Statistical Estimation of Effective Bandwidth

by

Peter Rabinovitch, B.A.
Concordia University

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master of Science in Information and Systems Sciences

Carleton University
Ottawa, Ontario
May 2, 2000

0-612-52403-5

The undersigned hereby recommend to

The Faculty of Graduate Studies and Research

acceptance of the thesis

STATISTICAL ESTIMATION OF EFFECTIVE BANDWIDTH

submitted by Peter Rabinovitch

in partial fulfillment of the requirements

for the degree of Master of Science

Thesis Supervisor

Director, School of Mathematics and Statistics

Carleton University
2000

Abstract

Effective bandwidth, a transformation of the large deviations rate function is presented. A survey of common teletraffic models follows. Confidence intervals for the effective bandwidth of a wide variety of traffic streams are developed, including Poisson streams, fractional Brownian motion streams, heavy-tailed streams and the Bellcore data set. Methods used include the moving blocks bootstrap, surrogate data, and the Dembo estimator.

Acknowledgements

Even though this thesis has only one author listed on the front page, many people contributed to it in many ways.

Ingemar Kaj taught a course on Stochastic Models of Broadband Networks in the Fall of '97. It is fair to say that this course sparked my interest in networking as applied mathematics, and the spark was largely due to the way Ingemar taught the course...a beautiful blend of math, probability, simulation, engineering and the freedom to pursue areas of individual interest.

Over a beer a few months later, Mike Devetsikiotis first suggested that statistics may be able to help in estimating effective bandwidth. This thesis is my attempt to prove him right.

Amit Bose has been the perfect thesis advisor, he has let me work in my own style, but gently brought me back to the right path. Any clarity and precision in this thesis is due to his suggestions and insistence. Any muddiness remains my own.

Most of all, I must thank Terri, Rebecca and Rachael who taught me to play, and never complained when "Daddy is going to the basement for a few hours." Without their love, patience and support, this thesis would consist only of blank pages.

Contents

Chapter 1

Introduction

The central problem of measurement based admission control in high speed communication networks is to determine whether or not a new connection can be admitted with a requested quality of service, without disturbing the existing traffic that the network is already servicing. One measure of quality of service is loss probability, which can be analyzed using the theory of large deviations, in particular by a transformation of the large deviations rate function called the "effective bandwidth".

Of course, one can estimate loss probability by explicitly calculating it for traffic models that are sufficiently simple. The problem is that for many traffic models the calculations are extremely difficult. Another option would be to simulate the system, but as in modern telecommunication networks the loss probability is typically of order 10^{-8}, simulation without techniques like importance sampling would take an extremely long time.

There have been many theoretical studies demonstrating the utility of the effective bandwidth concept, but little work has been done on the estimation of effective bandwidth of real traffic streams. Even less work has been done on the statistical estimation of effective bandwidth of long range dependent traffic, primarily due to the difficulty of obtaining sufficiently many independent, identically distributed samples of the long-range dependent process. If the autocorrelation is significant at lags the same order of magnitude as the length of the series itself, it is not possible to segment the series into independent parts, which is necessary for conventional statistical analysis. In this thesis we propose and evaluate a solution to this problem.

The question we address in this thesis is:

How can we construct confidence intervals for the effective bandwidth of a variety of traffic streams and how well do these confidence intervals perform in misspecified cases?

A few highlights of this thesis are as follows.

In Chapter 2, after defining the effective bandwidth as a function of two variables, we explain why we need to analyze the effective bandwidth only at the so-called "critical point" and not over the whole two dimensional surface, which makes statistical analysis much easier. Calculation of the critical point becomes one of the common threads throughout this thesis, as many of the estimators are sensitive to the value of the critical point. We prove several properties of effective bandwidth and explain and prove one part of the connection between the effective bandwidth and loss probability. We also briefly describe the measured buffer occupancy method for estimating loss probability, which is an alternative to effective bandwidth, but requires experience and judgement of the network analyst.

Chapter 3 discusses a variety of traffic models, and describe their characteristics, such as long range dependence, self-similarity and heavy tails. More importantly, we explain the connection between these three characteristics.

In Chapter 4 we develop methods for determining confidence intervals for the effective bandwidth of these traffic types. We begin by developing estimators analytically, but quickly run into complications. We then present the Dembo estimator of effective bandwidth and develop some of its properties. Of particular interest is a quick introduction to the bootstrap and moving blocks bootstrap. These two methods are well known in the statistical community, but not as well known in network engineering circles. We also discuss surrogate data methods which originated in nonlinear time series analysis. These methods have few theoretical results, but evidence presented in this chapter indicates that they may be a very useful tool in statistical analysis of long range dependent data. We use these bootstrap and surrogate data methods with the Dembo estimator to develop confidence intervals for the effective bandwidth of long range dependent and heavy-tailed traffic streams.

In Chapter 5, we compare all these estimators in a "shoot-out" to see how well the different estimators work on a variety of real and synthetic traffic streams, including the famous Bellcore data set. The main result here is that the different estimators perform very differently on the differing streams, and that there is not a clear "winner."

The main points of this thesis are:

1. The methods of Surrogate Data and Amplitude Adjusted Surrogate Data appear to work for a variety of traffic models, including long range dependent streams.

2. Although no recommendation can yet be made to teletraffic engineers on which estimator to use when, there are many promising avenues of research that may yield such recommendations.

1.1 Notation

In this section, we note several notations used throughout this thesis.

Capital letters will always denote random variables, corresponding lowercase letters will denote their observed values.

$X(n)$ will always denote the cumulative arrivals, after time 0 up to and including time n, of a traffic stream. Thus we always refer to the half-open intervals $(0, n]$.

$Y(n)$ will always denote the increment of a traffic stream: $Y(n) := X(n) - X(n-1)$.

$\mathbb{P}$ denotes a probability.

$\mathbb{E}$ denotes expectation.

$\mathbb{V}$ denotes variance.

IID denotes independent, identically distributed.

MLE denotes maximum likelihood estimator.

We often abbreviate $x_1, x_2, ..., x_n$ by $\mathbf{x}$.

$X \stackrel{\mathcal{D}}{=} Y$ denotes that $X = Y$ in distribution.

$X \stackrel{\mathcal{D}}{\rightarrow} Y$ denotes that X converges to Y in distribution.

$X \stackrel{P}{\rightarrow} Y$ denotes that X converges to Y in probability.

$X \stackrel{a.s.}{\rightarrow} Y$ denotes that X converges to Y almost surely (with probability one).

$N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2.

$\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Chapter 2

Effective Bandwidth

Effective bandwidth is a measure of a traffic stream used in dimensioning modern communication networks. One of the characteristics of these networks is that many connections are multiplexed, or aggregated, over a common cable. The issue then is to determine the number of connections that can be multiplexed without violating any service levels guaranteed to the customers. A given connection will have a mean rate and a peak rate. If we decide to allow the connection based solely on its peak rate, then clearly there will be wasted bandwidth on the cable, as the connection will likely not send bytes at the peak rate continuously (unless the peak rate is the mean rate). On the other hand, if we base our decision on the mean rate, there may be times where we are unable to provide service to this connection, as it will occasionally send bytes at its peak rate. Thus, we need to make connection admission decisions based on some parameter lying between the mean and peak rate [15, Section 6.1]. Effective bandwidth allows us to do this, as it is a number between these two extremes.

We take as given the work by Kelly in "Notes on Effective Bandwidths" [17], however we modify the notation slightly in order to be consistent with other symbols used later in this thesis. Note also that we work entirely in discrete time, while Kelly's original definition applies to continuous time. Our decision to restrict this study to discrete time is due to the fact that real networks operate in discrete time.

Definition 1 *Let $X(\tau)$ be a process with stationary increments, with X_i representing the amount of traffic that has arrived in the time period i (i.e. the increments). Thus* $X(\tau) = \sum_{i=1}^{\tau} X_i$.

*Then the **effective bandwidth** of X is defined as:*

$$eb_X(\theta,\tau) := \frac{1}{\theta\tau}\log \mathbb{E}\left[e^{\theta\sum_{i=1}^{\tau} X_i}\right]$$

for $0 < \theta, \tau < \infty$, *where* $\theta \in \mathbb{R}$, *and* $\tau \in \mathbb{N}$.

2.1 Properties of Effective Bandwidth

The following properties of effective bandwidth are used later in this thesis.

Proposition 2 *[17, 2.2 (i)] If $X(\tau)$ has independent increments, then $eb_X(\theta,\tau)$ does not depend on τ.*

Proof.

$$\begin{aligned}
eb_X(\theta,\tau) &= \frac{1}{\theta\tau}\log \mathbb{E}\left[e^{\theta\sum_{i=1}^{\tau} X_i}\right] \\
&= \frac{1}{\theta\tau}\log \mathbb{E}\left[\prod_{i=1}^{\tau} e^{\theta X_i}\right] \\
&= \frac{1}{\theta\tau}\log \prod_{i=1}^{\tau} \mathbb{E}\left[e^{\theta X_i}\right], \text{ due to independent increments} \\
&= \frac{1}{\theta\tau}\sum_{i=1}^{\tau}\log \mathbb{E}\left[e^{\theta X_i}\right] \\
&= \frac{\tau}{\theta\tau}\log \mathbb{E}\left[e^{\theta X_i}\right], \text{ due to stationarity of increments} \\
&= \frac{1}{\theta}\log \mathbb{E}\left[e^{\theta X_i}\right] \\
&= eb_X(\theta,1)
\end{aligned}$$

which does not depend on τ. ■

Remark 3 *This will prove useful later, because for traffic streams with independent increments, the dimension of the problem gets effectively reduced, i.e. τ does not play a role in the estimation of $eb_X(\theta,\tau)$.*

Proposition 4 *[17, 2.2 (iii)] Let $n < \infty$. If $X(\tau) = \sum_{j=1}^{n} X^j(\tau)$, where $X^j(\tau)$ represents the j^{th} traffic stream, and the $\left(X^j(\tau)\right)_{j=1}^{n}$ are independent, then*

$$eb_X(\theta,\tau) = \sum_{j=1}^{n} eb_{X^j}(\theta,\tau)$$

where $eb_{X^j}(\theta,\tau)$ is the effective bandwidth of $X^j(\tau)$.

Proof.

$$\begin{aligned} eb_X(\theta,\tau) &= \frac{1}{\theta\tau}\log \mathbb{E}\left[e^{\theta\sum_{j=1}^{n} X^j(\tau)}\right] \\ &= \frac{1}{\theta\tau}\log \mathbb{E}\left[\prod_{j=1}^{n} e^{\theta X^j(\tau)}\right] \\ &= \frac{1}{\theta\tau}\log\left(\prod_{j=1}^{n} \mathbb{E}\left[e^{\theta X^j(\tau)}\right]\right) \\ &= \sum_{j=1}^{n} \frac{1}{\theta\tau}\log \mathbb{E}\left[e^{\theta X^j(\tau)}\right] \\ &= \sum_{j=1}^{n} eb_{X^j}(\theta,\tau) \end{aligned}$$

■

As stated earlier, the effective bandwidth is a number between the mean and peak rates of the traffic source.

Proposition 5 *[17, 2.2 (iv)] For any fixed value of τ, $eb_X(\theta,\tau)$ is increasing in θ, and lies between the mean and peak of the arrival rate measured over an interval of length τ:*

$$\frac{\mathbb{E}X(\tau)}{\tau} \le eb_X(\theta,\tau) \le \frac{\tilde{X}(\tau)}{\tau}$$

where $\tilde{X}(\tau)$ is the (possibly infinite) essential supremum

$$\tilde{X}(\tau) = \sup\left\{x : \mathbb{P}\left\{X(\tau) > x\right\} > 0\right\}.$$

We illustrate this proposition with a simple, but typical example.

Example 6 *Let us consider an independent sequence of Bernoulli random variables as an arrival stream, with*

$$\mathbb{P}[X_i = 1] = p$$
$$\mathbb{P}[X_i = 0] = 1 - p$$

for all $i > 0$.

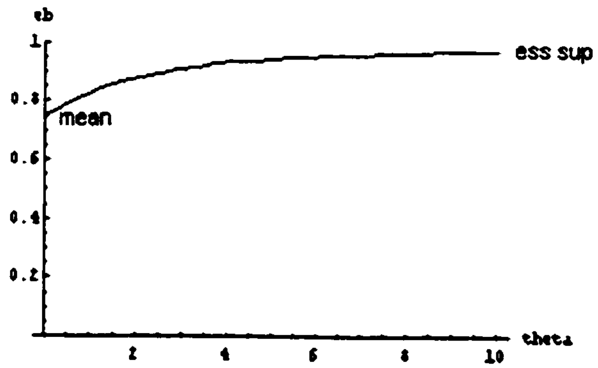
Then

$$\begin{aligned}
eb_X(\theta,\tau) &= \frac{1}{\theta\tau}\log\mathbb{E}[e^{\theta X(\tau)}] \\
&= \frac{1}{\theta\tau}\log\mathbb{E}\left[e^{\theta\sum_{i=1}^{\tau} X_i}\right] \\
&= \frac{1}{\theta\tau}\log\mathbb{E}\left[\prod_{i=1}^{\tau} e^{\theta X_i}\right] \\
&= \frac{1}{\theta\tau}\log\prod_{i=1}^{\tau}\mathbb{E}\left[e^{\theta X_i}\right], \textit{ due to independence} \\
&= \frac{1}{\theta\tau}\sum_{i=1}^{\tau}\log\mathbb{E}\left[e^{\theta X_i}\right] \\
&= \frac{1}{\theta}\log\mathbb{E}\left[e^{\theta X_i}\right], \textit{ due to the } X_i \textit{ being identically distributed} \\
&= \frac{1}{\theta}\log\left(pe^{\theta} + 1 - p\right)
\end{aligned}$$

Note that

$$
\begin{aligned}
eb_X(0,\tau) &= \lim_{\theta\to 0} \frac{1}{\theta} \log\left(pe^{\theta}+1-p\right) \\
&= \lim_{\theta\to 0} \frac{d}{d\theta} \log\left(pe^{\theta}+1-p\right), \text{ by l'Hospital's rule} \\
&= \lim_{\theta\to 0} \frac{pe^{\theta}}{pe^{\theta}+1-p} \\
&= \frac{p}{p+1-p} \\
&= p, \text{ the mean arrival rate of the stream, and} \\
eb_X(\infty,\tau) &= \lim_{\theta\to\infty} \frac{1}{\theta} \log\left(pe^{\theta}+1-p\right) \\
&= \lim_{\theta\to\infty} \frac{\frac{d}{d\theta}pe^{\theta}}{\frac{d}{d\theta}\left(pe^{\theta}+1-p\right)}, \text{ by repeated application of l'Hospital's rule} \\
&= \lim_{\theta\to\infty} \frac{pe^{\theta}}{pe^{\theta}} \\
&= 1, \text{ the essential supremum of the stream.}
\end{aligned}
$$

We illustrate with a graph, with $p = 0.75$.



2.2 On the Use of Effective Bandwidth in Discrete Time

In this section we state the large deviations rate function for overflow in a discrete time queue with service rate C and buffer size B, being fed by an input process X, the result being [40, Section 3.6]:

$$I = \inf_{\tau \geq 0} \sup_{\theta \geq 0} \theta(B + C\tau) - \theta\tau eb_X(\theta, \tau). \tag{2.1}$$

Definition 7 *[30] The optimizing θ^*, τ^* in the above equation are referred to as the* ***critical point*** *of the system.*

The critical point may be calculated explicitly in simple cases via calculus. In more complicated cases, but where the form of the effective bandwidth function is known, it must be estimated numerically. Finally in cases where the form of the effective bandwidth function is not known, as in a real traffic stream, it must be estimated statistically.

At the critical point, the traffic stream behaves (from the point of view of overflow probability) as though it were a constant traffic stream of rate $eb_X(\theta^*, \tau^*)$. Roughly, this says that as the number of traffic sources increases, and the service rate and buffer size increase in proportion, that

$$-\log \mathbb{P}(\text{overflow}) \approx I.$$

It is therefore used in measurement based admission control to decide whether or not a given quality of service (as measured by buffer overflow) can be maintained after the admission of a new stream, given the traffic that is currently being served.

The parameter θ^* describes the degree of multiplexing in the system. A large value of θ^* corresponds to a low degree of multiplexing. This can be understood by considering a link of fixed capacity and a number of connections with a high value of θ^*. As high values of θ^* correspond to an effective bandwidth closer to the essential supremum of the traffic flow (Prop. 5), it follows that a smaller number of connections can be serviced than if the connections had a lower value of θ^*, which corresponds to an effective bandwidth closer to the mean rate of the traffic stream.

