Mathematical Society Winter Meeting, Kingston, Canada
- International Conference on Stochastic Models, Ottawa, Canada, 1998
- 1997 IEEE International Conference on Communications, Montreal, Canada

AWARDS

1996 - 2001	Queen's University Graduate Award
1998	Fields Institute Graduate Student Fellowship
1988	Ontario Scholar
1987	Suncor Award for Scientific Excellence

AFFILIATIONS

American Mathematical Society
IEEE Communications Society
Expired member of Ontario College of Teachers