Another interpretation of θ^* that relates to importance sampling [39] is that "over the busy period preceding the buffer overflow the amount of work produced by a stream is exponentially tilted, with tilt parameter θ^*" [30, Q1].

The parameter τ^* is the most probable length of the busy period prior to buffer overflow and hence the minimum scale on which the traffic needs to be observed in order to be able to analyze buffer overflow. Large values of τ^* indicates that slow time scales are responsible for buffer overflow, whereas small values of τ^* indicate that fast time scales are responsible for buffer overflow.

Let us now derive formula 2.1. We follow both [10, Theorem 1] and [22, Pg. 36], however we focus on the case of IID arrivals for ease of exposition, and we operate in discrete time, which is the case for ATM traffic. Also, we only derive the upper bound, as this is the one that is more useful in admission control. First, we need the following well known result.

Theorem 8 *Chernoff's Bound*

Let X be a random variable, and assume that for all $\theta \in \mathbb{R}$ that $\mathbb{E}[e^{\theta X}] < \infty$, and define $\psi(\theta) := \log \mathbb{E}[e^{\theta X}]$. Then for $a \in \mathbb{R}$

$$\mathbb{P}[X > a] \leq \exp(-\sup_{\theta \geq 0}(a\theta - \psi(\theta)))$$

Proof. For $\theta \geq 0$, $a \in \mathbb{R}$, $X > a$ implies that $e^{\theta X} > e^{\theta a}$. Therefore

$$\begin{aligned}
\mathbb{P}[X > a] &= \mathbb{P}\left[e^{\theta X} > e^{\theta a}\right] \\
&\leq \frac{\mathbb{E}\left[e^{\theta X}\right]}{e^{\theta a}} \text{ by Markov's Inequality} \\
&= e^{-\theta a} e^{\log \mathbb{E}\left[e^{\theta X}\right]} \\
&= \exp(-(a\theta - \psi(\theta)))
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{P}[X > a] &\leq \inf_{\theta \geq 0} \exp(-(a\theta - \psi(\theta))) \\
&\leq \exp(\inf_{\theta \geq 0}(-(a\theta - \psi(\theta)))), \text{ because exp is monotone increasing} \\
&\leq \exp(-\sup_{\theta \geq 0}(a\theta - \psi(\theta))), \text{ because } \inf(-x_n) = -\sup(x_n)
\end{aligned}$$

■

Proposition 9 *Consider a single server queue operating in discrete time with serving N sources, with service rate c per source and buffer size b per source. Denote the IID (discrete time) arrivals by one source to the queue by X_i, and the cumulative arrivals by one source in $(0, \tau]$ by $X(\tau) := \sum_{i=1}^{\tau} X_i$. Then the probability of buffer overflow is bounded by*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}[Q > Nb] \leq -\{\theta^*(b + c\tau^*) - \theta^*\tau^* eb_X(\theta^*, \tau^*)\}$$

where Q_t denotes the contents of the queue at time t, and Q denotes the contents of the buffer in equilibrium.

Proof. First, observe that buffer overflow at time period 0 implies the existence of a time $-m$ at which the buffer was last empty, and since which at least $(Nb + Ncm)$ cells have arrived, so

$$\mathbb{P}[Q_0 > Nb \text{ in time period m}] = \mathbb{P}\left[\sum_{k=1}^{N} \sum_{i=-m}^{0} X_i > Nb + Ncm\right].$$

Now,

$$\begin{aligned}
\mathbb{P}[Q_0 > Nb \text{ for some } m > 0] &\leq \sum_{m=1}^{\infty} \mathbb{P}[Q_0 > Nb \text{ in time period m}] \\
&\leq \sum_{m=1}^{\infty} \mathbb{P}\left[\sum_{k=1}^{N} \sum_{i=-m}^{0} X_i > Nb + Ncm\right] \\
&\leq \sum_{m=1}^{m_1-1} \mathbb{P}\left[\sum_{k=1}^{N} \sum_{i=-m}^{0} X_i > Nb + Ncm\right] \\
&+ \sum_{m=m_1}^{\infty} \mathbb{P}\left[\sum_{k=1}^{N} \sum_{i=-m}^{0} X_i > Nb + Ncm\right]
\end{aligned}$$

By Chernoff's Bound

$$\mathbb{P}\left[\sum_{k=1}^{N}\sum_{i=-m}^{0} X_i > Nb + Ncm\right] \leq \exp\left(-\sup_{\theta\geq 0}\theta(Nb+Ncm) - \psi(\theta)\right),$$

$$\begin{aligned}
\text{where } \psi(\theta) &= \log\mathbb{E}\left[\exp\left(\theta\sum_{k=1}^{N}\sum_{i=-m}^{0} X_i\right)\right] \\
&= \log\mathbb{E}\left[\exp\left(\theta N\sum_{i=-m}^{0} X_i\right)\right] \\
&= \log\mathbb{E}\left[\left(\exp\left(\theta\sum_{i=-m}^{0} X_i\right)\right)^N\right] \\
&= \log\mathbb{E}\left[\exp\left(\theta\sum_{i=-m}^{0} X_i\right)\right]^N \\
&= N\log\mathbb{E}\left[\exp\left(\theta\sum_{i=-m}^{0} X_i\right)\right]
\end{aligned}$$

We also have that $\sum_{i=-m}^{0} X_i \stackrel{\mathcal{D}}{=} X(m)$, as the X_i are IID, so

$$\begin{aligned}
\psi(\theta) &= N\log\mathbb{E}\left[\exp\left(\theta X(m)\right)\right] \\
&= N\theta meb_X(\theta, m), \text{ so}
\end{aligned}$$

$$\mathbb{P}\left[\sum_{i=-m}^{0} X_i > Nb + Ncm\right] \leq \exp\left(-\sup_{\theta\geq 0}\theta(Nb+Ncm) - N\theta meb_X(\theta,m)\right).$$

We have that

$$\begin{aligned}
\sum_{m=1}^{m_1-1}\mathbb{P}\left[\sum_{i=-m}^{0} X_i > Nb + Ncm\right] &\leq \sum_{m=1}^{m_1-1}\exp\left(-\sup_{\theta\geq 0}\theta(Nb+Ncm) - N\theta meb_X(\theta,m)\right) \\
&\leq (m_1-1)\max_{1\leq m\leq m_1}\exp\left(-\sup_{\theta\geq 0}\theta(Nb+Ncm) - N\theta meb_X(\theta,m)\right) \\
&\leq (m_1-1)\exp\left(-\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(Nb+Ncm) - N\theta meb_X(\theta,m)\right)
\end{aligned}$$

The second sum (from m_1 to ∞) is shown to be negligible under technical conditions in [10, Theorem 1]. Finally,

$$\mathbb{P}[Q > Nb] \leq (m_1 - 1)\exp\left(-\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(Nb + Ncm) - N\theta m eb_X(\theta, m)\right)$$

$$\log\mathbb{P}[Q > Nb] \leq \log\left\{(m_1 - 1)\exp\left(-\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(Nb + Ncm) - N\theta m eb_X(\theta, m)\right)\right\}$$

$$\leq \log(m_1 - 1) - \inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(Nb + Ncm) - N\theta m eb_X(\theta, m).$$

Therefore

$$\frac{1}{N}\log\mathbb{P}[Q > Nb] \leq \frac{1}{N}\log(m_1 - 1) - \frac{1}{N}\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(Nb + Ncm) - N\theta m eb_X(\theta, m)$$

$$\leq \frac{1}{N}\log(m_1 - 1) - \inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(b + cm) - \theta m eb_X(\theta, m)$$

and we have

$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}[Q > Nb] \leq \limsup_{N\to\infty}\frac{1}{N}\log(m_1 - 1) - \limsup_{N\to\infty}\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(b + cm) - \theta m eb_X(\theta, m)$$

$$\leq 0 - \limsup_{N\to\infty}\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(b + cm) - \theta m eb_X(\theta, m)$$

$$\leq -\inf_{1\leq m\leq m_1}\sup_{\theta\geq 0}\theta(b + cm) - \theta m eb_X(\theta, m)$$

because the infimum over a bigger set is smaller, so the negative of the infimum over a bigger set is bigger. Thus, we have

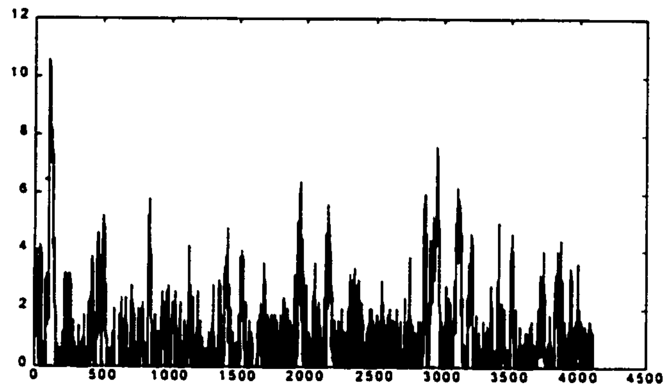
$$\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{P}[Q > Nb] \leq -\left\{\theta^*(b + c\tau^*) - \theta^*\tau^* eb_X(\theta^*, \tau^*)\right\}.$$

■

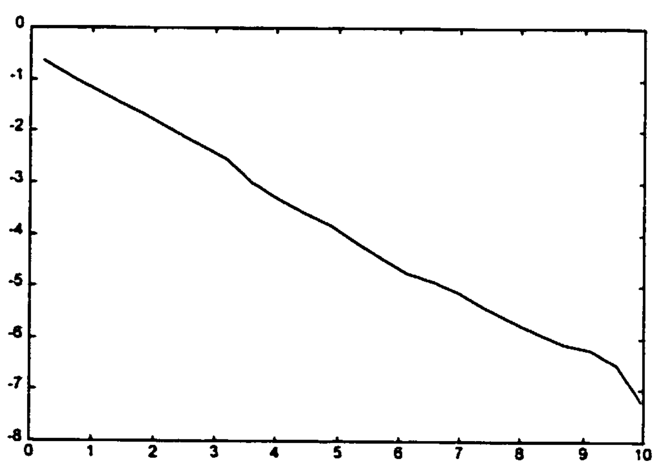
2.3 Graphical Method for Estimating Loss Probabilities Suggested by the Large Deviations Principle

Other than calculating effective bandwidth, one can estimate loss probability by simulating a queue with a fixed service rate and observing the queue length distribution. Then we plot $\log \mathbb{P}[\text{queue length} > B]$ vs. B, and use least squares regression to fit this curve. We obtain a slope δ, and then estimate the loss probability in a queue with buffer size B as $e^{-\delta B}$. This is described in [26, Pg. 3] and [22, Section 6], and we will refer to this as the measured buffer occupancy method.

Example 10 *Let us consider a simulated Poisson arrival stream with rate 0.3 to a single server queue with deterministic service rate 0.4. The buffer size is plotted below.*



Now let us plot $\log \mathbb{P}[\textit{queue length} > B]$ *vs.* B.



If we fit a linear regression to this curve, we get that $\delta = 0.6503$. *Thus the probability of loss in this system, if the buffer size were* 5, *would be approximately* $e^{-\delta B} = 0.0387$.

Note that for long range dependent data (defined later in this thesis), this method has to be modified to

$$\log \mathbb{P}(\text{queue length} > B) \approx -\delta B^{\gamma}$$

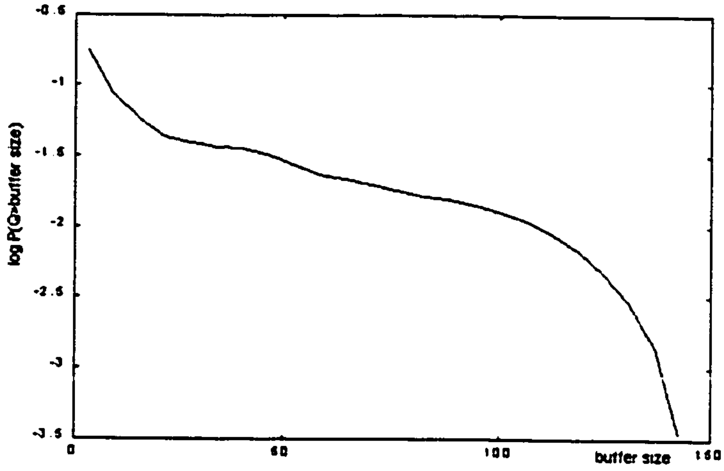
where $\gamma = 2(1 - H)$ [26, Pg. 38]. No similar formula is known for long range dependent, heavy tailed data, so in these cases we will also estimate γ from the data. We can do this by transforming the problem as follows.

$$\begin{aligned} \log \mathbb{P}(\text{queue length} > B) &\approx -\delta B^{\gamma} \\ \log\log \mathbb{P}(\text{queue length} > B) &\approx \log(-\delta B^{\gamma}) \\ &\approx \log(-\delta) + \gamma \log(B) \end{aligned}$$

Thus we can perform a linear regression to determine the parameters δ and γ. Note that $\log\log \mathbb{P}(\text{queue length} > B)$ will be complex numbers, but the estimates δ and γ are real.

In addition, this method also provides a quick, graphical tool to assess whether or not the traffic stream is long range dependent. If $\gamma \approx 1$, then the stream is likely short range dependent. One can also look at the plot of $\log \mathbb{P}[\text{queue length} > B]$ vs. B to see how well the curve is approximated by a straight line for large B.

The problem with this method is best illustrated by the following graph.



This is the graph of log $\mathbb{P}$[queue length > B] vs. B for a Brownian motion stream. The problem is that there appears to be at least three regions where one could fit a straight line, requiring judgement and experience of the modeler.

Chapter 3

Traffic Models

In this chapter we briefly review some of the most commonly used traffic models, as we will be estimating effective bandwidth for these models later in this thesis. We begin with the simplest model, the Poisson process, and then successively generalize to Brownian Motion, On/Off Markov Fluid, fractional Brownian motion, and an α-Stable FARIMA model.

These models exhibit a variety of features observed in various forms of real network traffic, such as independent increments (Poisson), short range dependent increments (On/Off Markov Fluid), long range dependence (fractional Brownian motion), self-similarity (fractional Brownian motion), and heavy tails (α-Stable FARIMA). These last three characteristics are frequently confused, so later in this chapter we explain the relationships between these features.

All models we consider in this thesis have stationary increments.

Definition 11 *A stochastic process $X(\tau)$, where $\tau \in \mathcal{T}$, a time index set, is said to have* ***stationary increments*** *if $\{X(\tau+h) - X(h), \tau \in \mathcal{T}\} \stackrel{\mathcal{D}}{=} \{X(\tau) - X(0), \tau \in \mathcal{T}\}$, for all $h \in \mathcal{T}$. Note that some authors refer to this as "strictly stationary increments".*

This definition says that the increments of the process depend only on the length of the interval over which they are observed, not on the absolute time that the observation begins. This is essential for statistical analysis, for if the distribution of the increments were changing over time, this would need to be modeled too.

3.1 Independent Increment Models

Definition 12 *A stochastic process $X(\tau)$ is said to have* ***independent increments*** *if for any time points $0 = \tau_0 < \tau_1 < \tau_2 \cdots < \tau_n$ the process increments $X(\tau_1) - X(\tau_0), X(\tau_2) - X(\tau_1), \cdots, X(\tau_n) - X(\tau_{n-1})$, are independent random variables.*

We now consider two processes with independent increments, the Poisson process, and Brownian motion. For these two cases, it is possible to analytically derive confidence intervals for the effective bandwidth, and we will do this in Chapter 4.

3.1.1 Poisson Traffic

The Poisson model for telephone traffic has been used successfully since Erlang's work in the early part of this century. The model has proved accurate for telephone conversation traffic, and has many theoretical properties that make it easy to analyze. Thus, any work on teletraffic must encompass Poisson traffic as a special case.

Definition 13 *A* ***Poisson process*** *of rate $\lambda > 0$ is an integer valued stochastic process $\{X(t); t \geq 0\}$ such that*

1. *$X(t)$ has independent increments*

2. *for $s \geq 0$ and $t > 0$,*

$$\mathbb{P}[X(s+t) - X(s) = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \text{ for } k = 0, 1, \ldots$$

3. *$X(0) = 0$.*

Effective Bandwidth of a Poisson Stream

Recall that the moment generating function of a Poisson random variable X with parameter λ is $M_X(\theta) = e^{\lambda(e^\theta - 1)}$. Now, let us calculate the effective bandwidth of a Poisson stream with parameter λ.

Proposition 14 *The effective bandwidth of a Poisson process with parameter λ is*

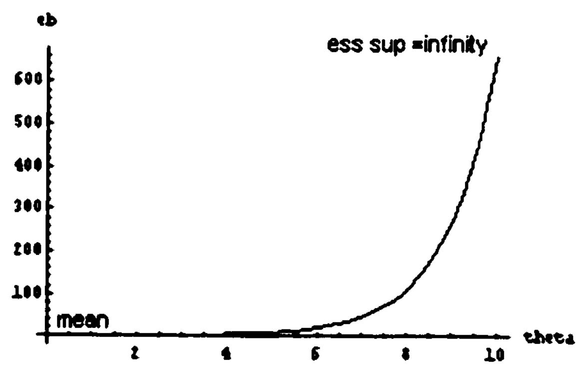
$$eb(\theta,\tau) = \lambda\frac{(e^\theta - 1)}{\theta}$$

Proof. Since $X(\tau)$ is a Poisson process, the number of arrivals in $(0,\tau]$ has a Poisson distribution with parameter $\lambda\tau$. Therefore

$$\begin{aligned} eb(\theta,\tau) &= \frac{1}{\theta\tau}\log \mathbb{E}[e^{\theta X(0,\tau]}] \\ &= \frac{1}{\theta\tau}\log e^{\lambda\tau(e^\theta - 1)} \\ &= \frac{1}{\theta\tau}\lambda\tau(e^\theta - 1) \\ &= \lambda\frac{(e^\theta - 1)}{\theta} \end{aligned} \tag{3.1}$$

■

Note that the effective bandwidth of a Poisson process does not depend on τ, in agreement with Proposition 2, and the fact that a Poisson process has independent increments. A graph of the effective bandwidth of a Poisson stream, as a function of θ looks like (for $\lambda = 0.3$):



Note that the effective bandwidth of a Poisson stream is unbounded, as required by Proposition 5, since the essential supremum of a Poisson random variable is unbounded.

Critical Point of a Poisson Stream

In order to determine the critical point of a Poisson stream of rate λ, we need to find the optimizing point in

$$\inf_{\tau\geq 0}\sup_{\theta\geq 0}\theta(B+C\tau)-\tau\lambda(e^{\theta}-1).$$

Consider, for fixed τ, the function

$$f(\theta):=\theta(B+C\tau)-\tau\lambda(e^{\theta}-1).$$

Note that

$$\frac{df(\theta)}{d\theta}=B+C\tau-\tau\lambda e^{\theta}$$

and that if $\theta^*=\log(\frac{B+C\tau}{\lambda\tau})$, then $\theta^*\geq 0$ because $C>\lambda$ is required for stability of the queue, $\frac{df(\theta^*)}{d\theta}=0$, and $\frac{d^2f(\theta^*)}{d\theta^2}=-\tau\lambda e^{\theta}<0$, so θ^* is the maximizing point of f.

Now consider

$$\begin{aligned}g(\tau)&:=\theta^*(B+C\tau)-\tau\lambda(e^{\theta^*}-1)\\&=-B-(C-\lambda)\tau+(B+C\tau)\log\left(\frac{B}{\lambda\tau}+\frac{C}{\lambda}\right)\end{aligned}$$

This equation can be minimized analytically, but yields a negative value for the minimizing τ, so we resort to numerically minimizing g to find τ^*. We then substitute this value of τ^* into the equation $\log(\frac{B+C\tau^*}{\lambda\tau^*})$ to determine θ^*.

3.1.2 Brownian Motion Traffic

Brownian motion has been used as a limiting case of heavy traffic, and also provides a stepping stone to the fractional Brownian motion traffic dealt with later in this thesis.

Definition 15 *A stochastic process $\{X(t);t\geq 0\}$ is said to be a* ***Brownian motion*** *if*

1. *$X(t)$ has stationary and independent increments*

2. *for $t > 0$, $X(t)$ is normally distributed with mean μ and variance $\sigma^2 t$ for some real numbers μ and σ.*

3. $X(0) \overset{a.s.}{=} 0$.

Since $X(t)$ is normally distributed, it may take on negative values, which is inapplicable to queues. We can view a Brownian motion as a perturbation to a constant rate stream, with rate so large that we can safely ignore issues of negative arrivals. Then Proposition 4 lets us consider the Brownian motion and the constant rate stream separately.

Effective Bandwidth of a Brownian Motion Stream

Theorem 16 *The effective bandwidth of a Brownian motion is*

$$eb(\theta, \tau) = \mu + \frac{\theta\sigma^2}{2}$$

where μ is the mean arrival rate, σ^2 is the variance of the arrival rate.

Proof. Since $X(\tau)$ is normally distributed with mean $\mu\tau$ and variance $\sigma^2\tau$, and the moment generating function of such a normally distributed random variable is

$$M(\theta) = \exp\left(\mu\tau\theta + \frac{\tau\sigma^2\theta^2}{2}\right)$$

we have that

$$\begin{aligned} eb(\theta, \tau) &= \frac{1}{\theta\tau} \log M(\theta) \\ &= \frac{1}{\theta\tau}\left(\mu\tau\theta + \frac{\tau\sigma^2\theta^2}{2}\right) \\ &= \mu + \frac{\theta\sigma^2}{2} \end{aligned}$$

∎

Critical Point of a Brownian Motion Stream

Consider

$$\inf_{\tau\geq 0}\sup_{\theta\geq 0}\theta(B+C\tau)-\theta\tau(\mu+\frac{\theta\sigma^2}{2}).$$

Now,

$$\frac{d}{d\theta}\theta(B+C\tau)-\theta\tau\left(\mu+\frac{\theta\sigma^2}{2}\right)=B+\tau\left(C-\mu-\theta\sigma^2\right)$$

and if we set this expression to 0, we get that

$$\theta^*=\frac{B+\tau(C-\mu)}{\sigma^2\tau}>0.$$

Substituting θ^* back into our original equation yields

$$\inf_{\tau\geq 0}\frac{B+\tau(C-\mu)}{\sigma^2\tau}(B+C\tau)-\frac{B+\tau(C-\mu)}{\sigma^2\tau}\tau(\mu+(\frac{B+\tau(C-\mu)}{\sigma^2\tau})\frac{\sigma^2}{2})$$

or after simplifying, $\inf_{\tau\geq 0}\frac{(B+\tau(C-\mu))^2}{2\sigma^2\tau}$

and

$$\frac{d}{d\tau}\frac{(B+\tau(C-\mu))^2}{2\sigma^2\tau}=\frac{-B+(C-\mu)^2\tau^2}{2\sigma^2\tau^2}.$$

If we set this expression to 0, we get that the positive solution is

$$\tau^*=\frac{B}{C-\mu}$$

and therefore, $\theta^*=\frac{2(C-\mu)}{\sigma^2}.$

3.2 Short Range Dependent Models

In this section we drop the assumption of independent increments of the previous section, and focus on short-range dependent models. To define short range dependence, we first need a few definitions.

Definition 17 *If $X(\tau)$ is a process such that $Var(X(\tau)) < \infty$ for each τ, then the* ***autocovariance function*** *of $X(\tau)$ is defined to be*
$\gamma_X(r,s) := Cov[X(r), X(s)]$.

Note that if the process is stationary, then $Cov[X(r), X(s)] = Cov[X(r-s), X(0)]$, and therefore we can and do consider $\gamma_X(r,s)$ to be a function of one variable $\gamma_X(r-s)$. Also, we have that

$$\gamma_X(-k) = \gamma_X(k)$$

as both sides are the covariance of random variables lagged by k time units.

Definition 18 *If $X(\tau)$ is a stationary process such that $Var(X(\tau)) < \infty$ for each τ, then the* ***autocorrelation function*** *of $X(\tau)$ is defined to be*
$\rho_X(k) := \gamma_X(k)/\gamma_X(0)$.

Definition 19 *A stationary stochastic process exhibits* ***short range dependence*** *if*

$$\sum_{k=-\infty}^{\infty} |\rho_X(k)| < \infty$$

A standard example of a short range dependent process is the following Markov On/Off fluid.

3.2.1 Discrete On-Off Markov Fluid

Definition 20 *A* ***Discrete On/Off Markov Fluid*** *is defined to be a two state, discrete time Markov chain, where the transition probability from state 1 to state 2 is μ and from state 2 to state 1 is λ. In state 1, work is produced at constant rate h, while in state 2, no work is produced. Thus the process has transition matrix*

$$P = \begin{bmatrix} 1-\mu & \mu \\ \lambda & 1-\lambda \end{bmatrix}$$

with the components of this matrix denoted by p_{ij}, *and the components of* P^k *denotes by* $p_{ij}^{(k)}$. *The stationary distribution of this chain is given by [5, Example 5.1]*

$$\pi_1 = \frac{\lambda}{\lambda+\mu}$$
$$\pi_2 = \frac{\mu}{\lambda+\mu}.$$

Theorem 21 *The Discrete On/Off Markov Fluid is short range dependent.*

Proof. *Let* x_i *denote the value of the chain when in state* i, *i.e.*

$$x_1 = h$$
$$x_2 = 0.$$

Then we have for $k > 0$

$$\begin{aligned}
\mathbb{E}[X(t+k)X(t)] &= \sum_{i,j\in\{1,2\}} x_i x_j \mathbb{P}[X(t+k) = x_i, X(t) = x_j] \\
&= h^2\mathbb{P}[X(t+k) = h, X(t) = h] \\
&= h^2\mathbb{P}[X(t+k) = h | X(t) = h]\mathbb{P}[X(t) = h] \\
&= h^2 p_{11}^{(k)} \pi_1 \\
&= h^2 \left\{ \frac{1}{\lambda+\mu} \left(\lambda + \mu (1-\mu-\lambda)^k \right) \right\} \frac{\lambda}{\lambda+\mu}
\end{aligned}$$

so

$$\begin{aligned}
\gamma_X(k) &= \mathbb{E}[X(t+k)X(t)] - E[X(t+k)]E[X(t)] \\
&= h^2 \left\{ \frac{1}{\lambda+\mu} \left(\lambda + \mu (1-\mu-\lambda)^k \right) \right\} \frac{\lambda}{\lambda+\mu} - \left(h\frac{\lambda}{\lambda+\mu} \right)^2 \\
&= \frac{h^2\lambda\mu}{(\lambda+\mu)^2} (1-\mu-\lambda)^k
\end{aligned}$$

and

$$\gamma_X(0) = \frac{h^2\lambda\mu}{(\lambda+\mu)^2}.$$

Therefore, for $k > 0$

$$\rho_X(k) = \gamma_X(k)/\gamma_X(0)$$
$$= (1-\mu-\lambda)^k$$

and

$$\sum_{k=-\infty}^{\infty} |\rho_X(k)| < \infty$$

because $|1-\mu-\lambda| < 1$, and $\gamma_X(-k) = \gamma_X(k)$. ∎

Effective Bandwidth of a Discrete On-Off Markov Fluid

In order to derive the effective bandwidth of a discrete on-off Markov fluid we need the following Ergodic Theorem for Markov chains.

Theorem 22 *[5, Theorem 4.1] Let $\{X_n\}$ be an irreducible positive recurrent Markov chain with stationary distribution π, and let $f : \{1, 2\} \to \mathbb{R}$ be such that*

$$|f(1)|\,\pi_1 + |f(2)|\,\pi_2 < \infty.$$

Then for any initial distribution

$$\frac{1}{N}\sum_{k=1}^{N} f(X_k) \stackrel{a.s.}{\to} f(1)\pi_1 + f(2)\pi_2$$

as $N \to \infty$.

One can interpret this result as saying that the average value of f per unit time is $f(1)\pi_1 + f(2)\pi_2$.

Theorem 23 $eb_X(\theta, \tau) = \frac{1}{\theta}\log\left(\frac{\lambda}{\lambda+\mu}e^{\theta h} + \frac{\mu}{\lambda+\mu}\right)$

Proof. For $x \in \{1, 2\}$ define

$$f(1) = e^{\theta h}$$
$$f(2) = 1$$

Then we have

$$\begin{aligned}
eb(\theta, \tau) &= \frac{1}{\theta\tau} \log \mathbb{E}\left[e^{\theta X(0,\tau]}\right] \\
&= \frac{1}{\theta\tau} \log \mathbb{E}\left[e^{\theta \sum_{i=1}^{\tau} X_i}\right] \\
&= \frac{1}{\theta\tau} \log \mathbb{E}_{\pi}\left[\{f(X)\}^{\tau}\right], \text{ by the ergodic theorem for Markov chains} \\
&= \frac{1}{\theta\tau} \log \{[(f(1)\pi_1 + f(2)\pi_2)]^{\tau}\} \\
&= \frac{1}{\theta} \log \left(e^{\theta h}\frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu}\right)
\end{aligned}$$

■

Critical Point of a Discrete On-Off Markov Fluid

In this example, analytically calculating the critical point of the system is difficult or impossible. We will numerically estimate the critical point by numerically searching for the optimizing (θ^*, τ^*) in

$$\begin{aligned}
I &= \inf_{\tau \geq 0} \sup_{\theta \geq 0} \theta(B + C\tau) - \theta\tau eb_X(\theta, \tau) \\
&= \inf_{\tau \geq 0} \sup_{\theta \geq 0} \theta(B + C\tau) - \tau \log \left(e^{\theta h}\frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu}\right)
\end{aligned}$$

by calculating I on a grid, and then repeatedly making the mesh finer until sufficient accuracy is reached.

3.3 Long Range Dependent Models

In this section, we allow long range dependent processes, as many recent studies suggest that modern network traffic is long range dependent.

Definition 24 *A stationary stochastic process $X(\tau)$ exhibits* ***long range dependence*** *if*

$$\sum_{k=-\infty}^{\infty} |\rho_X(k)| = \infty$$

A standard example of a long range dependent process is fractional Brownian motion, with Hurst parameter $H > \frac{1}{2}$.

3.3.1 Fractional Brownian Motion

Definition 25 *A stochastic process $\{X(t); t \geq 0\}$ is said to be a* ***fractional Brownian motion*** *with Hurst parameter H if*

1. *$X(t)$ has stationary increments*

2. *for $t > 0$, $X(t)$ is normally distributed with mean 0*

3. *$X(0) \overset{a.s.}{=} 0$.*

4. *The increments of $X(t)$, $Z(j) := X(j+1) - X(j), j = 0, 1, \ldots.$ satisfy*

$$\rho_Z(k) = \frac{1}{2}\left\{|k+1|^{2H} + |k-1|^{2H} - 2k^{2H}\right\}$$

Theorem 26 *Let $\{X(t); t \geq 0\}$ be a fractional Brownian motion with Hurst parameter $H > 1/2$. Then the increments of $X(t)$ are long range dependent.*

Proof.

$$\begin{aligned}
\rho_Z(k) &= \frac{1}{2}\left\{k^{2H-2}\left(\frac{|k+1|^{2H}}{k^{2H-2}} + \frac{|k-1|^{2H}}{k^{2H-2}} - \frac{2k^{2H}}{k^{2H-2}}\right)\right\} \\
&= \frac{1}{2}\left\{k^{2H-2}\left(k^2\left[\frac{|k+1|^{2H}}{k^{2H}} + \frac{|k-1|^{2H}}{k^{2H}} - 2\right]\right)\right\} \\
&= \frac{1}{2}\left\{k^{2H-2}\left(k^2\left[\left(1+\frac{1}{k}\right)^{2H} + \left(1-\frac{1}{k}\right)^{2H} - 2\right]\right)\right\} \\
&= \frac{1}{2}k^{2H-2}\left(k^2\left[\left(1+\frac{1}{k}\right)^{2H} + \left(1-\frac{1}{k}\right)^{2H} - 2\right]\right) \\
&\to \frac{1}{2}k^{2H-2}2H(2H-1) \text{ as } k \to \infty, \text{ by repeated applications of l'Hospital's rule} \\
&= k^{2H-2}H(2H-1).
\end{aligned}$$

Therefore

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} |\rho_X(k)| &\sim \sum_{k=-\infty}^{\infty} |H(2H-1)k^{2H-2}| \\
&= H(2H-1)\sum_{k=-\infty}^{\infty} |k^{2H-2}| \\
&= \infty
\end{aligned}$$

■

Effective Bandwidth of a Fractional Brownian Motion

Theorem 27 *The effective bandwidth of a fractional Brownian motion with Hurst parameter H is*

$$eb(\theta, \tau) = \mu + \frac{\theta\sigma^2}{2}\tau^{2H-1}$$

where μ is the mean arrival rate, σ^2 is the variance of the arrival rate.

Proof. Since $X(\tau)$ is normally distributed with mean $\mu\tau$ and variance $\sigma^2\tau^{2H}$, and the moment generating function of such a normally distributed random variable is

$$M(\theta) = \exp\left(\mu\tau\theta + \frac{\tau^{2H}\sigma^2\theta^2}{2}\right)$$

we have that

$$\begin{aligned} eb(\theta,\tau) &= \frac{1}{\theta\tau}\log M(\theta) \\ &= \frac{1}{\theta\tau}\left(\mu\tau\theta + \frac{\tau^{2H}\sigma^2\theta^2}{2}\right) \\ &= \mu + \frac{\theta\sigma^2}{2}\tau^{2H-1} \end{aligned}$$

∎

Critical Point of a Fractional Brownian Motion

Consider

$$\inf_{\tau\geq 0}\sup_{\theta\geq 0}\theta(B + C\tau) - \theta\tau(\mu + \frac{\theta\sigma^2}{2}\tau^{2H-1}).$$

Now,

$$\frac{d}{d\theta}\theta(B + C\tau) - \theta\tau\left(\mu + \frac{\theta\sigma^2}{2}\tau^{2H-1}\right) = B + C\tau - \mu\tau - \theta\sigma^2\tau^{2H}$$

and if we set this expression to 0, we get that

$$\theta^* = \frac{B + \tau(C - \mu)}{\tau^{2H}\sigma^2} > 0.$$

Substituting θ^* back into our original equation yields

$$\inf_{\tau\geq 0}\frac{B + \tau(C-\mu)}{\sigma^2\tau}(B + C\tau) - \frac{B + \tau(C-\mu)}{\sigma^2\tau}\tau(\mu + (\frac{B + \tau(C-\mu)}{\sigma^2\tau})\frac{\sigma^2}{2})$$

or after simplifying, $\inf_{\tau\geq 0}\frac{(B + \tau(C-\mu))^2}{2\sigma^2\tau^{2H}}$

and

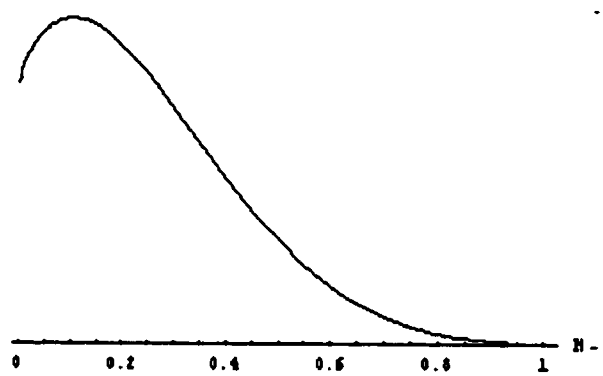
$$\frac{d}{d\tau}\frac{(B+\tau(C-\mu))^2}{2\sigma^2\tau^{2H}} = -\frac{\tau^{-1-2H}(B+(C-\mu)\tau)(BH+(H-1)(C-\mu)\tau)}{\sigma^2}.$$

If we set this expression to 0, we get that the positive solution is

$$\tau^* = \frac{B}{C-\mu}\frac{H}{1-H}$$

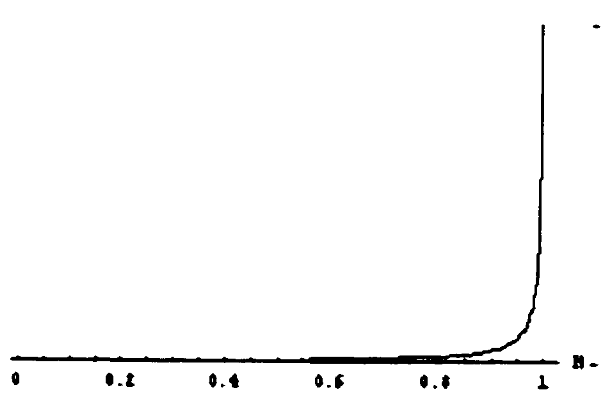
$$\text{and therefore, } \theta^* = \frac{B+(C-\mu)\tau^*}{\sigma^2\left(\tau^*\right)^{2H}}.$$

If we plot θ^* as a function of H, we get



For differing parameter values, the peak moves to the left or right, but always decreases to 0 as $H \to 1$. $\theta^* = 0$ implies that we should use only the mean in evaluating whether or not to accept this connection.

τ^* as a function of H looks like



As $H \to 1$, slower and slower time scales are responsible for buffer overflow, and τ^* becomes infinite, suggesting that we use as long a trace as possible for measurement based admission control.

These results seem counter-intuitive, as long range dependence is typically thought of as a bad thing, making statistics more difficult than in the independent increment case. In particular, we have that the variance of a fractional Brownian motion with Hurst parameter H, at time t is given by:

$$\sigma^2 t^{2H}. \tag{3.2}$$

At $H = 1$, this becomes $\sigma^2 t^2$. This would seem to imply that fluctuations get larger and larger, making prediction, statistical analysis, and measurement based admission control very difficult. However, the following theorem shows that in fact this is not the case at all.

Theorem 28 *[29, Lemma 7.2.1] Suppose that $X(t)$ is a fractional Brownian motion with Hurst parameter $H = 1$ and that $X(0) = 0$ a.s. Then $X(t) = tX(1)$ a.s.*

Thus, $\mathbb{V}(X(t)) = \mathbb{V}(tX(1)) = t^2\mathbb{V}(X(1)) = \sigma^2 t^2$, where all equalities are almost sure. This agrees with Eq. 3.2. However, the variance is not dynamic, i.e. there are no fluctuations with time.

Theorem 29 *[29, Example 7.2.5] Suppose that $X(t)$ is a fractional Brownian motion with Hurst parameter $H = 1$ and that $X(0) = 0$ a.s. Then $X(t) = t(Z + \mu)$ where $\mu \in \mathbb{R}$, and Z is a mean 0 normally distributed random variable.*

Note that, agreeing with the above, $\mathbb{V}(X(t)) = \mathbb{V}(t(Z+\mu)) = t^2\mathbb{V}(Z+\mu) = t^2\mathbb{V}(Z) = \sigma^2 t^2$.

For measurement based admission control then, we only have to sample once, at one time. This is enough to determine $Z + \mu$, after which the traffic is completely determined. In other words, all the variance is in the initial choice of Z, and once that choice is made, the traffic process is deterministic, and our admission control problem is solved.

This is another method to obtain the result of [25, Pg. 4] that rules out the limiting case of H=1, since the process is then deterministic with linear paths.

3.4 Heavy Tailed Processes

In this section, we allow for processes with infinite variance, as many empirical studies have shown the some characteristics of modern network traffic, such as fax call holding times, and web server file sizes, tend to be heavy tailed.

Definition 30 *A random variable X is said to have* ***heavy tails*** *if*

$$x^{\alpha} P[|X| > x] \rightarrow C \text{ as } x \rightarrow \infty$$

where C is a finite positive constant, and $0 < \alpha < 2$. α is called the index of the distribution. A process with heavy-tailed marginal distributions is called a heavy-tailed process.

Example of a Heavy Tailed, Long Range Dependent Process

In this section we explain how to simulate a process that is long range dependent and has heavy tails.

Definition 31 *[6, Definition 13.2.2] A stationary process X_t is called a* **FARIMA(p, d, q)** *process if*

1. *$d \in (-0.5, 0.5)$, and*
2. *$\phi(B)\nabla^d X_t = \theta(B) Z_t$*

   where B is the backward shift operator defined by

$$BX_t = X_{t-1},$$

$$\nabla^d = (1 - B)^d, \text{ is defined by the binomial expansion}$$
$$= \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} B^j,$$

Z_t are uncorrelated, zero mean normally distributed random variables, ϕ, θ are polynomials of degree p, q respectively, and Γ is the gamma function.

We base our simulation on a $FARIMA(0,d,0)$ sequence, but instead of using Gaussian innovations, we use innovations from a heavy tailed distribution. We refer the reader to [?, Section 4] for details. As there is no known algorithm to generate an exact $FARIMA$ model with heavy tailed innovations, we use the infinite moving average representation of this model, and approximate the infinite moving average by a large order moving average.

$$X(\tau) = \sum_{j=0}^{J} c_j Z_{t-j}, \text{ for } t = 1, \ldots, n.$$

where

$$c_0 = 1, \text{ and } c_{j+1} = \frac{j+d}{j+1} c_j$$

In our work, we chose $J = 100$, $d = 0.1$ to generate long range dependence, and used the algorithm in [29, Proposition 1.7.1] to generate the innovations from a heavy tailed $S_{1.8}(1,0,0)$ distribution, i.e. a distribution with characteristic function [29, Definition 1.1.6]

$$\mathbb{E}\left[e^{i\theta X}\right] = \exp\left(-|\theta|^{1.8}\right).$$

Effective Bandwidth and Critical Point of Heavy Tailed Processes

An issue of interpretation arises when we consider traffic that may have heavy tails - the distribution of the traffic stream we want to analyze may not have moments of high enough order. For example, if the traffic stream is made up of IID Cauchy random variables, then the stream does not even have a mean, and does not have a moment generating function, so the effective bandwidth formula is meaningless. However, [1] has circumvented this problem by considering the "adapted" moment generating function, defined for a random variable X by

$$M_X(\theta, T) := \mathbb{E}[e^{\theta X \mathbb{1}_{\{|X|<T\}}}].$$

We will not pursue this line of work any further, and have mentioned it only to explain how others are dealing with the problem of the non-existence of moments.

It seems that there is little chance of finding formulae for describing the effective bandwidth and critical point of a general heavy tailed process. Therefore, we will numerically optimize the rate function to determine the critical point of the system.

3.5 Heavy Tails, Self-Similarity, and Long Range Dependence

Self-similarity, although not used directly in this thesis, links heavy-tails to long range dependence. Therefore, we describe it and the link briefly in this section.

Definition 32 *A process $X(t)$ with continuous time parameter t is said to be* ***self-similar****, with self-similarity parameter H, if for any positive number c, we have*

$$c^{-H}X(ct) \stackrel{\mathcal{D}}{=} X(t)$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution (of processes).

Remark 33 *Other definitions are common, including second-order self similarity, which states that the first and second moments of the process scale as in the above equation.*

Remark 34 *A fractional Brownian motion with Hurst parameter H is self similar with self-similarity parameter H.*

Next, the link from heavy tails to long range dependence, via the self similarity of fractional Brownian motion is explored.

Theorem 35 *[35, Theorem 1] Consider a strictly alternating on/off process, i.e. arrivals are generated at a constant rate 1 when in the on state, followed by an off state where no arrivals are generated, and the states always alternate, i.e. an on state can not be followed by another on state without an off state in between, and conversely. Suppose that the sojourn time in the on state, and the sojourn time in the off state have the same distribution, and that this distribution has heavy tails with index α, and that the mean of this distribution is μ. Denote the sequence of arrivals by $\{W(t); t \geq 0\}$. Suppose M independent copies of this source are aggregated together, and denote the m^{th} stream by $\{W^{(m)}(t); t \geq 0\}$, and define*

$$W^*_M(Tt) = \int_0^{Tt} \left(\sum_{m=1}^{M} W^{(m)}(u) \right) du$$

the aggregated cumulative arrivals in $[0, Tt]$ *and let*

$$H = (3 - \alpha)/2.$$

Then

$$\mathcal{L} \lim_{T\to\infty} \mathcal{L} \lim_{M\to\infty} \frac{(W_M^*(Tt) - TM\frac{1}{2}t)}{T^H L^{1/2}(T) M^{1/2}} = \left[\frac{l}{2\mu(\alpha-1)(2-\alpha)(3-\alpha)}\right]^{1/2} B_H(t)$$

where $\mathcal{L}\lim$ *denotes convergence of finite dimensional distributions,* l *is a constant, and* L *is a function that is slowly varying at infinity, i.e.*

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \text{ for any } t > 0.$$

In words, the normalized aggregate arrival process tends to a constant times a fractional Brownian motion with Hurst parameter $(3-\alpha)/2$, *as the number of sources tends to infinity, and then a time rescaling parameter tends to infinity.*

Remark 36 *If* $\alpha < 2$, *the limiting process is long range dependent.*

Taking the above limit in the reverse order leads to a stable Lévy motion, a process with discontinuous sample paths and infinite variance. In [16], the question of finding a simultaneous time and space limit is addressed, avoiding the two limits required in the above result. Further results on the limiting behaviour in different regimes are presented in [24].

It is important to note that these three concepts (self-similarity, heavy tails, and long range dependence) are independent, although they interact in interesting ways. For example, self-similarity and Gaussian marginals characterize fractional Brownian motion, which is long range dependent, but in general, self similarity does not imply long range dependence.

Chapter 4

Estimating Effective Bandwidth

In this chapter we look at ways of estimating effective bandwidth for a variety of traffic streams. We begin with Poisson arrivals and then calculate confidence intervals for the effective bandwidth of a Brownian motion. Next we discuss the Dembo estimator of effective bandwidth, and develop some of its properties. Later, we generate bootstrap samples to which we apply the Dembo estimator to generate confidence intervals. The properties of the Dembo estimator developed that make it a "good" estimator are now used.

We briefly review bootstrap methods of independent identically distributed data as a precursor to discussing the Moving Blocks Bootstrap for weakly dependent data, which we use to estimate a confidence interval for the effective bandwidth of the discrete on-off Markov fluid. However, the Moving Blocks Bootstrap is known not to work in general for long range dependent data, so we next present the Surrogate Data method for generating bootstrap samples for dependent data with Gaussian marginals. We apply this method to Brownian motion and then discuss the Amplitude Adjusted Surrogate Data method which is similar, but generates samples with arbitrary marginals, not Gaussian ones. We then apply it to a long range dependent, heavy tailed process, the α-Stable FARIMA process discussed earlier.

4.1 Confidence Interval for the Effective Bandwidth of a Poisson Stream

Assume that we have observed a Poisson process $X(\tau)$ at the fixed time intervals $i = 1, 2, 3, ..., n$. Let us first derive a point estimator of the effective bandwidth of this traffic stream.

Proposition 37 *Let $X_1, X_2, ..., X_n$ be IID Poisson(λ) random variables. Then the maximum likelihood estimator (MLE) of λ, is $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$.*

Proof. *The likelihood function is $L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$. Therefore,*

$$\log L(\lambda|\mathbf{x}) = \sum_{i=1}^{n}(x_i \log\lambda - \lambda - \log(x_i!))$$

So, to find the extrema, set the partial derivative equal to zero and solve for λ:

$$\frac{\partial \log L(\lambda|\mathbf{x})}{\partial\lambda} = \sum_{i=1}^{n}(\frac{x_i}{\lambda} - 1) = 0$$

implies that

$$\sum_{i=1}^{n}\frac{x_i}{\lambda} = \sum_{i=1}^{n} 1 = n,$$

so

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and finally

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

■

Theorem 38 *[9, Theorem 7.2.1] Invariance Property for Maximum Likelihood Estimators If $\hat{\eta}$ is the MLE of η, then for any function $f(\eta)$, the MLE of $f(\eta)$ is $f(\hat{\eta})$.*

Corollary 39 *The MLE of $eb(\theta,\tau)$ for Poisson arrivals is $\widehat{eb}(\theta,\tau) = \frac{e^\theta-1}{\theta}\frac{X(n)}{n}$ where $X(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$.*

Proof. Recall that the effective bandwidth of a Poisson arrival stream with rate λ is $\lambda\frac{(e^\theta-1)}{\theta}$. Now, note that $X(n) = \sum_{i=1}^{n} X_i$, so $\hat{\lambda} = \frac{X(n)}{n}$. By the Invariance Property for Maximum Likelihood Estimators , the MLE of $\lambda\frac{(e^\theta-1)}{\theta}$ is $\widehat{\lambda}\frac{(e^\theta-1)}{\theta} = \frac{X(n)}{n}\frac{(e^\theta-1)}{\theta}$ ■

Note that this estimator does not depend on τ. This allows us to use as much data as we have to calculate $\hat{\lambda}$, and therefore a more accurate estimate. We now show that this estimator is unbiased.

Proposition 40 *For Poisson arrivals with rate λ, $\mathbb{E}\left[\widehat{eb}(\theta,\tau)\right] = eb(\theta,\tau)$*

Proof.

$$\begin{aligned}
\mathbb{E}\left[\widehat{eb_P}(\theta,\tau)\right] &= \mathbb{E}\left[\frac{e^\theta-1}{\theta}\frac{X(n)}{n}\right] \\
&= \frac{e^\theta-1}{n\theta}\mathbb{E}\left[X(n)\right] \\
&= \frac{e^\theta-1}{n\theta}n\lambda \\
&= \lambda\frac{(e^\theta-1)}{\theta}
\end{aligned}$$

■

We can also show that $\hat{\lambda}$ is a sufficient statistic [9, Definition 6.1.1] for λ, a fact which will be useful in deriving a confidence interval for $eb_P(\theta,\tau)$

Proposition 41 *$\widehat{\lambda}$ is a sufficient statistic for λ.*

Proof. *Let $X_1, X_2, ..., X_n$ be IID Poisson(λ) random variables with parameter λ. Define*

$T(\mathbf{X}) = \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$. *The joint probability mass function of the sample* $\mathbf{X}$ *is*

$$f(\mathbf{x}|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} = \frac{e^{-n\lambda}\lambda^{nT(\mathbf{x})}}{\prod_{i=1}^{n} x_i!} = g(T(\mathbf{x})|\lambda)h(\mathbf{x})$$

where $g(T(\mathbf{x})|\lambda) = e^{-n\lambda}\lambda^{nT(\mathbf{x})}$, *and* $h(\mathbf{x}) = \frac{1}{\prod_{i=1}^{n} x_i!}$

Therefore, by the Factorization Theorem [9, Theorem 6.1.2], $\widehat{\lambda}$ *is a sufficient statistic for* λ. ∎

To develop a confidence interval for $eb(\theta, \tau)$, we use the fact that "when looking for a confidence set we need to consider only sets based on sufficient statistics" [9, Pg 458].

Proposition 42 *[9, following Example 9.2.10] Let* $X_1, X_2, ...X_n$ *be IID Poisson(*λ*) random variables with parameter* λ, *and let* $X(n) = \sum_{i=1}^{n} X_i$. *Then a* $(1-\alpha)$ *confidence interval for* λ *is given by*

$$\left\{\lambda : \frac{1}{2n}\chi^2_{2X(n),1-\frac{\alpha}{2}} \leq \lambda \leq \frac{1}{2n}\chi^2_{2(X(n)+1),\frac{\alpha}{2}}\right\}$$

Proof. *$X(n)$ is sufficient for* λ, *and* $X(n) \stackrel{\mathcal{D}}{=} \text{Poisson}(n\lambda)$. *Now if* $X(n) = x_0$ *is observed, then we need to solve for* λ *in the following equations*

$$\sum_{k=0}^{x_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = \frac{\alpha}{2}, \text{ and} \tag{4.1}$$

$$\sum_{k=x_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = \frac{\alpha}{2} \tag{4.2}$$

Now, recall the fact [9, Example 3.2.1]: if $V \stackrel{\mathcal{D}}{=} \text{gamma}(\alpha, \beta)$ *where* α *is an integer, and if* $W \stackrel{\mathcal{D}}{=} \text{Poisson}(v/\beta)$, *then for any* v

$$\mathbb{P}[V \leq v] = \mathbb{P}[W \geq \alpha].$$

We can rewrite Equation 4.1 as

$$\begin{aligned}
\frac{\alpha}{2} &= \sum_{k=0}^{x_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!} \\
&= \mathbb{P}[X(n) \le x_0] \\
&= \mathbb{P}[X(n) < x_0 + 1] \\
&= 1 - \mathbb{P}[W \ge \alpha],\ \textit{where } W \overset{\mathcal{D}}{=} \textit{Poisson}(v/\beta),\ v = 2n\lambda,\ \alpha = 2(x_0+1),\ \textit{and } \beta = 2 \\
&= 1 - \mathbb{P}[V \le 2n\lambda],\ \textit{where } V \overset{\mathcal{D}}{=} \textit{gamma}(2(x_0+1), 2) \\
&= \mathbb{P}[V > 2n\lambda] \\
&= \mathbb{P}\left[\chi^2_{2(x_0+1)} > 2n\lambda\right]
\end{aligned}$$

where $\chi^2_{2(x_0+1)}$ *is a chi-squared random variable with* $2(x_0+1)$ *degrees of freedom. This is because a gamma distribution with parameters* $\alpha = p/2$, *where p is an integer, and* $\beta = 2$ *is a chi-squared distribution with p degrees of freedom. The solution to this equation is*

$$\lambda = \frac{1}{2n}\chi^2_{2(x_0+1),\alpha/2}$$

where $\chi^2_{n,\alpha}$ *satisfies* $P[\chi^2 < \chi^2_{n,\alpha}] = \alpha$, *and* χ^2 *is a chi-square random variable with n degrees of freedom.*

Similarly,

$$\begin{aligned}
\frac{\alpha}{2} &= \sum_{k=x_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!} \\
&= \mathbb{P}[X(n) \ge x_0] \\
&= \mathbb{P}[W \ge \alpha],\ \textit{where } W \overset{\mathcal{D}}{=} \textit{Poisson}(v/\beta),\ v = 2n\lambda,\ \alpha = 2x_0,\ \textit{and } \beta = 2 \\
&= \mathbb{P}[V \le 2n\lambda],\ \textit{where } V \overset{\mathcal{D}}{=} \textit{gamma}(2x_0, 2) \\
&= \mathbb{P}[V < 2n\lambda]\ \textit{because the gamma distribution is continuous} \\
&= \mathbb{P}\left[\chi^2_{2x_0} < 2n\lambda\right].
\end{aligned}$$

The solution to this equation is

$$\lambda = \frac{1}{2n}\chi^2_{2x_0,\alpha/2}.$$

Finally, a $(1-\alpha)$ *confidence interval for* λ *is*

$$\left\{\lambda : \frac{1}{2n}\chi^2_{2x_0,1-\frac{\alpha}{2}} \le \lambda \le \frac{1}{2n}\chi^2_{2(x_0+1),\frac{\alpha}{2}}\right\}.$$

■

We now extend this to $eb(\theta,\tau)$.

Corollary 43 *Let* $X(n)$ *be a Poisson process with parameter* λ. *Then a* $(1-\alpha)$ *confidence interval for* $eb(\theta,\tau)$ *is given by*

$$\left(\frac{e^\theta - 1}{\theta}\frac{1}{2n}\chi^2_{2X(n),1-\frac{\alpha}{2}},\ \frac{e^\theta - 1}{\theta}\frac{1}{2n}\chi^2_{2(X(n)+1),\frac{\alpha}{2}}\right). \tag{4.3}$$

Proof. If

$$\begin{aligned} \frac{1}{2n}\chi^2_{2x_0,1-\frac{\alpha}{2}} \le \lambda \le \frac{1}{2n}\chi^2_{2(x_0+1),\frac{\alpha}{2}}&, \text{ then} \\ \frac{e^\theta - 1}{\theta}\frac{1}{2n}\chi^2_{2x_0,1-\frac{\alpha}{2}} \le \lambda\frac{e^\theta - 1}{\theta} \le \frac{e^\theta - 1}{\theta}\frac{1}{2n}\chi^2_{2(x_0+1),\frac{\alpha}{2}}& \\ \frac{e^\theta - 1}{\theta}\frac{1}{2n}\chi^2_{2x_0,1-\frac{\alpha}{2}} \le eb_P(\theta,\tau) \le \frac{e^\theta - 1}{\theta}\frac{1}{2n}\chi^2_{2(x_0+1),\frac{\alpha}{2}}& \end{aligned}$$

■

4.2 Confidence Interval for the Effective Bandwidth of a Brownian Motion Stream

Since the marginal distribution of a Brownian motion is Gaussian, we can estimate μ by $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$, and σ^2 by $S^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. One can then use normal distribution theory to calculate confidence intervals for both μ and σ^2, and then for $eb(\theta,\tau)$ as follows.

First, the Bonferroni inequality allows us to combine confidence intervals for μ and for σ^2

Theorem 44 *Bonferroni's Inequality*

For any events A and B, $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1$

Proof. $1 \geq \mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$ *so, after rearranging we have* $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1$ ■

Now, note that if p and q are such that $(1-p)+(1-q)-1 = (1-\alpha)$, then in order to find a $(1-\alpha)$ confidence interval for $eb(\theta, \tau)$ we can simply find a $(1-p)$ confidence interval for μ and a $(1-q)$ confidence interval for σ^2.

We have that

$$\mathbb{P}\left[\overline{X} - t_{n-1,p/2}\sqrt{\frac{S^2}{n}} \leq \mu \leq \overline{X} + t_{n-1,p/2}\sqrt{\frac{S^2}{n}}\right] = 1 - p$$

where $\mathbb{P}\left[T_{n-1} \leq t_{n-1,p/2}\right] = 1 - \frac{p}{2}$, and T_{n-1} is a Student's T distributed random variable with $n-1$ degrees of freedom [9, Pg. 415].

Also

$$\mathbb{P}\left[\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right] = 1 - q$$

where we choose a and b such that $\mathbb{P}[a \leq Q \leq b] = 1 - q$ where Q is a χ^2_{n-1} [9, Pg. 416].

Finally, we combine these two confidence intervals by taking the lower bounds for each and substituting them into the formula for the effective bandwidth of a Brownian motion stream to obtain the lower bounds for $eb(\theta, \tau)$, and similarly for the upper bounds. This leads to, using the suffix U (L) to denote upper (lower) bound on the appropriate confidence interval, that a $(1-\alpha) = (1-p)+(1-q)-1$ confidence interval for $eb\theta, \tau)$ is

$$\left(\mu_L + \frac{\theta\sigma_L^2}{2}, \mu_U + \frac{\theta\sigma_U^2}{2}\right) \tag{4.4}$$

Also note that we can choose different combinations of p and q that satisfy $(1-p)+(1-q)-1 = (1-\alpha)$. Therefore, we can choose values that numerically minimize the length of this interval, but for the purposes of this thesis we will choose $p = q$ and leave the selection of optimal choice of these parameters for further work.

4.3 The Dembo Point Estimator of Effective Bandwidth

For traffic streams that are more complicated than Poisson or Brownian motion processes, we will use resampling methods, defined later in this chapter. In order to apply these methods, we need a point estimator of effective bandwidth that will work for any stream we wish to analyze. The most commonly used is the Dembo estimator.

In [12], the authors propose the following estimator, originally suggested by Amir Dembo, for the scaled cumulant generating function of a traffic stream: first choose a block size b for which the block sums

$$\widetilde{X_1} := \sum_{k=1}^{b} X_k, \quad \widetilde{X_2} := \sum_{k=b+1}^{2b} X_k, \ldots$$

are approximately independent and identically distributed. Then, use as the estimator

$$\frac{1}{b} \log \frac{1}{\lfloor n/b \rfloor} \sum_{i=1}^{\lfloor n/b \rfloor} e^{\theta \widetilde{X}_i} \tag{4.5}$$

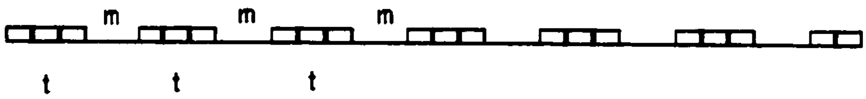
We translate this estimator into that of the effective bandwidth as in [30]

$$\widehat{eb_D}(\theta, \tau) := \frac{1}{\theta\tau} \log \left[\frac{1}{\lfloor T/\tau \rfloor} \sum_{i=1}^{\lfloor T/\tau \rfloor} e^{\theta \tilde{X}_i} \right] \tag{4.6}$$

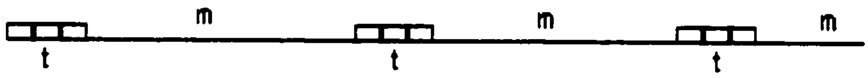
where T is the duration of the trace, $\lfloor a \rfloor$ denotes the largest integer less than or equal to a. We will refer to this estimator as the "Dembo Estimator". We will also use the notation $\widehat{eb^n_D}(\theta, \tau)$ when needed, to indicate the dependence of the estimator on the sample size n. Note that in effect, τ has become our block size.

If our traffic is m-dependent, i.e. that points more than m time units apart are independent, and the trace (sample) is long enough, we can select the $\tilde{X}_i$ to be independent by choosing blocks of length τ and leaving a gap of m points in between the blocks.

For example, if our data is known to be 2-dependent, and we are trying to analyze $\widehat{eb_D}(\theta, 3)$, we get the following picture.



One problem that can arise is that there may not be enough data to be able to choose blocks in this manner, i.e. the length of the trace may be the same size as the lag over which the sample is dependent. This illustrated in the following picture, where we have chosen $m = 10$ and $t = 3$, and clearly in this case, the variance of any estimates will be much larger, due to the fact that we have many fewer samples.



We will use the method of surrogate data (developed later in this chapter) to solve this problem.

4.3.1 Properties of the Dembo Estimator

Theorem 45 *The Dembo estimator of effective bandwidth is found by replacing the population distribution of the traffic stream by the empirical distribution of the traffic stream.*

Proof. Let us fix τ, θ and write $eb(\theta, \tau)$ as $eb(F) = \frac{1}{\theta\tau} \log \mathbb{E}_F \left[e^{\theta \tilde{X}_i} \right]$ to indicate its dependence on the underlying distribution of the traffic stream, which we have denoted by F. Also, $\tilde{X}_i$ are IID F. Then if we denote the empirical distribution by F_n, we have

$$
\begin{aligned}
eb(F_n) &= \frac{1}{\theta\tau} \log \mathbb{E}_{F_n} \left[e^{\theta \tilde{X}_i} \right] \\
&= \frac{1}{\theta\tau} \log \left[\frac{1}{n} \sum_{i=i}^{n} e^{\theta \tilde{X}_i} \right] \\
&= \widehat{eb_D}(\theta, \tau)
\end{aligned}
$$

■

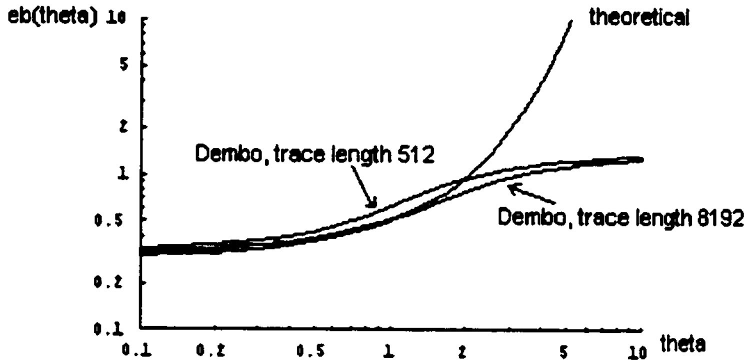
Theorem 46 *The Dembo estimator is biased in general, i.e. for any fixed θ, τ and any finite n, $\mathbb{E}_F\left[\widehat{eb^n_D}(\theta,\tau)\right] \neq eb(\theta,\tau)$.*

Proof.

$$\begin{aligned}
\mathbb{E}_F\left[\widehat{eb^n_D}(\theta,\tau)\right] &= \mathbb{E}_F\left[eb(F_n)\right], \text{ by Theorem 45} \\
&= \mathbb{E}_F\left[\frac{1}{\theta\tau}\log \mathbb{E}_{F_n}\left[e^{\theta\tilde{X}_i}\right]\right] \\
&= \frac{1}{\theta\tau}\mathbb{E}_F\left[\log \mathbb{E}_{F_n}\left[e^{\theta\tilde{X}_i}\right]\right] \\
&\leq \frac{1}{\theta\tau}\log \mathbb{E}_F\left[\mathbb{E}_{F_n}\left[e^{\theta\tilde{X}_i}\right]\right], \text{ by Jensen's Inequality for concave functions} \\
&= \frac{1}{\theta\tau}\log \mathbb{E}_F\left[\frac{1}{n}\sum_{i=i}^{n} e^{\theta\tilde{X}_i}\right] \\
&= \frac{1}{\theta\tau}\log \left[\frac{1}{n}\sum_{i=i}^{n} \mathbb{E}_F\left[e^{\theta\tilde{X}_i}\right]\right] \\
&= \frac{1}{\theta\tau}\log \mathbb{E}_F\left[e^{\theta\tilde{X}_i}\right], \text{ because the } \tilde{X}_i \text{ are IID } F. \\
&= eb(\theta,\tau)
\end{aligned}$$

So we have that $\mathbb{E}_F\left[\widehat{eb^n_D}(\theta,\tau)\right] \leq eb(\theta,\tau)$. But we also know that there is equality in Jensen's Inequality if and only if the concave function is affine, which *log* is not. Therefore the inequality is strict, and the conclusion follows. ■

In the following graph, we have plotted the theoretical effective bandwidth of a Poisson stream, as well as the effective bandwidth as estimated by the Dembo estimator, for two different trace lengths, each with the same parameter ($\lambda = 0.3$) and the same random seed for the simulation. The estimate based on the longer trace tracks the theoretical value better, but the trace is sixteen times as long. This clearly illustrates that attempting to estimate the effective bandwidth of a stream for all values of θ would require huge amounts of data. However, estimating effective bandwidth for all values of θ is not required, as we use it only at the critical point.



4.4 Resampling Methods

In general, statistical estimation of complicated models, such as the models used for network traffic, can only be carried out in simple cases, or with strong assumptions. In particular, dependent data, such as the discrete on-off Markov fluid and fractional Brownian motion are difficult to analyze statistically. Resampling methods allow for a much wider variety of models to be analyzed, as they replace theoretical derivations with extensive computation. One must still validate the results of these models, but it is frequently easier to validate a result (through observation of the system being studied) than it is to derive results about the system.

In this section, we look at the bootstrap method for IID random variables, as well as the moving blocks bootstrap method for weakly dependent random variables. The moving blocks bootstrap method is known to fail in some long range dependent cases; for this reason we propose using the method of surrogate data, first developed for hypothesis testing in the analysis of nonlinear time series. This method appears to work well even in long range dependent cases, but requires that the data have a marginal distribution that is Gaussian, which is clearly not the case for many examples of network traffic. Thus we turn to the method of amplitude adjusted surrogate data, which appears to work with arbitrary marginal distributions.

The general pattern we follow is:

Algorithm 47 *Resampling algorithm for estimating confidence intervals for effective bandwidth*

1. *Input a realization of an arrival stream* $\{x_i\}_{i=1}^{N}$

2. *Choose a large number B*

3. *Generate B synthetic data sets* $\{y_i\}_{i=1}^N$ *that are similar to the input stream*

4. *Calculate the Dembo estimate of the effective bandwidth of the synthetic data sets*

$$\widehat{eb_D(\theta,\tau)}_j = \frac{1}{\theta\tau}\log\left[\frac{1}{\lfloor T/\tau\rfloor}\sum_{i=1}^{\lfloor T/\tau\rfloor} e^{\theta y((i-1)\tau, i\tau]}\right], j = 1, 2, \ldots, B$$

5. *Form the histogram of these estimates, and use as a* $(1-\alpha)$ *confidence interval of the effective bandwidth a* $(1-\alpha)$ *interval of the histogram of these estimates. This can be done in many ways, here we select symmetric intervals, i.e. upper and lower cut-offs have equal probability.*

This algorithm raises two questions immediately.

- How many samples are enough? i.e. when is B large enough? This question is usually addressed by either fixing a number thought to be sufficiently large in advance, say $B = 10,000$, or by continuing to generate synthetic data sets until the final answer changes by less than a desired amount, such as the bounds for our interval not changing by more than 0.01%.

- How do we generate synthetic data sets, and what do we mean by "similar" to the input stream? This question is addressed in the following sections.

4.4.1 Bootstrap Methods for IID Data

The bootstrap was developed by Bradley Efron in the late 1970's to estimate parameters of distributions in situations where theoretical derivation of the sampling distribution is difficult or impossible. For an accessible introduction, see [13]. Rather than depend on a model, as in normal distribution theory, one forms a reference distribution by resampling with replacement from the observed data. We will not use this method other than as an stepping stone to the Moving Blocks Bootstrap discussed later.

Algorithm 48 *[13, Pg. 35] Generate a Bootstrap sample for IID Data*

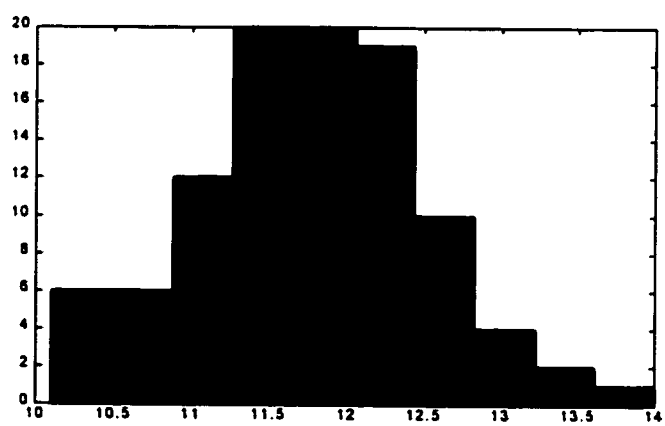
Denote the observed sample by $x_1, x_2, ..., x_n$.

Form the empirical distribution $\widehat{F}$ *of the sample by putting mass* $\frac{1}{n}$ *for each* x_i, *for* $i = 1, 2, ..., n$. *i.e.*

$$\mathbb{P}[X_i^* = x_j] = \frac{1}{n}$$

and draw a sample from this distribution. In other words, we sample with replacement from our observed data.

Example 49 *Suppose we wish to estimate the mean of a sample, and we want to know what the sampling error is. If our observed sample is {13, 11.5, 13.5, 14, 10, 9, 15, 14.5, 8, 9.5}, the average is* 11.8. *But how much would this vary if the experiment were repeated? Bootstrap gives us one way to answer the question. Generate a bootstrap sample from the original data by sampling with replacement from the original data. Let our first bootstrap sample be {13, 11.5, 11.5, 8, 9.5, 9.5, 8, 9, 8, 14} with an average of* 10.2. *Now repeat this procedure a large number of times (say* 100*) and from the resulting data, we generate a histogram of the observed averages of the bootstrap samples.*



We find that the results lie between 10.1 *and* 14, *with* 90% *of the data lying between* 10.45 *and* 12.9. *This then is our bootstrap* 90% *confidence interval for the mean* (10.45, 12.9). *In this situation, normal distribution theory would have worked with fewer computations, and yield* (10.33, 13.27), *but the point of bootstrap is to work when the sampling distribution is difficult to derive.*

In [3], the question of whether or not the bootstrap gives the correct answer is addressed.

Theorem 50 *[3, Theorem 2.1] Suppose $X_1, X_2, \ldots$ are IID, and have finite positive variance σ^2. Along almost all sample sequences $X_1, X_2, \ldots$, given $(X_1, X_2, \ldots, X_n)$, as n and m tend to ∞:*

1. *The conditional distribution of $\sqrt{m}(\frac{1}{m}\sum_{j=1}^{m} X_j^* - \frac{1}{n}\sum_{j=1}^{n} X_j) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$, where X_j^* is the j^{th} sample from F^*.*
2. *For $\varepsilon > 0$, $\mathbb{P}[|s_m^* - \sigma| > \varepsilon | X_1, X_2, \ldots, X_n] \xrightarrow{a.s.} 0$, where $s_m^* = \frac{1}{m}\sum_{i=1}^{m}\left(X_i^* - \left(\frac{1}{m}\sum_{j=1}^{m} X_j^*\right)\right)^2$*

What this means is that the bootstrap estimate of the mean and the standard (normal distribution based) estimate of the mean coincide asymptotically. Furthermore, [3] also prove the following result about the convergence in probability of the bootstrap empirical distribution function to the (unknown) population distribution function F.

Theorem 51 *[3, Corollary 4.1] For almost all sample sequences $X_1, X_2, \ldots$, given $(X_1, X_2, \ldots, X_n)$, as n and m tend to ∞, $||F_{nm} - F|| \xrightarrow{P} 0$. Here F_{nm} is the empirical distribution function of the m resampled data points, from the original n data points, and $||\cdot||$ is the sup-norm.*

Remark 52 *This is needed for our problem, as we have shown that the Dembo estimate of effective bandwidth depends on the empirical distribution, so applying the Dembo estimator to bootstrap samples should yield a reasonable confidence interval.*

These results have been extended to the case of the mean of IID α-stable random variables, for $1 < \alpha < 2$, by changing the bootstrap sample size [28, Pg. 15].

4.4.2 Moving Blocks Bootstrap for Short Range Dependent Data

We now outline an extension of the bootstrap to weakly dependent data that was developed independently by Künsch [18] and Liu & Singh [23], and is known as the Moving Blocks Bootstrap. It is designed to capture the dependence structure of non-IID data, which the IID-bootstrap would destroy. In the next chapter, we will use this algorithm to estimate the effective bandwidth of the discrete on-off Markov fluid discussed earlier.

Algorithm 53 *Generate a Moving Blocks Bootstrap sample for short range dependent data*

Denote the observed sample by $x_1, x_2, ..., x_n$, and choose a block size b. Here we assume that b divides n evenly. This assumption can be eliminated, at the cost of complicating the exposition.

1. *Form blocks of size b of adjacent points, and denote these blocks by y_i, for $i = 1, 2, ..., n - b + 1$. Note that these blocks overlap.*

2. *Form the empirical distribution $\widehat{F}$ of the sample by putting mass $\frac{1}{n-b+1}$ for each y_i, for $i = 1, 2, ..., n - b + 1$, i.e.,*

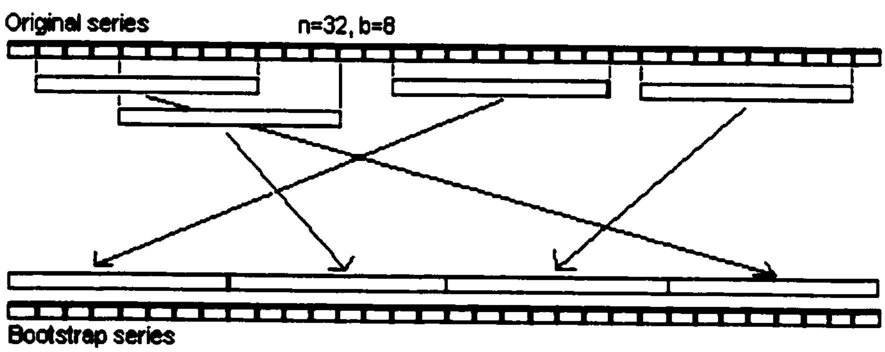
$$\mathbb{P}[Y_i^* = y_j] = \frac{1}{n - b + 1}$$

3. *Draw a sample from $\widehat{F}$:*

$$Y_1^*, Y_2^*, \ldots, Y_{\lfloor n/b \rfloor}^* \overset{IID}{\sim} \widehat{F}$$

In words, choose blocks of size b, uniformly with replacement from the original data, until a series the same length as the original series is obtained.

A picture shows all.



Note that if we choose a block size of 1, then the Moving Blocks Bootstrap is the IID Bootstrap.

Example 54 *Suppose we are given a time series and we want an estimate of the autocorrelation at lag* 1. *Let the time series be (13, 11.5, 13.5, 14, 10, 9, 15, 14.5, 8, 9.5, 9, 10), which has* *autocorrelation at lag 1 of 0.885193. For illustrative purposes, choose a block size of* 3. *So, we now sample with replacement from the set of blocks of length 3, which are: [13, 11.5, 13.5], [11.5, 13.5, 14], [13.5, 14, 10], [14, 10, 9], [10, 9, 15], [9, 15, 14.5], [15, 14.5, 8], [14.5, 8, 9.5], [8, 9.5, 9], [9.5, 9, 10] For example, a moving blocks bootstrap sample of the time series would be (13, 11.5, 13.5, 13, 11.5, 13.5, 9.5, 9, 10, 14.5, 8, 9.5), which has autocorrelation at lag 1 of 0.889565. Similarly, repeat this process a large number of times (say* 100*) and from the resulting data, we generate a histogram of the observed autocorrelations of the moving blocks bootstrap samples. We find that the results lie between 0.815578 and 0.927526, with* 90% *of the data lying between 0.837993 and 0.91411. This then is our moving blocks bootstrap* 90% *confidence interval for the first order autocorrelation of the time series* (0.837993, 0.91411).

Note that this procedure (largely) preserves the autocorrelation of lags less than b in the time series, because the majority of terms used to calculate the autocorrelation with lag less than b will be within blocks, but destroys the correlation structure at larger lags because the majority of terms used to calculate the autocorrelation with lag greater than b will be outside the blocks. In practice one has to choose a suitable block size that balances the length of the correlation structure with the need for many samples i.e. if one has a correlation structure that is significant at lags the same length as the time series, then using a small block size will destroy the correlation properties of the boostrapped samples, while using a large block size will reduce the number of possible different samples. Analogous to Theorem 50, it has been shown that this procedure works for smooth functions of the sample mean for weakly dependent data [18, Theorem 3.5], the conclusion of which is:

$$\sup_x |\mathbb{P}[(T_N^* - T_N) \leq x | X_1, ..., X_N] - \mathbb{P}[(T_N - \mu) \leq x]| \overset{a.s.}{\to} 0$$

where T_N^* is the moving blocks bootstrap statistic, T_N is the standard (non-bootstrap) statistic, and μ is the parameter we are seeking. In other words the bootstrap estimate of the statistic and the standard (non-bootstrap) estimate of the parameter coincide asymptotically.

Analogous to Theorem 51, it has been shown that this procedure works for the empirical process [27, Theorem 2.3] for strongly mixing stationary data. However, it fails in some cases for long range dependent data, as "the MBB (moving blocks bootstrap) provides a valid approximation to the distribution of $d_n^{-1}S_n$ only when $d_n^{-1}S_n$ is asymptotically normal" [20, Corollary 2.1], where d_n^{-1} is a normalizing factor and S_n is the (partial) sum of the centered random variables. This is why we look at the method of surrogate data in the next section.

Another complication with this method is choosing the "best" block size, and although in practice choosing a block size of (length of series)$^{1/3}$ has proven effective, research continues in this area [7].

Much further work done on a variety of bootstrap techniques for dependent data has appeared in the literature, see [21], [41] for examples.

4.4.3 Surrogate Data Method for Input Streams with Gaussian Marginals

Theiler [36] proposed a method of generating synthetic data that preserves the mean and the autocorrelation of the original series. The mean is preserved by subtracting it from the original series, and then adding it back to the synthetic series. This is standard, as it is frequently also a step in moving blocks bootstrap, or even classical time series analysis (ARMA modeling, etc.). For the autocorrelation, he proposed taking the Fourier transform of the series, and then scrambling the phases of the transformed data, while ensuring that the phases remain symmetric so that when the data is transformed back it is real valued. The phases can be either sampled with replacement from the original phases (i.e. bootstrapped), or they can be sampled uniformly on the $(0, 2\pi]$ interval. Here, we use the latter choice for its simplicity.

Now, because the autocorrelation function of a time series is determined by the power of its Fourier transform (i.e. the square of the coefficients), and the power of the Fourier transform does not depend on the phase of the data, the output of this procedure has the same autocorrelation as the input does. Furthermore, this procedure requires only that the data be a realization of a linear Gaussian process, and a particular such process does not have to be specified.

In the next chapter, we will use this algorithm to estimate the effective bandwidth of the fractional Brownian motion, as it has Gaussian marginals. Since fractional Brownian motion is long range dependent, this method is not guaranteed to work. However, it is one of the main points of this thesis that the method does appear to work in long range dependent cases too. This will be explained further later in this section.

First, here is how we generate a single Surrogate Data Sample.

Algorithm 55 *[37] Generate a Surrogate sample for short range dependent data with Gaussian marginals*

1. *Input a time series* $y[t]$, $t = 1..N$
2. *Compute the discrete Fourier transform of the data:* $z[t] := DFT(y[t])$. *Note that* $z[t]$ *has both real and imaginary parts.*
3. *Randomize the phases:* $z\prime[t] := z[t]e^{i\phi[t]}$, *where* $\phi[t]$ *is uniformly distributed on* $(0, 2\pi]$
4. *Symmetrize the phases to ensure that* $z''[t] = -z''[-t]$. *The mechanics of this will depend on the way in which your software stores the components of the FFT.*
5. *Invert the discrete Fourier transform:* $y\prime[t] := DFT^{-1}z''[t]$. *Note that because of the symmetry of the phases, the resulting time series* $y\prime[t]$ *is real.*
6. *Output* $y\prime[t]$

Example 56 *Suppose we want to generate a surrogate data set for the time series (0.707, 1.0, 0.707, 0.0, -0.707, -1.0, -0.707, 0.0, 0.707, 1.0, 0.707, 0.0, -0.707, -1.0, -0.707, 0.0). The DFT of this series is approximately (0, 0, 1.414 + 1.414 i, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1.414 - 1.414 i, 0). Random, symmetric phases are (0.0, 1.433, 4.543, 1.996, 0.878, 4.724, 2.792, 0.736, 0.0, -0.736, -2.792, -4.724, -0.878, -1.996, -4.543, -1.433). Our data to inverse transform is then (0.0, 0.0, 1.155 - 1.632 i, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.155 + 1.632 i, 0.0). Finally, our inverse DFT'd data, i.e our surrogate data set, is (0.577, 0.985, 0.816, 0.169, -0.577, -0.985, -0.816, -0.169, 0.577, 0.985, 0.816, 0.169, -0.577, -0.985, -0.816, -0.169). In order to calculate a statistic, one would normally generate many surrogate data sets of the original, calculate the statistic of each, assemble them into a histogram, and proceed as in the bootstrap methods.*

This method is frequently used in nonlinear time series analysis to perform hypothesis tests, however little theoretical work has been done to assess its limits. The following two results are notable exceptions.

Theorem 57 *[8, Theorem 1] Suppose the autocovariance function of the original process is absolutely summable. Let*

$$f_{j,Y^*} = \sum_{t=1}^{n} Y_t^* e^{\frac{2\pi i t j}{n}}$$

Then

$$\left| \frac{1}{2\pi n} \mathbb{E}\left[|f_{j,Y^*}|^2 \right] - f\left(\frac{2\pi j}{n} \right) \right| \rightarrow 0$$

as $n \rightarrow \infty$, where f is the spectral density of the original process (which determines the autocorrelation of the process), and Y^* is a surrogate data version of the original process.

In other words, the surrogate data has the same Fourier spectrum as the original data, asymptotically. The next result says that (roughly) the marginal distribution of the surrogate data converges in distribution to a Gaussian distribution.

Theorem 58 *[8, Corollary 1] If $\{Y_h, h \geq 1\}$ is a stationary ergodic sequence such that $\{Y_h^2, h \geq 1\}$ is ergodic with $\mathbb{E}[Y_1^2] = \sigma^2$, then*

$$Y_j^* \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

almost surely, for $j = 1, 2, ...$

Remark 59 *The surrogate data method generates synthetic data that has the same* ***sample*** *autocorrelation structure as the original series, not the same autocorrelation as the population. i.e. we may be reproducing artifacts of the sample, and not the true distribution we are seeking.*

Remark 60 *The autocorrelation we are talking about is the* circular *autocorrelation, defined by*

$$\rho^c(T) := \frac{1}{N} \left(\sum_{t=1}^{N-T} x_t x_{t+T} + \sum_{t=N-T+1}^{N} x_t x_{t+T-N} \right),$$

which in the limit of infinite sample size approaches the population autocorrelation function, but *in the case of finite sample size is only approximately equal to the population autocorrelation function [38].*

Remark 61 *The original data may be obviously* non-*Gaussian, which is what the method of amplitude adjusted surrogate data was created to address. This is discussed later in this chapter.*

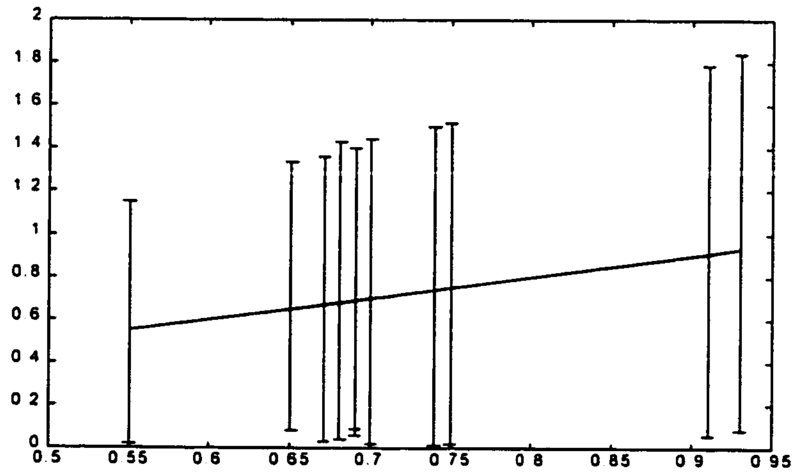
4.4.4 Surrogate Data and Long Range Dependence

The fact that the surrogate data method generates synthetic data sets that are linear and have Gaussian marginals is not as restrictive as it seems, as "a long-memory process can always be approximated by an ARMA(p,q) process" [6, Pg. 520], i.e. a sufficiently complicated linear Gaussian process. Also, a linear, Gaussian time series is determined by its autocorrelation function, so we have the following chain of reasoning:

- A fractional Brownian motion (long memory process) can be approximated arbitrarily well by an ARMA(p,q) process with sufficiently large p and q. It is determined by its mean and autocovariance function.
- The approximating ARMA (p,q) process is a linear process with Gaussian marginals, and as such it is determined by its mean and autocovariance function.
- The surrogate data method outputs a series with the same autocovariance as its input, and approximately Gaussian marginals.
- Therefore, the output is a fractional Brownian motion with the same long range dependence properties as the input.

The only obstacle to proving this is that the proof of [8, Theorem 1] requires that the autocovariance function of the series must be absolutely summable, which is not the case for long range dependent data.

A small simulation was performed to see how these facts are reflected in actual situations. A fractional Brownian motion trace of length 5,000 was generated according to the method of superposition of heavy-tailed on/off sources [35, Section 5], where for this simulation 100 sources were used. Next, one hundred surrogate data sets were generated from the original data set, and then the Hurst parameter of each surrogate data set was estimated by the variance time plot method. The upper five and lower five were dropped, with the remainder forming a 90% confidence interval. This experiment was repeated eleven times with different Hurst parameters. The results follow.



Thus, it seems that the absolute summability hypothesis of Theorem 57 may be stronger than is needed to achieve the result, and that the method of surrogate data may work in the long range dependent case too.

4.4.5 Amplitude Adjusted Surrogate Data Method for Input Streams with Arbitrary Marginals

The amplitude adjusted surrogate data method is designed to extend the surrogate data method by generating a synthetic data set that has exactly the same marginal distribution as the original, and a similar autocorrelation structure. In the next chapter, we will use this algorithm to estimate the effective bandwidth of the long range dependent, heavy tailed α-stable FARIMA process.

Algorithm 62 *[37] Generate an Amplitude Adjusted Surrogate Sample for data with arbitrary marginals*

1. *Input a time series* $x[t]$, $t = 1..N$

2. *Sort the data:* $Sx[k], k = 1..N$

3. *Make a ranked time series $Rx[t]$, defined to satisfy $Sx[Rx[t]] = x[t]$. Note that $Sx[k]$ is a monotone function with a well-defined inverse, so $Rx[t] = Sx^{-1}[t]$ is a static rescaling of $x[t]$.*

4. *Create a random Gaussian data set $g[t], t = 1..N$*

5. *Sort the Gaussian series $Sg[k], k = 1..N$*

6. *Define a new time series $y[t] = Sg[Rx[t]]$. The new series is a static rescaling of $x[t]$, with the property that the amplitude distribution is Gaussian.*

7. *Use the Surrogate Data algorithm to make a surrogate data set of this Gaussian time series $y\prime[t]$*

8. *Make a ranked time series for $y\prime[t]$, call it $Ry\prime[t]$*

9. *Let $x\prime[t] := Sx[Ry\prime[t]]$*

10. *Output $x\prime[t]$*

Remark 63 *Note that since the output is just a shuffling of the input data, that the resulting data set has exactly the same marginal distribution as the original data.*

With this method we are assuming that the observed data is a monotonic rescaling of a Gaussian process, and we call this rescaling h. Since it is monotonic, it is invertible, and we need to find both h and its inverse. Once we have them, we can take our observed data, transform it by h^{-1} to get Gaussian data, and we can apply the standard surrogate data method to this set to get new synthetic data. For each of these synthetic sets, we simply apply h to get a rescaled version of the synthetic data.

In order to estimate h and its inverse, we generate a Gaussian white noise data set and sort it, then map it to the sorted original data. This gives us our approximation to h.

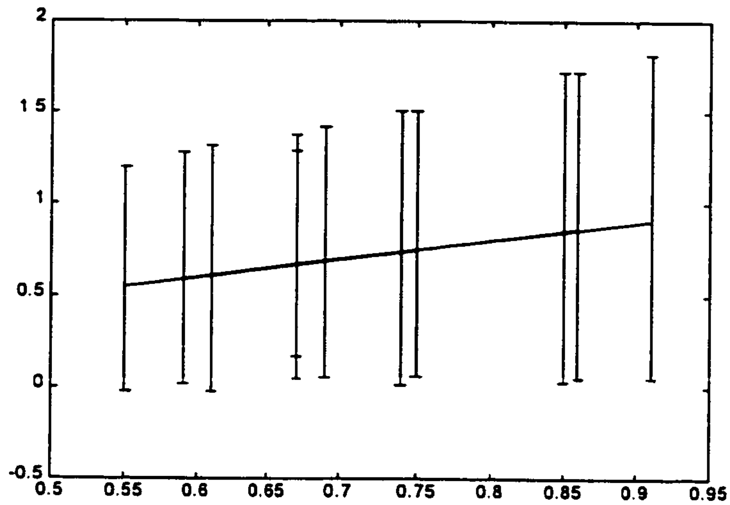
Example 64 *Suppose we want to generate an amplitude adjusted surrogate data set for the time series (0.7, 1.0, 0.8, 0.0, -0.7, -1.0, -0.8, 0.2). The sorted version of this series is Sx=(-1.0, -0.8, -0.7, 0.0, 0.2, 0.7, 0.8, 1.0), with ranked series Rx=(6, 8, 7, 4, 3, 1, 2, 5). Our Gaussian series is g=(1.5, -0.56, 0.85, 1.04, -1.55, -2.9, -0.44, 0.57), and the sorted Gaussian* *series is Sg=(-2.9, -1.55, -0.56, -0.44, 0.57, 0.85, 1.04, 1.5). The static rescaling of the original data is y=Sg[Rx]=(0.85, 1.5, 1.04, -0.44, -0.56, -2.95, -1.55, 0.57). The surrogate of y is y'=(-1.0, -0.87, 0.89, 1.52, 1.33, -0.64, -1.05, -1.76), and the ranked series of y' is Ry'=(3, 4, 6, 8, 7, 5, 2, 1). Finally, x'=Sx[Ry']=(-0.7, 0.0, 0.7, 1.0, 0.8, 0.2, -0.8, -1.0) is our amplitude adjusted surrogate data set.*

Another way to look at this algorithm is in terms of order statistics, where for a sample $x_1, x_2, ..., x_n$ the order statistics are the sample values arranged in ascending order, and denoted by $x_{(1)}, x_{(2)}, ..., x_{(n)}$, i.e. $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq .... \leq x_{(n)}$.

Example 65 *Suppose we want to generate an amplitude adjusted surrogate data set for the time series (0.7, 1.0, 0.8, 0.0, -0.7, -1.0, -0.8, 0.2). The order statistics of this series are* $x_{(1)} = -1.0$, $x_{(2)} = -0.8$, $x_{(3)} = -0.7$, $x_{(4)} = 0.0$, $x_{(5)} = 0.2$, $x_{(6)} = 0.7$, $x_{(7)} = 0.8$, $x_{(8)} = 1.0$, *which means our original series is* ($x_{(6)}$, $x_{(8)}$, $x_{(7)}$, $x_{(4)}$, $x_{(3)}$, $x_{(1)}$, $x_{(2)}$, $x_{(5)}$) *Our Gaussian series is still (1.5, -0.56, 0.85, 1.04, -1.55, -2.9, -0.44, 0.57), and the order statistics of this series are* $g_{(1)} = -2.9$, $g_{(2)} = -1.55$, $g_{(3)} = -0.56$, $g_{(4)} = -0.44$, $g_{(5)} = 0.57$, $g_{(6)} = 0.85$, $g_{(7)} = 1.04$, $g_{(8)} = 1.5$. *We then form* $y = (g_{(6)}, g_{(8)}, g_{(7)}, g_{(4)}, g_{(3)}, g_{(1)}, g_{(2)}, g_{(5)})$. *and generate a surrogate data set of this set y to get y'=(-1.0, -0.87, 0.89, 1.52, 1.33, -0.64, -1.05, -1.76). The order statistics of y' are* $y'_{(1)} = -1.76$, $y'_{(2)} = -1.05$, $y'_{(3)} = -1.0$, $y'_{(4)} = -0.87$, $y'_{(5)} = -0.64$, $y'_{(6)} = 0.89$, $y'_{(7)} = 1.33$, $y'_{(8)} = 1.52$, *and our series y' is then* ($y'_{(3)}$, $y'_{(4)}$, $y'_{(6)}$, $y'_{(8)}$, $y'_{(7)}$, $y'_{(5)}$, $y'_{(2)}$, $y'_{(1)}$). *We then form* $(x_{(3)}, x_{(4)}, x_{(6)}, x_{(8)}, x_{(7)}, x_{(5)}, x_{(2)}, x_{(1)})$ = *(-0.7, 0.0, 0.7, 1.0, 0.8, 0.2, -0.8, -1.0) as our amplitude adjusted surrogate data set.*

This algorithm produces a data set with the same marginal distribution as the original data, as it is just the original data, shuffled in a particular way so as to have autocorrelation properties similar to the original data. There is little research explaining why this method should work, nor when it does not work, the exception being a footnote in [31]. However, results quoted in [38], [36], and [37] indicate that in some cases, it works very well.

We repeated the earlier experiment to see how well long range dependence properties are preserved by the amplitude adjusted surrogate data method. The results follow.



Thus it appears that this method preserves long range dependence also.

Further developments of the method have been proposed in [31] and [33]. In [31] an iterative procedure is proposed modify the amplitude adjusted surrogate data method so that the resulting data will have almost the same autocorrelation and marginal distribution as the original data, rather than just similar autocorrelation and identical marginal distribution. Then in [33], the author proposes using an optimization procedure to create surrogate data sets that fulfill given constraints, but are otherwise random. For a recent survey, see [32].

Chapter 5

Comparison of Estimators

In this chapter, we compare the confidence intervals generated by the different methods discussed earlier in this thesis for a variety of traffic streams. Specifically, we will analyze Poisson arrivals, a Brownian motion, the On/Off fluid source model, a fractional Brownian motion, the α-FARIMA process with long range dependent, heavy tailed increments, and a real traffic trace: the Bellcore data set. These traces were chosen to represent the range of traffic considered, from classical telephone traffic (Poisson) to modern network traffic (fluid model) to as wild as we can analyze (heavy tailed and long range dependent). The real trace is included to see how well the estimators may perform in the real world.

For each of these traces, we calculate confidence intervals based on the Poisson estimator, the Brownian motion estimator, and using the Dembo estimator as the point estimator, the Moving Blocks Bootstrap, the Surrogate Data method and the Amplitude Adjusted Surrogate Data method.

For all experiments, we will calculate 90% confidence intervals of effective bandwidth at the critical point (θ^*, τ^*) with a trace of length 512.

Abbreviations for tables that follow are:

P $\rightarrow$ Poisson.

BM$\rightarrow$Brownian Motion.

MBB$\rightarrow$ Moving Blocks Bootstrap.

SD $\rightarrow$ Surrogate Data using the Dembo estimator.

AASD $\rightarrow$ Amplitude Adjusted Surrogate Data using the Dembo estimator.

LB → Lower Bound of Confidence Interval

UB → Upper Bound of Confidence Interval

Width → Width of Confidence Interval

Includes Theoretical Value?→Does the estimated confidence interval include the theoretical value (if known)?

5.1 Poisson Traffic

Here we consider a simulated Poisson stream of length 512, with parameters $\lambda = 0.3$, a queue with service rate $C = 0.4$, and buffer size $B = 5$. We calculate $(\theta^*, \tau^*) = (1.172, 9)$, and the value of the effective bandwidth is $eb(\theta^*, \tau^*) = 0.571$.

We compare the estimates based on the methods developed in this thesis

Estimator	LB	UB	Width	Includes Theoretical Value?
P	0.519	0.670	0.151	Yes
BM	0.441	0.563	0.122	No
MBB	0.436	0.857	0.421	Yes
SD	0.469	0.595	0.126	Yes
AASD	0.683	1.285	0.602	No

All methods provide reasonable answers, with, as expected, the Poisson estimator having the shortest interval that covers the actual value.

5.2 Brownian Motion

Here we consider a simulated Brownian motion stream of length 512, with estimated parameters $\widehat{\mu} = 50.83$, and $\widehat{\sigma^2} = 30.18$, a queue with service rate $C = 52$, and buffer size $B = 5$. We calculate $(\theta^*, \tau^*) = (0.0775, 4)$, and the value of the effective bandwidth is $eb(\theta^*, \tau^*) = 52$. The estimates based on the methods developed in this thesis are as follows.

Estimator	LB	UB	Width	Includes Theoretical Value?
P	52.316	53.392	1.076	No
BM	51.489	52.531	1.042	Yes
MBB	51.279	52.696	1.417	Yes
SD	51.897	52.251	0.354	Yes
AASD	51.863	52.187	0.324	Yes

Again, all values are reasonable, with the amplitude adjusted surrogate data method having the shortest interval of those that cover the actual value. This is somewhat surprising, as one would expect that the Brownian motion estimator would perform best on a Brownian motion stream. It is possible that the SD and AASD confidence intervals are too short. This would be an area for further research.

5.3 On-Off Fluid Source

Here we consider the simulated On-Off fluid source model of length 512, with parameters $\mu = 0.4$, $\lambda = 0.3$ and $h = 3.0$, a queue with service rate $C = 2$, and buffer size $B = 15$. We calculate $(\theta^*, \tau^*) = (0.7, 28)$, and the value of the effective bandwidth is $eb(\theta^*, \tau^*) = 2.006$.

We compare the estimates based on the methods developed in this thesis.

Estimator	LB	UB	Width	Includes Theoretical Value?
P	2.322	2.582	0.260	No
BM	2.282	2.660	0.378	No
MBB	2.003	2.425	0.422	Yes
SD	2.128	2.594	0.466	No
AASD	2.559	2.888	0.329	No

In this example, the moving blocks bootstrap performs best, which is not surprising since the Markov chain exhibits short range dependence, and non-Gaussian marginals.

5.4 Fractional Brownian Motion Source

Here we consider a simulated fractional Brownian motion stream of length 512 and Hurst parameter $H = 0.75$, with estimated parameters $\widehat{\mu} = 35.09$, and $\widehat{\sigma^2} = 25.53$, a queue with service rate $C = 37$, and buffer size $B = 5$. We calculate $(\theta^*, \tau^*) = (0.0355, 8)$, and the value of the effective bandwidth is $eb(\theta^*, \tau^*) = 36.364$. We compare the estimates based on the methods developed in this thesis.

Estimator	LB	UB	Width	Includes Theoretical Value?
P	35.288	36.162	0.874	No
BM	35.134	35.964	0.830	No
MBB	35.310	36.314	1.004	No
SD	35.752	35.959	0.207	No
AASD	35.849	36.208	0.359	No

None of the methods perform acceptably, although the moving blocks bootstrap is closest. This is surprising, as the moving blocks bootstrap generally does not work well for long range dependent data. Perhaps this is due to the moving blocks bootstrap confidence interval being the widest.

5.5 α-Stable FARIMA

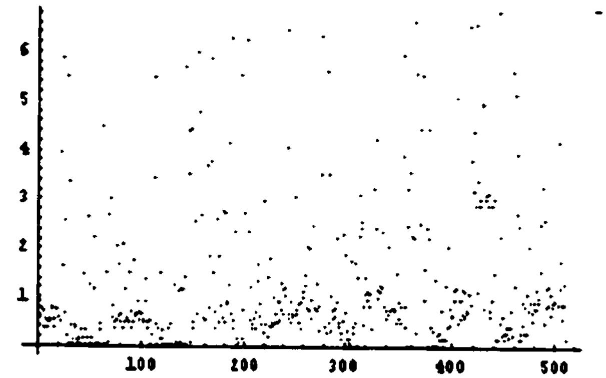
Here we consider the αStable FARIMA stream of length 512, which is long range dependent, and has heavy tailed increments. We have estimated parameters $\widehat{\mu} = 24.968$, and $\widehat{\sigma^2} = 4.058$, a queue with service rate $C = 25$, and buffer size $B = 4$. Since there is no known formula for the critical point of this traffic stream, we use the same formula as in the Brownian motion case, and calculate $(\theta^*, \tau^*) = (0.0159, 124)$, and we don't know the value of the effective bandwidth $eb(\theta^*, \tau^*)$. However, let us compare the estimates based on the methods developed in this thesis.

Estimator	LB	UB	Width
P	24.804	25.534	0.730
BM	24.850	25.150	0.300
MBB	24.792	25.172	0.380
SD	24.948	25.064	0.116
AASD	24.942	25.040	0.098

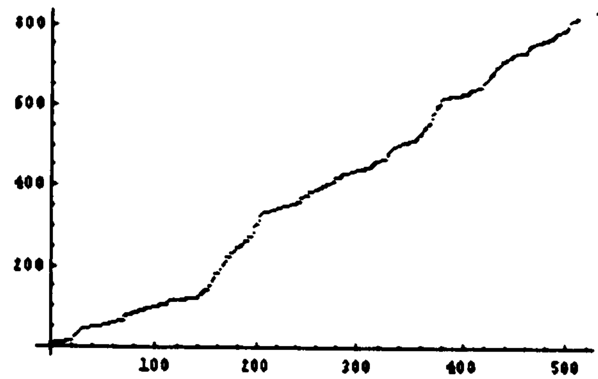
All methods give similar answers, and as intuitively one would expect the effective bandwidth to be substantially larger than the mean arrival rate in the heavy tailed case, perhaps none of these estimators is performing acceptably well.

5.6 Bellcore

The following graph shows the first 512 msec of the Bellcore data set, in 1000's of bytes per msec.



The cumulative arrivals are plotted next.



We have estimated parameters $\widehat{\mu} = 1.591$, and $\widehat{\sigma^2} = 4.169$, a queue with service rate $C = 2.9$, and buffer size $B = 0.3$ (300 bytes). Since there is no known formula for the critical point of this traffic stream, we use the same formula as in the Brownian motion case, and calculate $(\theta^*, \tau^*) = (0.628, 19)$, and we don't know the value of the effective bandwidth $eb(\theta^*, \tau^*)$. However, we compare the estimates based on the methods developed in this thesis

Estimator	LB	UB	Width
P	2.090	2.343	0.253
BM	2.627	3.195	0.568
MBB	2.252	3.753	1.510
SD	2.790	4.050	1.260
AASD	2.871	4.754	1.883

These estimates are significantly larger than the mean arrival rate, but one could conservatively estimate the confidence interval as $(2.090, 4.754)$.

Chapter 6

Conclusion

The main contributions of this thesis are:

1. A unified exposition of effective bandwidth for a variety of traffic models.
2. An emphasis on knowing the critical point of a system.
3. Presentation of several statistical methods (Bootstrap, Surrogate Data methods) that may be largely unknown to the network engineering community.
4. Evidence that the surrogate data method, and the amplitude adjusted surrogate data method are capable of working with long-range dependent data.
5. In the comparison of estimators, it is apparent that knowing the specific traffic model that generates the data has a large impact on the performance of the estimator. Knowing the model allows choosing an estimator that is best suited to the traffic stream.

The following are new:

1. The proof of Proposition 2 has not appeared in print before.
2. The proof of Proposition 4 has not appeared in print before.
3. Proposition 9, reformulated in terms of the effective bandwdith function, has not appeared in print before.

4. Statement and proof of Proposition 14.
5. Calculation of the critical point of a Poisson stream.
6. Proof of Proposition 16.
7. Calculation of the critical point of a Brownian motion stream.
8. Theorem 23.
9. Proof of Theorem 27.
10. Calculation of the critical point of a fractional Brownian motion stream.
11. Connection between θ^* and Theorem 28 for fractional Brownian motion.
12. Corollary 43.
13. Confidence interval for a Brownian motion traffic stream.
14. Theorem 45.
15. Proof of Theorem 46.
16. The evidence that, and explanation of why, the Surrogate Data, and the Amplitude Adjusted Surrogate Data methods appear to work in the long range dependent case.

Further work on the following areas would be useful and make interesting results:

1. The sampling properties of the Dembo estimator. Currently, little is known about this estimator, even though it is the most commonly used estimator of effective bandwidth. Worthwhile investigations include: determining the order of magnitude of bias of the estimator as a function of trace length; developing a bias-corrected version of the estimator; clarifying the use of the time parameter τ^* and the block size parameter b.
2. Developing theoretical justification for the applicability of the surrogate data method, and the amplitude adjusted surrogate data method to the long-range dependent case.

3. Studies comparing the estimates of loss probabilities obtained by the methods of this thesis to explicit calculations, simulation, and measured buffer occupancy methods. This would require traffic traces of sufficient length and variety so that recommendations about which estimator(s) to use can be made to teletraffic engineers.

Bibliography

[1] Bates, Stephen and Steve McLaughlin (1998). *The Effective Bandwidth of Stable Distributions.* Available from http://www.ee.ed.ac.uk/~sb/res_pubs.html

[2] Beran, Jan (1994). *Statistics for Long-Memory Processes.* Chapman & Hall

[3] Bickel, Peter J. & David Freedman (1981). Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics*, Vol. 9, No. 6, 1196-1217

[4] Billingsley, Patrick (1986). *Probability and Measure, 2^{nd} Edition.* Wiley

[5] Brémaud, Pierre (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer Verlag

[6] Brockwell, Peter J. and Richard A. Davis (1991). *Time Series: Theory and Methods 2^{nd} Edition.* Springer-Verlag

[7] Bühlmann, Peter and Hans R. Künsch (1994). *Block Length Selection in the Bootstrap for Time Series.* Available from http://www.stat.math.ethz.ch/Research-Reports/72.ps.Z

[8] Braun, W. J. & R. J. Kulperger (1997). Properties of a Fourier Bootstrap Method for Time Series. *Communications in Statistics - Theory & Methods.* 26(6), 1329-1336

[9] Casella, George & Roger L. Berger (1990). *Statistical Inference.* Duxbery Press

[10] Courcoubetis, Costas and Richard Weber (1996). Buffer Overflow Asymptotics for a Buffer Handling Many Traffic Sources. *Journal of Applied Probability*, Vol. 33. Also available from http://www.ics.forth.gr/netgroup/publications/CW95_largeN.ps.gz

[11] Crosby, Simon, Ian Leslie, John T. Lewis, Neil O'Connell, Raymond Russell and Fergal Toomey (1995). *Bypassing Modelling: an Investigation of Entropy as a Traffic Descriptor in the Fairisle ATM Network.* Available from http://www.stp.dias.ie/DAPG/dapg9513.ps

[12] Duffield, N.G, J.T. Lewis, Neil O'Connell, Raymond Russell & F. Toomey (1995). Entropy of ATM Traffic Streams: A Tool for Estimating QoS Parameters. *IEEE Journal on Selected Areas In Communications*, August 1995, Vol. 13 pp. 981-990. Also available from http://www.stp.dias.ie/DAPG/dapg9430.ps

[13] Efron, Bradley & Robert J Tibshirani (1993). *An Introduction to the Bootstrap.* Chapman & Hall

[14] Gibbens, R. J. (1996). Traffic Characterization and Effective Bandwidths for Broadband Network Traces. In *Stochastic Networks: Theory and Applications* (Editors F.P. Kelly, S. Zachary and I.B. Ziedins) Royal Statistical Society Lecture Notes Series, 4. Oxford University Press, 169-179. Also available from http://www.statslab.cam.ac.uk/Reports/1996/1996-9.ps

[15] Kaj, Ingemar (1999). *Stochastic Modeling in Broadband Communications Systems.* Lecture Note Series of the Laboratory for Research in Statistics and Probability, Carleton University. No. 327.

[16] Kaj, Ingemar (1999). *Convergence of Scaled Renewal Processes to Fractional Brownian Motion.* Available from http://www.math.uu.se/~ikaj/preprints/fbm.ps.gz

[17] Kelly, Frank (1996). Notes on Effective Bandwidth. In *Stochastic Networks: Theory and Applications* (Editors F.P. Kelly, S. Zachary and I.B. Ziedins). Royal Statistical Society Lecture Notes Series, 4. Oxford University Press, 141-168. Also available from http://www.statslab.cam.ac.uk/~frank/eb.ps

[18] Künsch, Hans R (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics, Vol.* 17 1217-1241

[19] Kokoszka, Piotr S. & Murad S. Taqqu (1996). Parameter Estimation for Infinite Variance Fractional ARIMA. *The Annals of Statistics*, Vol. 24 1880-1913

[20] Lahiri, Soumendra Nath (1993). On the Moving Block Bootstrap Under Long Range Dependence. *Statist. Probab. Lett.* 18, no. 5, 405–413. Also available from http://emlab.berkeley.edu/wp/nsf96/lahiri.ps

[21] LePage, Raoul & L. Billard, eds (1992). *Exploring the Limits of Bootstrap*. John Wiley & Sons, Inc. New York

[22] Lewis, John T. & Raymond Russell (1996). An Introduction to Large Deviations for Teletraffic Engineers. Available from http://www.stp.dias.ie/DAPG/LDtut96.ps

[23] Liu, R. Y & K. Singh (1992). Moving Blocks Jackknife and Bootstrap Capture Weak Dependence. *Exploring the Limits of Bootstrap*, Eds. Raoul LePage & L. Billard. John Wiley & Sons, Inc. New York, 225-248

[24] Mikosch, Thomas, Sidney Resnick, Holder Rootzén, and Alwin Stegeman (1999). *Is Network traffic Approximated by Stable Lévy Motion or Fractional Brownian Motion?* Available from ftp://ftp.orie.cornell.edu/pub/techreps/TR1247.ps

[25] Norros, Ilkka (1995). On the use of Fractional Brownian Motion in the Theory of Connectionless Networks. *IEEE Journal on Selected Areas In Communications*, August 1995, Vol. 13 pp. 953-962

[26] O'Connell, Neil (1999). *Large Deviations with Applications to Telecommunications*. Available from http://www.stp.dias.ie/LDtut96.ps http://www.maths.lth.se/mathstat/research/asn/ldcourse99notes.ps

[27] Peligrad, Magda (1998). On the Blockwise Bootstrap for Empirical Processes for Stationary Sequences. *The Annals of Probability*, Vol. 26, No. 2, Pg 877-901

[28] Politis, Dimitris N, Joseph P. Romano and Michael Wolf (1999). *Subsampling*. Springer-Verlag

[29] Samorodnitsky, Gennady and Murad Taqqu (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall

[30] Siris, Vasilios A (1999). *Large Deviation Techniques for Traffic Engineering*. Available at http://www.ics.forth.gr/netgroup/msa

[31] Schreiber, Thomas & Andreas Schmitz (1996). Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters*, Volume 77, Number 4, Pg. 635-638.

[32] Schreiber, Thomas & Andreas Schmitz (1999). *Surrogate Time Series.* Available at http://xxx.lanl.gov/ps/chao-dyn/9909037

[33] Schreiber, Thomas (1998). Constrained Randomization of Time Series Data. *Physical Review Letters*, Volume 80, Number 10, Pg.2105-2108.

[34] Shwartz, Adam and Alan Weiss (1995). *Large Deviations for Performance Analysis: Queues, Communications and Computing.* Chapman Hall

[35] Taqqu, Murad S, Walter Willinger and Robert Sherman (1997). Proof of a Fundamental Result in Self-Similar Traffic Modeling. *Computer Communications Review*, 27, 5-23

[36] Theiler, James, S. Eubank, André Longtin, B Galdrikan & J. Doyne Farmer (1992). Testing for Nonlinearity in Time Series: the Method of Surrogate Data. *Physica D* 58

[37] Theiler, James, Bryan Galdrikan, André Longtin, Stephen Eubank & J. Doyne Farmer (1992). Using Surrogate Data to Detect Nonlinearity in Time Series. *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity*, Eds. Martin Casdagli and Stephen Eubank, Addison Wesley

[38] Theiler, James, Paul S. Linsay and David M. Rubin (1994). Detecting Nonlinearity in Data with Long Coherence Times. *Time Series Prediction: Forecasting the Future and Understanding the Past, SFI Studies in the Sciences of Complexity*, Eds. Andreas S. Weigend and Neil A. Gershenfeld, Addison Wesley

[39] Townsend, J. Keith, Zsolt Haraszti, James A. Freebersyser and Michael Devetsikiotis (1998). Simulation of Rare Events in Comunications Networks. *IEEE Communications Magazine*, August 1998, 36-41

[40] Wischik, Damon (1999). *Large deviations and Internet Congestion.* Available from http://www.statslab.cam.ac.uk/~djw1005/Stats/Research/phd.html

[41] Young, G.A. (1994). Bootstrap: More than a Stab in the Dark? *Statistical Science* 9 382-415